The influence of memory on driving behavior: How route familiarity is related to speed choice. An on-road study

Original Citation:
The influence of memory on driving behavior: How route familiarity is related to speed choice. An on-road study / Colonna, Pasquale; Intini, Paolo; Berloco, Nicola; Ranieri, Vittorio. - In: SAFETY SCIENCE. - ISSN 0925-7535. - 82:(2016), pp. 456-468. [10.1016/j.ssci.2015.10.012]

Availability:
This version is available at http://hdl.handle.net/11589/78070 since: 2021-03-03

Published version
DOI:10.1016/j.ssci.2015.10.012

ABSTRACT

Differences in driving behavior due to the presence of users familiar (or unfamiliar) with the road are considered in the road and traffic engineering. However, although considered, the matter is largely unexplored: there is a lack of theoretical foundations and data on determining the impact of route familiarity on accident rates, speed choice and risk perception. On the other hand, some literature studies confirm that route familiarity is influential on driving behavior, encouraging research in this sense.

This paper reports the results of an on-road test carried out on a two lane rural road in the District of Bari in the Puglia Region (Italy) over six days of testing by following this time schedule: first four tests in four consecutive days, the fifth test in the ninth day after the first test and the sixth test in the twenty-sixth day after the first test. The main aim of the experiment was to find relationships between route familiarity and speed choice. In particular, speed data were analyzed by considering the influence of road geometry and human factors.

The main finding is that speed choice seems to be affected by route familiarity: speed increases with the repetition of travels on the same route. The particular schedule used for the tests allows to consider the influence of memory on the speed behavior of the test drivers. Moreover, some relationships between changes in speed over days, road geometry and drivers' attitudes were shown.

KEYWORDS

Route familiarity
Road safety
Speed choice
Driving behavior
On-road study

1. INTRODUCTION

Traffic safety policies can be implemented in different ways: enforcement, increasing user awareness, and engineering countermeasures. These policies should be defined by technicians and different experts: engineers, psychologists and economists in cooperation.
The engineering part of the matter involves interventions on existing roads in order to reduce the expected number of accidents (see for example [1], [2], [3]). After sites of intervention have been identified, a countermeasure should be implemented. If the road infrastructure is recognized as the supposed or real main cause of accidents, the countermeasure should come from engineering.
However, apart from the method employed for choosing countermeasures, there is a lack of theoretical approaches able to take into account users' reactions to modifications in infrastructure. This phenomenon is not secondary because risk compensation is considered by different sources as a problem influential in safety [4], [5], and in particular if the safety countermeasure is visible to drivers [6] (see van der Horst [7] for a recent summary about experimental evidences of behavioral adaptation to countermeasures). If an engineering safety measure modifies user behavior, who acts pursuing the aim of minimizing travel disutility, which depends on several factors [8], [9], [10], then the countermeasure could be useless or detrimental. In fact, in the case of adaptation, the possible increase in speed could lead to a mobility benefit (reduction in travel time) but also to a worsening of accident risk [11], [12]. Moreover, the relationship between speed and accident risk is well-known. It can be considered as a power function [13] or as an exponential function (especially for injury accidents [14]): the accident risk increases more if speed is higher.
Hence, in order to forecast the effectiveness of a countermeasure, it is necessary to consider driver behavior. However, driver behavior is not characterized by a universally accepted theory, because of the various factors involved in the process [15]. For example, the zero-risk model [16], the risk homeostasis theory [5], the rule-based model [17], the risk allostasis theory [15] and/or the risk monitor model [18] could be taken into account.
Speed choice is one of the main indicators of driver behavior and it is influenced in turn by many factors, among which risk perception is crucial [19]. The way in which users perceive accident risk while they are driving is a topic currently studied, a perplexing topic due to the lack of consensus about measuring risk and users' risk misperceptions [20]. One method to measure risk is cognitive heuristics: in uncertain conditions, decisions are not deterministic but they are influenced by experience acquired over time through empirical observations. This process is recognized as influential in risk perception and as closer to reality [21], [22], even if sometimes this method could lead to errors or imprecision [23], [24]. The heuristic approach is coherent with the process of speed selection (connected to the risk perception) which it is often based on users' misperception of risk and travel time [25].
By applying the cognitive heuristic concept to driver behavior, it is possible to identify one influential feature in drivers' behavior: the familiarity with a route (on which this paper is focused) determined by the habit of driving on it, while acquiring experience and information. There is some research about the relationships between route familiarity and driving performances. Yanko and Spalek [26] e.g. carried out an experiment involving 20 drivers and a driving simulator. They found that route familiar users (users who had driven on the experimental route four times before the test) needed greater reaction times than route unfamiliar users (users who drove on the experimental route for the first time during the test) in order to respond to unexpected external stimuli simulated in the presented scenarios. The results obtained from the presented experiment are similar to what Martens and Fox [27] suggest about route familiarity: it can lead to a greater distraction while driving, probably because familiarity could increase the effect of "mind wandering". Mind wandering occurs when the mind is occupied by thoughts not concerning the task being undertaken and so, responses to external stimuli are potentially slowed down. This interpretation is coherent with the MART theory presented by Young and Stanton [28], which assumes that driving performance varies as a function of mental workload and that in low demand conditions (normal driving tasks) attention capacity is reduced. The matter of risk underestimation related to route familiarity was considered also by Rosenbloom et al.

[29], who observed the driving behavior of a sample of female drivers in both familiar and unfamiliar locations. They found that drivers performed more traffic violations, more dangerous behaviors and speeding while driving in more familiar locations, confirming that risk perception could change with the acquired route familiarity.
From an engineering point of view, the matter of familiarity is considered in the traffic flow theory and in the road design guidelines.
In fact, within the framework of the level of service (LOS) calculation for highways and freeways, the Highway Capacity Manual [30] suggests the following formula in order to calculate the equivalent flow rate (higher equivalent flow rates correspond to lower LOS), taking into account vehicular composition of traffic flow:

$$Vp = \frac{V}{PHF*N*f_{HV}*f_p} \tag{1}$$

where**:**
V_p = 15-minute passenger-car equivalent flow rate (pcphpl);
V = hourly volume (pc/hr);
PHF = Peak Hour Factor;
N = number of lanes in one direction;
fHV = heavy-vehicle adjustment factor;
fp = driver population adjustment factor.

The introduction of the f_p factor in the equivalent flow rate (V_p) calculation makes it possible to implicitly consider users as divided into two categories according to their familiarity with a route:

• Users familiar with the route: in general all those who drive on a given route almost daily (regular users), such as commuters;

• Users not familiar with the route: all those who infrequently drive on the route, such as tourists or other non-habitual (recreational) drivers.

HCM 2010 considers f_p = 1 in the case of traffic mainly consisting of regular users and a value between 0.85 and 1 for traffic with a more or less significant component of recreational users. This means that other conditions being equal, a decrease in f_p down to a minimum of 0.85, corresponds to an increase in the V_p of up to about the 20% more than the value calculated for f_p equal to 1. In the context of uninterrupted flows, an increase in the V_p (equivalent traffic flow rate) is related to an increase in the car density (equivalent passenger cars/km) and consequently this leads to worsening in the level of service of the road. Therefore, according to this method, the presence of recreational users leads to an evident deterioration in the LOS of the road.
Considering that differences between users familiar and unfamiliar with a given route are influential on flow rate, it could be assumed that accident rates should also be different between the two categories of users. In fact, it is commonly accepted that route familiarity is a factor influencing speed choice and trade-offs between travel time and safety (see e.g. [31]). However, accident rates have not been largely related to familiarity in literature studies. Instead, this relationship would conduct to noticeable results, as can be verified by considering e.g. Blatt and Furman [32], who found that people are most likely to be involved in crashes on roads on which they traveled most frequently (among the considered sample, most of the rural residents involved in fatal crashes were traveling on rural roads while urban residents were primarily involved in urban accidents).
Moreover, a good practice for road designers should be the consideration that users are driving on a roadway for the first time and that they have no familiarity with its features [31].
So, even if theoretically assumed as an influential factor in the road and traffic engineering, the impact of route familiarity on driving behavior and traffic safety was not adequately studied by measuring, for example, accident rates for different compositions of traffic flow (tourist/commuters), by understanding the process responsible for making an unfamiliar user familiar with a given route and/or by estimating possible variations in speed choice based on on-road experiments.

As explained, the crux of the problem in dividing the familiar drivers from the unfamiliar ones is represented by the habit connected to a given route. From a merely psychological point of view, the effect of habituation has been explained by various theories, such as the early study by Groves and Thompson [33]. They supposed the existence of two parallel and interacting processes in the central nervous system: the habituation process and the sensitization process. Both processes handle external inputs and generate behavioral outputs: the response to an external stimulus depends on which process is prevailing. In the habituation process, the response decreases with the repetition of the same stimuli over time until it reaches an asymptotic constant value (habituation effect). When the stimulus is withheld after response decrement, the response recovers at least partially over the observation time. However, some stimuli repetitions may result in response decrement that last hours, days or weeks: this persistence is called long-term habituation [34]. Instead, in the sensitization process, in a first phase the response increases with the repetition of the same stimuli over time and after it decreases. The behavioral response to repeated stimuli is the output of the interaction of these two processes and one of them can prevail. The final behavioral output depends on the stimulus presented. Moreover, if a novel stimulus presents itself at the end of the habituation process, then the response increases again and after decays to its previous habituated level, independently of the interruption or not of the habituation stimulus after the novel stimulus presentation (dishabituation effect, which can be seen as sensitization to the novel stimulus [33]). The three explained effects are summarized in Figure 1.
Considering the above cited studies and the diagrams in Fig. 1, the hypothesis advanced is that the dual-process theory could be applied to the case of driving, in which the habituation effect could prevail. In fact, driving on a familiar route is mostly an automatic process, in which skill-based tasks are unconscious [35]. In other words, in this case, driving on the same route can be identified as the stimulus repeated many times over time for a driver, who would get the habituation condition corresponding to the asymptotic response value. Hence, this theoretical asymptote should correspond to the acquired familiarity, equivalent to a condition of minimum energy, which seems the normal driving condition for route familiar users [36] and in which users potentially reduce danger levels and underestimate risk because driving task is simpler (see e.g. [37]). For these reasons, in this study, the driving familiarization process will be compared with the habituation effect, since the sensitization effect implies that drivers should be more responsive to external stimuli in the first stages of the process.
During the habituation process, the decrease in response of users getting familiar with the route (who leads to the explained changes in risk perception) could be related also to speed choice. In fact, speed choice depends on risk perception and so, a decrease in response to external stimuli connected to underestimating risk could lead to an increase in speed over time until a theoretical asymptotic value corresponding to the acquired familiarity condition. Some indications in this sense come from Colonna et al. [38], [39], [40] who also studied relationships between route familiarity, risk inclination and road geometry.
However, the presented matter should be studied more in depth in order to provide theoretical support for technical choices regarding the different composition of traffic flow and its impact on safety. The first step of this in-depth analysis should be the accurate definition of the condition in which users can be defined as familiar with a given route. In order to obtain this result, the process responsible for making an unfamiliar user familiar with a given route and the related variations in speed choice and risk perception should be studied before (see also Colonna et al. [38]). For this purpose, the main aim of this paper is to investigate in detail the relationships between variations in speed choice and increasing route familiarity by considering an on-road experiment. The investigation of speed behavior based on a real world setting has the advantage of producing data with the greatest validity in comparison with those obtained in a simulated scenario (see e. g. [41]).
The remainder of the paper summarizes the methods employed for the on-road experiment and the data obtained (section 2), the methods employed for data analysis (section 3) and the presentation and discussion of results (section 4). In particular, some important features such as the variations in speed

over the days of testing will be related to route familiarity and controlled for other variables as a result of the analysis.

2. METHODS

In order to investigate speed changes over time due to the acquired route familiarity, an on-road experiment was planned. Details about participants, routes, apparatus, procedure and measures are given below.

2.1 Participants

The road familiarization process and its effects on driver behavior needed to be observed. So, the aims of the study required that, before the experiment, all drivers had not to be confident with the road chosen for the tests. Furthermore, the sample of drivers should be at least age-homogeneous in order to minimize confounding factors, and gender-representative (drivers with similar ages and with genders enough equally represented).

For this reason, participants were recruited among students of the Polytechnic University of Bari by using advertisements requesting volunteers for an experimental study on driving behavior. A questionnaire about general information and driving habits was submitted to all respondents. Age, sex, driving experience, mileage, availability of their own car and unfamiliarity with the route selected were chosen as selection criteria for the definition of the final sample.

Students under 22 years of age were not included in the final sample due to their possible lack of experience in respect to the other drivers. Students at least 22 years old but licensed for less than 3 years were not included for the same reason (Italian Road Code consider as “new-licensed” people who are in their first three years of license). Furthermore, drivers who declared to experience a mileage of less than 10 km (on average) per week on rural roads (the roads chosen for the experiment) were not included in the final sample. All these exclusions were made in order to meet the following general principle: the “unfamiliar with the road” condition must not be confused with the “inexperienced driver” condition. The unfamiliarity with the route was checked by choosing students who reported never having driven in the past within the municipality of Cassano delle Murge (in which the selected route lies).

Finally, the last condition to meet was the availability of the car that they were driving usually. In fact, drivers, even if experienced, could modify their driving behavior when using another car for the first time. This occurrence was incompatible with the aims of the experiment for the same reasons explained above.

The final sample was defined by meeting all the above explained requirements. However, due to the few subscriptions made by female students to the experiment, genders were not perfectly equally represented.

So, the test sample was composed by 20 drivers, characterized by the following features: age: 24.45 ± 1.10 years old, 16 males and 4 females, years licensed: 5.75 ± 1.25 years. The sample size is consistent with a previous similar study which involved a driving simulator [26]. Both the age and the experience of the drivers are relatively low compared to the actual driving population. However, Martens and Fox [27] found some relationships between familiarity and driving behavior indicators for a sample of drivers with age included between 21 and 46. Therefore, familiarity could probably affect behavior of drivers of different ages. Further studies could help in understanding if the relationship between familiarity and driving behavior can be influenced by the variable age.

2.2 Driving routes

Two stretches of two-lane two-way rural roads (SP31 and SP18, situated in the municipality of Cassano delle Murge, district of Bari, Italy) were chosen as driving test routes.

As road characteristics can affect the results of the study, two isolated roads and with very low traffic conditions were chosen for the experiment in order to ensure free flow characteristics during tests.

Furthermore, two different stretches were chosen, instead of selecting only one of them, because in this way both horizontal and vertical variabilities in the alignments were taken into account. Some segments of the stretches were not analyzed due to the presence of driveways and intersections (see also 2.5).
In fact, the stretch 1, belonging to the road SP31 (see Fig. 3 and 5), is mainly horizontally varying and vertically homogeneous. It is composed by seven sharp curves (four of them are characterized by a radius of curvature less than 100 m) and seven small tangents, while the elevation profile is mainly flat (maximum grade is about 0.01). On this stretch, the posted speed limit is 50 km/h.
The stretch 2, belonging to the road SP18 (see Fig. 3 and 5), is mainly horizontally homogeneous and vertically varying. It is composed by three curves (radius of curvature greater than 150 m) and four tangents (one of them is longer than 1000 m), while the elevation profile is characterized by steep grades (maximum grade is about 0.07) and counter-slopes. On this stretch, the posted speed limit is 70 km/h.
The road stretches analyzed after rejecting segments near intersections and significant driveways were composed by two curves and two tangents for the stretch 1, and by one tangent and one curve for the stretch 2 (see Fig. 5). Anyway, speed choice is influenced by the overall road geometry and the road coherence (e.g.: by the sequence of horizontal and vertical curves [42]). Therefore, even after the rejection of some segments, a greater horizontal variability of the alignment can still be related to the stretch 1 and a greater vertical variability of the alignment can still be related to the stretch 2.

2.3 Apparatus

All users selected for the driving test used the car that they were driving usually.
Speed data were collected by using the Differential Positioning GPS technology (Dynamic Method). This technology allowed to orientate any point with respect to a fixed one, by calculating the baseline vector for the two points. That vector has been transformed into three parts with each part being directed along three perpendicular coordinate axes, with the aim of obtaining three-dimensional information. This way, distance was measured along every coordinate with an accuracy of a few millionth parts of the distance. That accuracy was better than the one derivable from the same measurement made with other standard geodetic surveys.
Two receivers were necessary to achieve the GPS Differential Positioning. Each of them was put into the baseline's extremities and they worked during all the survey campaign. Thanks to this technology, it was not necessary that the two receivers were always visible with one another.
The first receiver, the fixed one, was composed of an adjustable height tripod and a GPS antenna on the tripod top. Once the tool was assembled, it was necessary to align the instrument with a survey point (a point which has highly accurate GPS coordinates) by using a viewfinder. Furthermore, in order to obtain an almost perfect horizontal system, it was necessary to adjust the bull's eye level by using the dedicated screws. Finally, the height of the antenna above ground was measured.
The second receiver, the mobile one, was composed by a Rover antenna and a recorder. The antenna was placed on each car used by the test drivers in order to guarantee visibility and functionality. The recorder, connected to car battery, was located inside the car and it had the task of recording the antenna position with respect to the fixed point (baseline) on a USB pen drive.
Data were collected by using the software released by the company producing the GPS antenna. This software solved the GPS fundamental equation by using three-points triangulation considering visible satellites, fixed antenna and Rover antenna. In this way, during the test, the exact positioning was obtained by repeating measurement every second. This technique leads to an average location accuracy to within 10 cm and an average speed accuracy to less than 1 km/h.

2.4 Procedure

Before the driving test, each user was trained by a researcher on how to prepare the instrumentation (the on-board receiver, the recorder and the fixed antenna) and he received instructions on the driving task to accomplish.

The complete driving test (see Fig. 3) consisted on traveling along a route composed of the above mentioned two stretches of road (see 2.2) in the following order:

- Stretch 1, from the starting point (Start) to the intersection with stretch 2 – way there;
- Stretch 2, from the intersection with stretch 1 to the end point (End) – way there;
- Stretch 2, from the end point to the intersection with stretch 1 – way back;
- Stretch 1, from the intersection with stretch 2 to the starting point again – way back.

The total trip length is about 14 kilometers, from the Start to the Start again.
Drivers were asked to drive freely on this route without any other instruction.
Furthermore, users drove alone without any conditioning due to the presence of researchers during all tests.
For the purpose of the study, users were asked to repeat the same driving test described above six times in six different days. The chronological schedule (Fig. 2) of the test repetitions over time were fixed for each driver. The first four tests were scheduled in four consecutive days (1st, 2nd, 3rd and 4th days of testing). The other two tests were fixed in the ninth day after the first test (5th day of testing) and in the twenty-sixth day after the first test (6th day of testing). The chosen schedule is similar to the experimental plan used by Martens and Fox [27]. However, as in this study also the possible presence of a long term memory effect after some interruptions in administering stimuli (represented by driving tests) was investigated, the fifth test was postponed and a sixth test more distant in time was introduced.
All driving tests were planned in spring, from April to June, in order to limit weather variability (average monthly rainfall in that area for that period is 42 mm). Furthermore, all tests were performed in the daylight condition for the same purpose of homogeneity of the environmental variables. However, drivers were asked to report any adverse weather conditions during tests. In the same way, they were also asked to report particular traffic situations (including car-following) impeding the free-flow condition.

2.5 Measures

Driver behavior was observed in terms of speed and trajectory. For the purpose of this study, only speed data were considered.
Due to the employed technology (see 2.3) and the frequency of data collecting (one measurement per second), values of punctual speed were obtained. Speed profiles were drawn for each user and each test by putting punctual speed on the Y axis and distance on the X axis (see Fig. 4).
Data belonging to one of the twenty drivers were discharged from all the further analyses due to the large amount of missing data (more than 40 % of the total).
After, cross-sections were positioned along the driving routes each 25 meters. Cross sections were not placed in segments of the stretches near to intersections or significant driveways. In fact, those areas could have an unverifiable influence on the speed of each driver.
Finally, 61 road cross-sections on the stretch 1 and 76 road cross-sections on the stretch 2 were identified along the driving routes (marked with the red color in Fig. 5).
Speed data were assigned to each road cross-section so defined by connecting the value of distance corresponding to each cross-section to the respective value of speed in the speed profile.
However, before the data analysis, speed data corresponding to situations of adverse weather or adverse traffic conditions, based on experience reported by the drivers, were discharged from the dataset.
Commonly, road geometric features are highly related with speeds (see e. g. [43]). Therefore, the variable “road geometry” should be controlled while studying changes in speed over days of testing. However, the detailed consideration of all the road geometric elements could be misleading in respect of the aims of this study.
So, the available sight distance was used as a synthetic variable representing road geometric characteristics. The available sight distance is the unhindered length of road section that the driver can see ahead without considering the influence of traffic, weather and lighting. Sight distance takes

into account both the horizontal and vertical alignments and can be computed for each direction of travel.

A value of the sight distance was assigned to each road cross-section for both directions of travel by using the method of the Italian Road Design Standard [44] and video recordings of the paths in both directions. In this way, the sight distance profile (Fig. 6), for each stretch of road and for both directions of travel, was obtained.

Cross-sections were clustered into four classes in respect to their computed value of sight distance. The four visibility classes were so defined:

- class 1: cross-sections with low sight distance (0-100 m);
- class 2: cross-sections with medium-low sight distance (100-200 m);
- class 3: cross-sections with medium sight distance (200-400 m);
- class 4: cross-sections with high sight distance (400-600 m).

The low visibility interval was chosen considering that sight distances of about 100 m are indicated as critical sight distances, that is, the accident rate increases rapidly for smaller sight distances [45]. Furthermore, the high visibility interval was chosen according to Lamm et al. [46] who found that accidents related to passing maneuvers increase when the sight distance is less than 400 m to 600 m. The intermediate interval was split into two classes (medium-low, from 100 m to 200 m; and medium, from 200 m to 400 m) in order to divide the remaining cross-sections into subsets more numerically homogeneous. In fact, cross-sections with sight distances included between 100 m and 400 m were the most numerous.

From this classification, 53 cross-sections with a low available sight distance, 110 with a medium-low available sight distance, 38 with a medium and 73 with a high available sight distance were obtained. (The same 137 road cross sections were considered two times because sight distances were computed in the two different directions of travel).

3. DATA ANALYSIS

The purpose of this paper is to show how the driving speed behavior is influenced by the memory of the road and how this relationship can be conditioned by other factors such as the road geometry and the human factors.

The analysis of experimental speed data was divided into three phases. In the first phase, statistical tests were performed to verify if speeds of the various days of testing and visibility classes were significantly different. In the second phase, cluster analysis was employed to categorize individual drivers into groups with similar behaviors. Classification of speed profiles is necessary to allow the interpretation of speed measurements in terms of road user behavior, as long as changes in driving behavior could be influenced by driving characteristics. In the third phase, the relationship between speed and days of testing, in regard to the different visibility classes and drivers' clustering was analyzed. A preliminary analysis has showed that a simple regression analysis could have been unsuitable, so the chosen technique was the piecewise linear regression, in which regression lines are fitted with breaks in the slope.

3.1 Statistical tests of speed data

Speed data were pre-processed by testing the normality and homoscedasticity assumptions. Since different tests of normality often produce different results [47], the normality assumption was verified using the Anderson-Darling, Jarque–Bera, Kolmogorov-Smirnov, Lilliefors, and Shapiro-Wilk tests. The homoscedasticity assumption of the speed data distribution was verified using the Fisher’s test. Results of the tests carried out suggest that the normality and homoscedasticity assumptions cannot be rejected at the 5 % level of significance.

Given that data distributions are normal and homoscedastic, speed data were compared by parametric tests. To evaluate the presence of an overall effect, the mixed ANOVA test was performed considering, separately, all days of testing and all visibility classes, while Bonferroni post-hoc tests were carried out to isolate where the differences are. In detail, in the first analysis, it was tested

whether there is a difference in mean speed between the six days of testing, whereas the six days of testing as fixed effect, and the 19 drivers as random effect. The mixed ANOVA was chosen as long as the individual process of speed choice can be influenced by the human factors, and this was thought as an idiosyncratic factor affecting all responses from the same subject. Thus, in this way, the different responses can be rendered as inter-dependent rather than independent.
In the second analysis, it was tested whether there is a difference in speed mean between visibility classes, whereas visibility classes as fixed effect, and the 19 drivers and the six days as random effects.

3.2 Cluster analysis of speed data

Cluster analysis was carried out in order to group drivers into clusters characterized by similar speed behaviors.
The pattern recognition techniques perform quite well in classifying the behavior types compared to classification by a human observer. The advantage of these techniques is the automation of the classification process which allows for analyzing large datasets. Pattern recognition techniques might help reveal the relations between this subjective dimension and objective variables, contribute to standardization and therefore allow for larger comparability between analyses made by different individuals.
Cluster analysis is a multivariate statistical methodology aimed at partitioning N observations into K disjoint groups in order to obtain their maximal internal homogeneity and their external heterogeneity [48]. In the definition of these groups, a distance measure must be defined. As distance measure, the Euclidean distance was adopted.
Consistently with previous studies [49], [50] a nonhierarchical cluster analysis was performed, namely the K-means algorithm. This algorithm works as follows: (a) Step 1, the number of clusters K is chosen; (b) Step 2, random N initial means are selected as starting points for the clusters; (c) Step 3, for each series, the similarity measure with each mean series is computed; each series is assigned to the cluster whose mean series has the highest similarity with the time series; (d) Step 4, means are updated; and (e) Step 5, Step 3 is repeated until no reallocation in the cluster occurs after the updating step or a maximum number p of iterations is performed.
After cluster analysis, a silhouette analysis was performed. The silhouette analysis is a method used to validate and interpret the results of clustering. After clustering, it is assumed that data have been divided into K clusters. For each object I, the average distance of *i* from all other data within the same clusters is called a(i). This measure represents how well matched *i* is to the cluster it is assigned (as this quantity decreases, the matching is improved). Instead, let b(i) the minimum average distance of *i* from the data of other clusters. Now it is possible to define the silhouette value s(i) as:

$$s(i) = \frac{b(i)-a(i)}{max(a(i),b(i))}\ ,\ where: -1 \leq s(i) \leq 1. \qquad (2)$$

The average s(i) value of a cluster is a measure of the tight grouping of data in the cluster. Instead, the silhouette mean value related to all the objects in the sample is a powerful tool for determining how reliably data were clustered. Mean values greater than 0.6 are considered acceptable [48]. Silhouette plots can be used, jointly with mean values, in order to individuate narrower silhouettes of some objects among the cluster. The number of partitions corresponding to the greatest silhouette mean value related to all the objects belonging to the sample was selected as the most effective.

3.3 Piecewise linear regressions speed/days

Finally, it was studied how the speeds change depending on the days. Analyzing the relationship between a response variable, speed, and an explanatory variable, day, it was observed that for different ranges of days, different linear relationships occur. In these cases, a single linear model may not provide an adequate description and so, piecewise linear regression was carried out [51], [52]. Piecewise linear regression is a form of regression allowing multiple linear models to be fit to data for different ranges of the independent variable, x. Breakpoints (c) are values on the x-axis where a change in the slope of the linear relationships can be identified. When there is only one breakpoint (at $x = c$), the model can be written as: $E[y] = \alpha_1+\beta_1 x$ for $x \leq c$, $E[y] = \alpha_2 + \beta_2 x$ for $x > c$. This can be extended to cases where more breakpoints are present. The actual number and the location of the breaks are not known, and both physical and statistical criteria should be considered in determining the number of breaks in a slope. It is rather difficult to fit a truly optimal number of knots since the possible permutations of knot placements would quickly increase. The piecewise linear regression procedure consists of the following iterative steps: (1) the location of the break points; (2) estimating a linear regression model; (3) linear hypothesis testing; and (4) the assessment of the statistical significance of the break points in the model.

In this study, breakpoints were searched through a graphical analysis consistently to the previous findings. Instead, for each segment, the parameters α_i and β_i were estimated through simple linear regressions.

For each segment, the linearity hypothesis was tested through the t-student test using the bootstrap method. The null hypothesis is $H_0 : \beta_i = 0$, and the corresponding alternative hypothesis is $H_1 : \beta_i \neq 0$. The significance level α was set to 0.05. If the null hypothesis is true, the mean of population of y is α_i for every x value, which tells us that x has no effect on y, i.e. speed is constant. The alternative is that changes in x are associated with changes in y. Therefore, rejecting the null hypothesis equates to concluding that there is a linear relationship between speed and the days. If the p value is larger than α, than the null hypothesis is accepted.

Standard tests of significance do not state whether the inclusion of a single breakpoint provides a real improvement in overall fit compared with a simple linear specification. Such an improvement would only be apparent if H_1: $\beta_i \neq \beta_{(i+1)}$. Bootstrap procedures were used in order to answer this question [53]. The entire procedure was performed in Matlab environment.

4. RESULTS AND DISCUSSION

Results from the data analysis are presented and discussed in this section, by considering separately the three analyses performed.

4.1 Speed variations over days of testing

4.1.1 Differences in mean speed between the six days of testing

Means, standard deviations and percentage differences of observed speed for each day of testing are shown in Table 1.

Results from the one-way mixed ANOVA (testing differences in mean speed between the six days of testing, whereas the six days of testing as fixed effect, and the 19 drivers as random effect) are reported below.

A significant effect of days of testing on speed at the $p <.05$ level was found [$F (5, 90.612) = 14.939$, $p < 0.001$].

Furthermore, results from the Bonferroni test (Table 2) revealed that speed is statistically significantly lower in the first day of testing (79.068 ± 12.649 km/h) compared to all other days. Similarly, speed is statistically significantly lower in the second day of testing (83.802 ± 15.459 km/h) compared to days 3, 4, 5 and 6, in the third day of testing (87.205 ± 16.195 km/h) compared to days 4, 5 and 6 and in the fifth day of testing (88.442 ± 16.832 km/h) compared to day 6. Instead, there are no statistically significant differences between the fourth day (88.802 ± 15.845 km/h) and days 5 and 6 (89.515 ± 14.735 km/h).

Findings from the statistical analysis can be verified by looking at boxplots of speeds in the six days of testing (see Fig. 7).

A significant increase of mean speed over days can be noted while going from the first to the fourth day of testing. Instead, there are only slight differences between days 4, 5 and 6.

Moreover, a significant effect of the driver factor on speed at the $p < .05$ level was found [$F(18, 88.165) = 16.795$, $p < 0.001$]. Furthermore, there was a statistically significant interaction between drivers and days of testing on speed, [$F(88, 27682) = 36.615$, $p < 0.001$].
Results showed that, on average, the route learning process of the first four days of testing leads to an increase in speed. The repetition of the stimulus "driving test" on the same route for four consecutive days significantly affects driving behavior and in particular drivers' speeds. Furthermore, when the repetition of stimuli is interrupted and restarted after two longer time intervals (six days between the fourth and fifth days of testing and seventeen days between the fifth and the sixth days of testing), a significant long-term memory effect can be noted. In fact, speed does not vary significantly in the fifth and sixth days of testing in respect to the fourth day. Hence, the memory of the drivers who acquired familiarity with the route seems to influence speed, which maintains almost constant over time, independently from the number of days between one stimulus and the following one. Moreover, the highlighted memory effect indicates that four consecutive days seem to be a sufficient time to become confident with the chosen route. This finding is confirmed by the values of percentage differences in Table 1. In fact, even if speed increases over the first four days, the increasing rate decreases over days until it reaches the minimum in the fourth day of testing, where it is closer to zero (0.018).
Furthermore, the significant effect of drivers on speed and the significant interaction between drivers and days of testing on speed indicate that behavioral differences among the test drivers and the evolution of behaviors over time should be studied at a more detailed level (see 4.2 and 4.3).

4.1.2 Influence of road geometry (in terms of visibility)

Results shown in the previous paragraph did not consider the road geometric layout as a variable able to predict speed.
So, cross-sections were clustered into four classes in regard to their value of sight distance (see 2.5). Means and standard deviations of observed speed for each day of testing and for each visibility class are shown in Table 3. Speed/days diagrams were drawn for each visibility class (Fig. 8).
Results from the one-way mixed ANOVA (testing differences in mean speed between visibility classes, whereas visibility classes as fixed effect, and the 19 drivers and the six days as random effects) are reported below.
A significant effect of visibility on speed at the $p < .05$ level was found [$F(3, 26.869) = 217.599$, $p < 0.001$]. Furthermore, a Bonferroni post-hoc test showed that differences between the speed in each visibility class and the speed in all other classes are statistically significant.
Those results confirm that the chosen clustering of cross-sections into visibility classes is consistent and that road geometric characteristics have a strong impact on speed.
Nevertheless, there is no statistically significant interaction between visibility and days of testing on speed, [$F(15, 258.032) = 1.111$, $p = 0.346$]. So, even if globally speed is affected by days of testing and visibility, the way in which drivers modify their speed over days with the acquired route familiarity seems to be not influenced by the different visibility classes.
Instead, there is a statistically significant interaction between drivers and visibility on speed, [$F(54, 258.012) = 1.924$, $p < 0.001$] and a statistically significant interaction between drivers, days of testing and visibility on speed, [$F(258, 27352) = 4.631$, $p < 0.001$].
Results showed that, on average, as expected, speed increases with the sight distance. At the same time, speed increases over days (see 4.1.1), but it happens independently from the visibility class (no interaction was found between visibility and days of testing). This means that, on average, acquiring familiarity with a route leads drivers to increase their speed in both higher and lower visibility conditions. This finding confirms what found in literature: route familiarity leads to less cautious behaviors [26], and it was obtained by a more naturalistic study than the previous ones based on driving simulators [26], [27]. In fact, as an example, in low visibility condition (cross-sections in which available sight distance is less than 100 m), in the fourth day of testing (when familiarity seems to be already acquired) mean speed (79.449 ± 14.252 km/h) is significantly higher compared to the

first day of testing (72.195 ± 11.768 km/h). Moreover, the increased speed is maintained over time in both the fifth and the sixth days of testing (see Table 3).
However, those phenomena were investigated at a more detailed level by considering that all drivers could not have homogeneous behaviors, as long as interactions between drivers, visibility and days of testing on speed were found (see also 4.1.1).

4.2 Users clustering

Cluster analysis with K ranging between 2 and 5 was performed by considering for each driver speed data belonging to all six days of testing and to all road cross-sections. The silhouette plots (showing silhouette values for each object belonging to the dataset) and the overall mean silhouette values are shown below (Fig. 9, Table 4).
Based on literature review [48], an acceptable value of the mean silhouette related to all objects belonging to the sample was found only for K = 3. Furthermore, as noted in Fig. 9, for K = 3, all silhouette values are positive and the majority of them are greater than 0.6. So, speed data were clustered according to this result. Six drivers were assigned to the cluster A, eight drivers to the cluster B and the last five drivers to the cluster C.
Therefore, speed/days diagrams were drawn for each drivers’ cluster (Fig. 10).
Hence, drivers were clustered into three groups (A, B, C) according to their speed in the six days of testing (considering all observations in the different days and cross-sections as a whole sample). This means that the cluster B is composed by the drivers who are consistent with the mean speed of the sample. Instead, the cluster A is composed by the drivers who are more cautious, on average, than the other drivers (speeds lower than the mean speed of the sample), and the cluster C is composed by the drivers who are more aggressive, on average, than the other drivers (speeds greater than the mean speed of the sample).
This classification will be useful in order to investigate the different evolution of speed behaviors over days for each drivers’ cluster corresponding to a different speed behavior.

4.3 Piecewise linear regression speed/days

As obtained from the analyses performed, the breakpoint was located at $t = 4$. So, the time domain was split into two sub-domains: $t \leq 4$ and $t > 4$.
Results of the tests conducted for $t \leq 4$ allow to reject the null hypothesis for each combination of visibility classes and drivers’ cluster (always showing: $p < 0.001$), that is, the speed and days are linearly related. Therefore, for this time interval, the model can be structured in each combination as: $E(y) = \alpha_1 + \beta_1 t$. The estimated values of α and β are reported in Table 5.
Results of the tests conducted for $t > 4$ allow to reject the null hypothesis for some combinations of visibility classes and drivers’ cluster: in all visibility classes for the drivers belonging to the cluster A (always showing: $p < 0.001$), in the medium visibility class for the drivers belonging to the cluster B ($p = 0.011$), in the medium-low and high visibility classes for the drivers belonging to the cluster C (always showing: $p < 0.001$). In the highlighted cases, considering $t > 4$, speed and days are linearly related and the model can be structured as: $E(y) = \alpha_2 + \beta_2 t$. In all the other cases, for the same time interval, the null hypothesis is accepted, that is, speed is constant over days. So, in those cases, the model can be structured as: $E(y) = \beta_2$. The estimated values of α and β are reported in Table 5.
Results showed that, for each combination of visibility class and drivers’ cluster, there is a linear relationship between speed and days for $t \leq 4$ (see Fig. 11). In general, in this time domain, speed increases over days. However, at this time, some considerations about the behavioral differences between drivers can be made.
Drivers belonging to cluster A (the “cautious” drivers of the sample) show a light speed increasing tendency characterized by an angular coefficient β included between 1.946 and 2.211 (see Table 5). Furthermore, the increasing tendency seems unaffected by the visibility class. On the other hand, drivers belonging to cluster C (the “aggressive” drivers of the sample) show a strong speed increasing tendency characterized by an angular coefficient β included between 3.897 and 5.743 (see Table 5).

Moreover, in this case, the visibility class has a clear influence on the increasing tendency: drivers are more prone to increase their speed in high visibility conditions (in which speed choice process is characterized by more degrees of freedom). Drivers belonging to cluster B (the “mean” drivers of the sample) show an intermediate speed increasing tendency characterized by an angular coefficient β included between 1.796 and 3.188 (see Table 5). Also in this case, visibility has such an influence on the increasing tendency, even if it is less important than for the cluster C.

Hence, the main finding is that all drivers react to the repetition of the driving tests with an increase in speed over days. However, this effect is more evident for aggressive drivers. In fact, not only they show greater average speeds, but they also show steeper speed increasing tendencies. For example, even in low visibility conditions, the β value for cluster C is about two times the β value for cluster A. This effect is maximum in high visibility conditions.

On the other hand, results showed that, in general, there is not a linear relationship between speed and days for $t > 4$ (see Fig. 11). However, when linearity is confirmed, the β value is very small. Since relationships between speed and days could be various and the regression for $t > 4$ was only based on data from the fifth and the sixth days of testing, results from the regression in this time domain can be used, jointly with speed data in the fourth day of testing, only to argue that the speed could be considered constant or variable after the fourth day of testing. Further studies could help in obtaining more detailed speed/time relationships based on different chronological measurements.

However, some considerations about the behavioral differences between drivers can be made also for $t > 4$, by considering regressions and differences between speed values.

Since for cluster B the speed is mainly constant over days in the considered time domain, only the differences between the two extreme behaviors have been considered.

In general, for cautious drivers (cluster A), the estimated speed value for the fifth day of testing is smaller compared to the fourth day of testing, while speed in the sixth day is similar to that value. Therefore, it seems that cautious drivers do not trust on their memory after the stimuli interruption of six days between the fourth and the fifth test and so they need a “route re-test”.

However, after this re-test, the route seems to be completely acquired and so, in the last day of testing (even if more distant in time) speed returns comparable with the speed in the fourth day of testing. This observed speed behavior is consistent with the possibility of a partial response recovery for cautious drivers. Instead, for drivers belonging to cluster C, in both the high and medium visibility classes, speed in the fifth day is similar to the speed in the fourth day, while speed in the sixth day is considerably smaller. On the other hand, for the other visibility conditions, speed is constant over days in the considered time domain. This could mean that aggressive drivers trust their memory in the lower visibility conditions (without needing a “re-test” as for the cautious drivers), while this not happens in the higher visibility conditions. In fact, probably, if the stimulus is not constantly repeated over time, aggressive drivers do not seem to be able to maintain their considerably high level of speed.

5. CONCLUSIONS

The on-road experiment carried out has provided important elements concerning the evolution of speed choice with the increased familiarity with a given route. In particular, in regard to the influence of acquired route familiarity, these results have been highlighted:

- on average, speed progressively increases in the early four days of testing and then settles on a constant value, even if the last days of testing are more distant in time than the others. This confirm the hypothesis that the speed choice can be influenced by the habituation process.
- Speed increases with visibility; but, on average, the increase in speed over days is present in both higher and lower visibility conditions.
- Dividing drivers into clusters based on their mean speeds allows to consider that there are significant behavioral differences among the sample of drivers. So, the increase in speed over days need to be analyzed at a more detailed level.
- Aggressive drivers show greater speed increasing rates than the cautious and the mean drivers in the first four days of testing; that is, speeds of aggressive familiar drivers are relatively

higher than speeds of cautious familiar drivers. This occurrence indicates that the habituation process is more evident for aggressive drivers.

- At a more detailed level, differences between visibility conditions can be related to the speed increasing rates for the aggressive and the mean drivers. In fact, in those cases, speed increasing rates are generally greater in higher visibility conditions than in the lower ones.
- Generally, the memory of drivers guarantees that speed does not decrease over days even if stimuli are not repeated with the same time interval. However, experimental data show that this long-term memory fails: (a) if the time interval is longer, for aggressive drivers in higher visibility conditions, (b) after the first pause in the stimuli repetition, for prudent drivers.

More in general, results from the on-road experiment show that route familiar drivers increase their speed in respect to their unfamiliar condition. This is consistent with similar findings from a driving simulator study [27]. The hypothesis advanced that the driving behavior could be affected by the habituation process can be confirmed. Drivers seem to get the habituation condition corresponding to the asymptotic response value which in turn coincides with a low demand condition in which attention capacity is reduced. This means that, on average, familiar drivers seem to go faster. Since this happens in both higher and lower visibility conditions, familiar drivers seem to be at the same time more unfocused on the driving task consistently with the results of a similar study [26]: an issue to consider in safety matters. Moreover, these results indicate that the study of the impact of the differences between familiar and unfamiliar drivers on road and traffic engineering ([30], [31]) could be deepened.

Hence, results shown stimulate experimental research in the same direction. However, for future research it will be better to overcome limitations of the study by employing a greater sample of users and by considering other variables such as acceleration, lateral positioning and/or the influence of other boundary conditions, in order to deepen knowledge on this still largely unexplored behavioral aspect.

REFERENCES

[1] American Association of State Highway and Transportation Officials (AASHTO). Highway safety manual (1st ed.), Washington, DC., 2010.

[2] Elvik, R, Vaa, T. The Handbook of Road Safety Measures. Elsevier Science, 2004.

[3] Kononov, J., Allery, B. Level of Service of Safety, Conceptual Blueprint and Analytical Framework. Transportation Research Record: Journal of the Transportation Research Board, Vol. 1840, Transportation Research Board of the National Academies, Washington, D.C., 2003, pp. 57–66.

[4] Taylor, D. H. Drivers' galvanic skin response and the risk of accident. Ergonomics, Vol. 7, 1964, pp. 439-451.

[5] Wilde, G. J. S. The theory of risk homeostasis: implications for safety and health. Risk Analysis, Vol. 2, 1982, pp. 209-225.

[6] Hedlund, J. Risky business: safety regulations, risk compensation, and individual behavior. Injury Prevention, Vol. 6, 2000, pp. 82-90.

[7] van der Horst, A. R. A. Behavioral adaptation to roadway engineering countermeasures. In: Rudin-Brown, C., Jamson, S. (eds.) Behavioral adaptation and road safety: theory, evidence and action. CRC Press, Boca Raton, 2013.

[8] O'Neill, B. A decision-theory model of danger compensation. Accident Analysis and Prevention, Vol. 9, 1977, pp. 157-165.

[9] Blomquist, G. A utility maximization model of driver traffic safety behavior. Accident Analysis and Prevention, Vol. 18, Issue 5, 1986, pp. 371-375.

[10] Tarko, A. P. Estimating Subjective Risk Revealed Through Driver Speed Choice. Paper presented at the Road Safety and Simulation International Conference RSS, Rome, 2007.

[11] Dulisse, B. Methodological issues in testing the hypothesis of risk compensation. Accident Analysis and Prevention, Vol. 29, Issue 3, 1997, pp. 285-289.

[12] Noland B. From theory to practice in road safety policy: Understanding risk versus mobility. Research in Transportation Economics, Vol. 43, 2013, pp. 71-84.

[13] Nilsson, G. Traffic safety dimensions and the power model to describe the effect of speed on safety (Doctoral dissertation, Lund University), 2004.

[14] Elvik, R. A re-parameterisation of the Power Model of the relationship between the speed of traffic and the number of accidents and accident victims. Accident Analysis & Prevention, 50, 2014, pp. 854-860.

[15] Fuller, R. Driver control theory: from task difficulty homeostasis to risk allostasis. In: Porter, B. E., Handbook of traffic psychology, Academic Press, Waltham, 2008.

[16] Näätänen, R., Summala, H. A model for the role of motivational factors in drivers' decision-making. Accident Analysis and Prevention, Vol. 6, 1974, pp. 243-261.

[17] Michon, J. A. Explanatory pitfalls and rule-based driver models. Accident Analysis and Prevention, Vol. 21, Issue 4, 1989, pp. 341-353.

[18] Vaa, T. Proposing a driver behavior model based on emotions and feelings: exploring the boundaries of perception and learning. In: Regan, M. A., Lee, J. D., Victor, T. W. (eds.). Advances in research and countermeasures, Vol. 1, Ashgate, Farnham, 2013.

[19] Tarko, A. P, Figueroa Medina, A. M. Implications of Risk Perception Under Assumption of Rational Driver Behavior, Transportation Research Record: Journal of the Transportation Research Board, 1980, Washington, D.C., 2006, pp. 8–15.

[20] Slovic, P., Fischhoff, B., Liechtenstein, S. Why study risk perception? Risk Analysis, Vol. 2, Issue 2, 1982, pp. 83-93.

[21] Slovic, P. Perception of risk. Science, Vol. 236, 1987, pp. 280-285.

[22] Hogarth, R. M. Beyond discrete biases: Functional and dysfunctional aspects of judgmental heuristics. Psychological Bulletin, Vol. 90, 1981, pp. 197-217.

[23] Sjöberg, L., Moen, B., Rundmo, T. Explaining risk perception. An evaluation of the psychometric paradigm in risk perception research, Rotunde, Vol. 84, Trondheim, 2004.

[24] Tversky, A., Kahneman, D. Judgment under uncertainty: Heuristics and biases. Science, 185, 1974, pp. 1124-1131.

[25] Elvik, R. A restatement of the case for speed limits. Transport Policy, Vol. 17, 2010.

[26] Yanko, M. R., Spalek, T. M. Route familiarity breeds inattention: A driving simulator study. Accident Analysis and Prevention, Vol. 57, 2013, pp. 80-86.

[27] Martens, M. H., Fox, M. R. J. Do familiarity and expectations change perception? Drivers' glances and response to changes. Transportation Research Part F: Traffic Psychology and Behavior, Vol. 10, 2007, pp. 476–492.

[28] Young, M., Stanton, N. Malleable attentional resources theory: a new explanation for the effects of mental underload on performance. Human Factors, Vol. 44, Issue 3, 2002, pp. 365–375.

[29] Rosenbloom, T., Perlman, A., Shahar, A. Women drivers' behavior in well-known versus less familiar locations. Journal of Safety Research, vol. 38, Issue 3, 2007, pp. 283-288.

[30] Highway Capacity Manual. Transportation Research Board, Washington, D.C., 2000.

[31] Milliken, J. G., Council, F. M., Gainer, T. W., Garber, N. J., Gebbie, K. M., Hall, J. W., et al. Managing speed: review of current practice for setting and enforcing speed limits. Transportation Research Board, Special Report 254. Washington DC, 1998.

[32] Blatt, J., Furman, S. M., Residence location of drivers involved in fatal crashes. Accidents analysis and prevention, Vol. 30, No. 6, 1998, pp. 705-711.

[33] Groves, P. M., Thompson, R. F. Habituation: A dual-process theory. Psychological Review, Vol. 77, Issue 5, 1970, pp. 419-450.

[34] Rankin, C. H., Abrams, T., Barry, R. J., Bhatnagar, S., Clayton, D. F., Colombo, J., Coppola, G., Geyer, M. A., Glanzman, D. L., Marsland, S., McSweeney, F. K., Wilson, D. A., Wu C., Thompson, R. F. Habituation revisited: an updated and revised description of the behavioral characteristics of habituation. Neurobiology of learning and memory, 92(2), 2009, pp. 135-138.

[35] Rasmussen, J. Information Processing and Human-Machine Interaction. An Approach to Cognitive Engineering, 1986, Elsevier, New York.[36] Colonna, P., Berloco, N. External and internal risk of the user in road safety and the necessity for a control process. XXIV PIARC world road congress, Mexico City, 2011.

[37] Mahalel, D., Szternfeld, Z. Safety improvements and driver perception. Accident Analysis and Prevention, Vol. 18, Issue 1, 1986, pp. 37-42.

[38] Colonna, P., Aquilino A. R., Berloco, N., Intini, P., Ranieri, V. The influence of memory on road user behavior. Paper accepted for poster presentation at the 93rd Annual Meeting of the Transportation Research Board. Washington, DC, 2014.

[39] Colonna, P., Berloco, N., Intini, P., Ranieri, V. Route familiarity in road safety: speed choice and risk perception based on a on-road study. 94th Annual Meeting of the Transportation Research Board Compendium of papers. Washington, DC, 2015.

[40] Colonna, P., Aquilino, A., Berloco, N., Ranieri, V. Relationships between road geometry, drivers' risk perception and speed choice: an experimental study. 92nd Annual Meeting of the Transportation Research Board Compendium of Papers. Washington, DC, 2013.

[41] Godley, S. T., Triggs, T. J., Fildes, B. N. Driving simulator validation for speed research. Accident Analysis and Prevention, Vol. 34, Issue 5, 2002, pp. 589-600.

[42] World Road Association (PIARC). Road Safety Manual (1st ed.), Paris, 2004.

[43] Fitzpatrick, K., Elefteriadou, D. W., Harwood, J. M., Collins, J. M., McFadden, J., Anderson, I., Krammes, R. A., Irizarry, N., Parma, K. D., Bauer, K. M., Passetti, K. Speed Prediction for Two-Lane Rural Highways. Report FHWA-RD-99-171. FHWA. U.S. Department of Transportation, Washington DC, 1999.

[44] Guidelines for the Design of Road Infrastructures. Italian Ministry of Infrastructures and Transports. D. M. n. 6792, 5/11/2001. Rome, 2001.

[45] Fambro, D. B., Fitzpatrick, K., Koppa, R. J. Determination of Stopping Sight Distances. Transportation Research Board, Report 400. Washington DC, 1997.

[46] Lamm, R., Psarianos, B., Mailaender, T. Highway Design and Traffic Safety Engineering Handbook. Mc Graw-Hill Professional Publishing, New York City, 1999.

[47] Razali, N., M, Wah, Y., B. Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests. Journal of Statistical Modeling and Analytics 2.1, 2011, pp. 21-33.

[48] Everitt, B. S., Landau S., Leese, M., Stahl, D. Cluster Analysis, 5th Edition. Wiley and Sons Ltd., Hoboken, 2011.

[49] Galante, F., Mauriello, F., Montella, A., Pernetti, M., Aria, M., D'Ambrosio, A. Traffic calming along rural highways crossing small urban communities: Driving simulator experiment. Accident Analysis and Prevention, Vol. 42, 2010, pp. 1585-1594.

[50] Montella A., Pariota, L., Galante, F., Imbriani, L. L., Mauriello, F. Prediction of drivers' speed behavior on rural motorways based on an instrumented vehicle study. Transportation Research Board. Journal of the Transportation Research Board, n. 2434, Transportation Research Board of the National Academies. Washington DC, 2014, pp. 52-62.

[51] White, P., Webber, D. J., Helvin, A. Modelling structural change using broken sticks. Paper provided by University of the West of England, Department of Economics in its series Discussion Papers with number 801, 2008.

[52] Montella, A., Aria, M., D'Ambrosio, A., Galante, F., Mauriello, F., Pernetti, M. Simulator evaluation of drivers' speed, deceleration and lateral position at rural intersections in relation to different perceptual cues. Accident Analysis and Prevention, Vol. 43, 2011, pp. 2072-2084.

[53] White, P., Webber, D. J., Helvin, A. Testing for a structural change of gradient in regression modelling. Empirical Economics Letters, Vol. 9 (8), 2010, pp. 819-826.

CAPTIONS TO FIGURES

FIGURE 1. Habituation process (blue), sensitization process (red) and dishabituation effect (orange), based on [33].

FIGURE 2. Example of driving test schedule for a user belonging to the sample.

FIGURE 3. Layout of the test driving routes.

FIGURE 4. An example of speed profile.

FIGURE 5. Horizontal alignments and elevation profiles of the stretch 1 and the stretch 2.

FIGURE 6. Example of sight distance profile.

FIGURE 7. Boxplots of speeds in the six days of testing (legend to the boxplot on the right).

FIGURE 8. Speed/days diagrams for each visibility class.

FIGURE 9. Silhouette plots related to the drivers' clustering.

FIGURE 10. Speed/days diagram for each drivers' cluster.

FIGURE 11. Piecewise linear regressions speed/days for drivers' cluster A and C.

TABLE 1. Means, standard deviations of speed and speed percentage differences for each day of testing.

	Day 1 (Test 1)	Day 2 (Test 2)	Day 3 (Test 3)	Day 4 (Test 4)	Day 10 (Test 5)	Day 27 (Test 6)
Mean speed (km/h)	79.068	83.802	87.205	88.802	88.442	89.515
Standard deviation (km/h)	12.649	15.459	16.195	15.845	16.832	14.735
Mean speed percentage difference[a]	-	0,060	0,041	0,018	-0,004	0,012

[a]Mean speed percentage differences are computed as: (Speed, day i+1 – Speed, day i)/(Speed, day i).

TABLE 2. Results of the ANOVA test and Bonferroni post-hoc tests (*p*-values)[a].

Days	Day 1 (Test 1)	Day 2 (Test 2)	Day 3 (Test 3)	Day 4 (Test 4)	Day 10 (Test 5)	Day 27 (Test 6)
Day1	1	**< 0.001**	**< 0.001**	**< 0.001**	**< 0.001**	**< 0.001**
Day2		1	**< 0.001**	**< 0.001**	**< 0.001**	**< 0.001**
Day3			1	**< 0.001**	**< 0.001**	**< 0.001**
Day4				1	1	<u>0.083</u>
Day5					1	**0.001**
Day6						1
ANOVA test						
F-statistic = 14.939						
p-value < 0.001						

[a]Boldface indicates statistically significant values with 5 % level of significance. Underlined indicates values with 10 % level of significance.

TABLE 3. Means and standard deviations of speed for each day of testing and each visibility class.

	Visibility class	Days of testing						
		Day 1 (Test 1)	**Day 2 (Test 2)**	**Day 3 (Test 3)**	**Day 4 (Test 4)**	**Day 10 (Test 5)**	**Day 27 (Test 6)**	**Mean**
Mean speed (km/h)	**Low**	72.195	74.996	77.841	79.449	79.673	81.742	77.668
	Medium-low	77.537	81.308	85.079	87.347	86.666	88.006	84.365
	Medium	79.558	86.366	89.647	89.624	90.229	90.253	87.650
	High	84.932	90.996	94.427	95.829	95.145	95.858	92.893
St. Dev. (km/h)	**Low**	11.768	13.601	14.871	14.252	15.832	13.985	14.460
	Medium-low	11.406	13.987	14.477	14.021	15.600	13.812	14.431
	Medium	12.830	15.959	16.702	16.198	16.755	14.947	16.093
	High	12.075	14.594	15.412	15.629	16.218	13.566	15.171

TABLE 4. Silhouette overall mean values for each number of groups K attempted.

	K = 2	**K = 3**	**K = 4**	**K = 5**
Silhouette mean values[a]	0.56	**0.60**	0.58	0.48

[a]Boldface indicates silhouette mean values considered as acceptable [48].

TABLE 5. Values of the intercept α and the slope coefficient β resulting from the piecewise linear regressions for each combination of visibility class and drivers' cluster.

t ≤ 4									
		VISIBILITY							
		LOW		**MEDIUM-LOW**		**MEDIUM**		**HIGH**	
		α	**β**	**α**	**β**	**α**	**β**	**α**	**β**
DRIVERS' CLUSTER	**A**	62.048	1.946	68.111	2.132	71.091	2.211	77.651	2.071
	B	72.796	1.796	76.855	2.686	79.911	2.567	83.068	3.188
	C	75.577	3.897	79.132	5,454	83.925	5.558	94.138	5.743
t > 4									
		VISIBILITY							
		LOW		**MEDIUM-LOW**		**MEDIUM**		**HIGH**	
		α	**β**	**α**	**β**	**α**	**β**	**α**	**β**
DRIVERS' CLUSTER	**A**	64.833	0.312	73.284	0.229	75.077	0.285	81.968	0.288
	B	81.781		87.085		92.583	-0.144	93.693	
	C	91.481		100.005		110.828	-0.397	115.618	-0.369

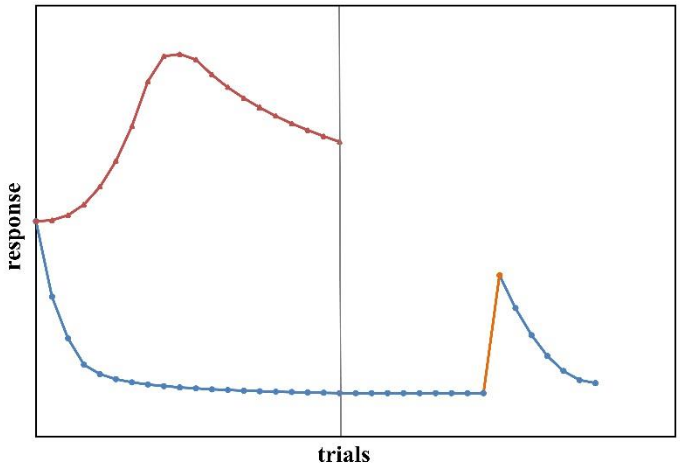


Figure 1

MAY							
week	M	T	W	T	F	S	S
18						1	2
19	3	4	5	6	7	8	9
20	10	11	12	13	14	15	16
21	17	18	19	20	21	22	23
22	24	25	26	27	28	29	30
23	31						

Figure 2

Figure 3

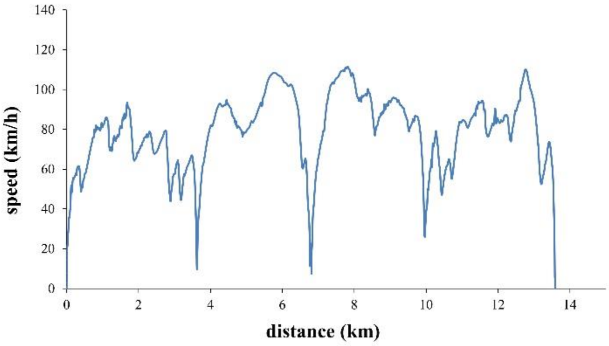
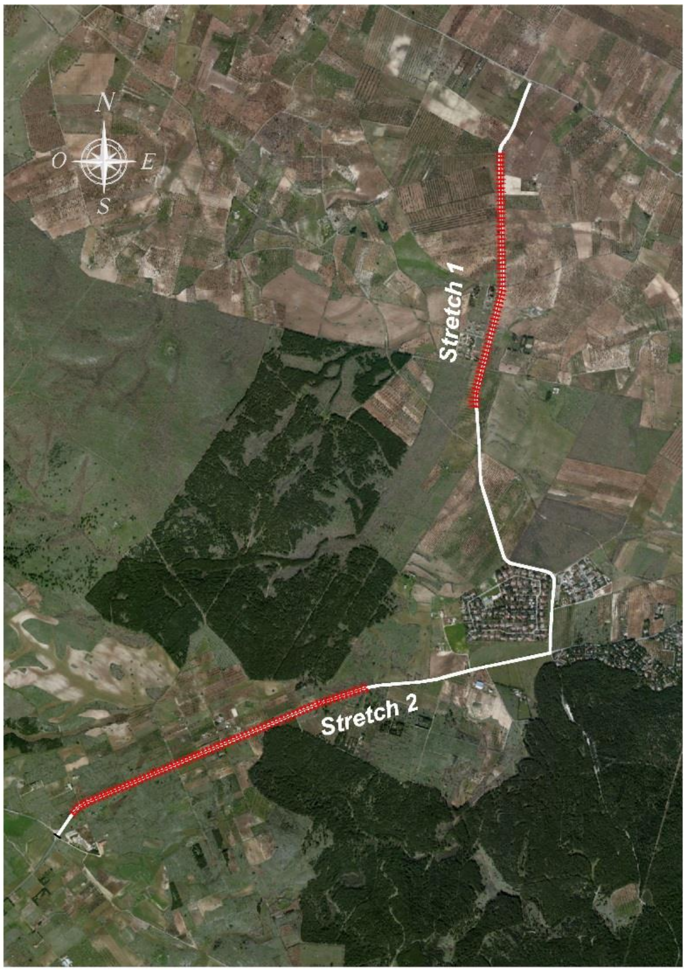


Figure 4



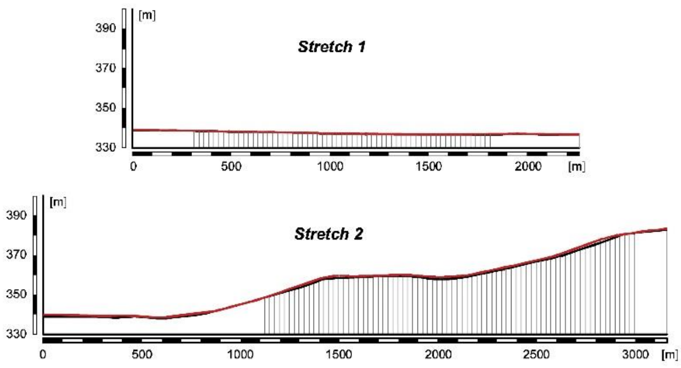


Figure 5

Figure 6

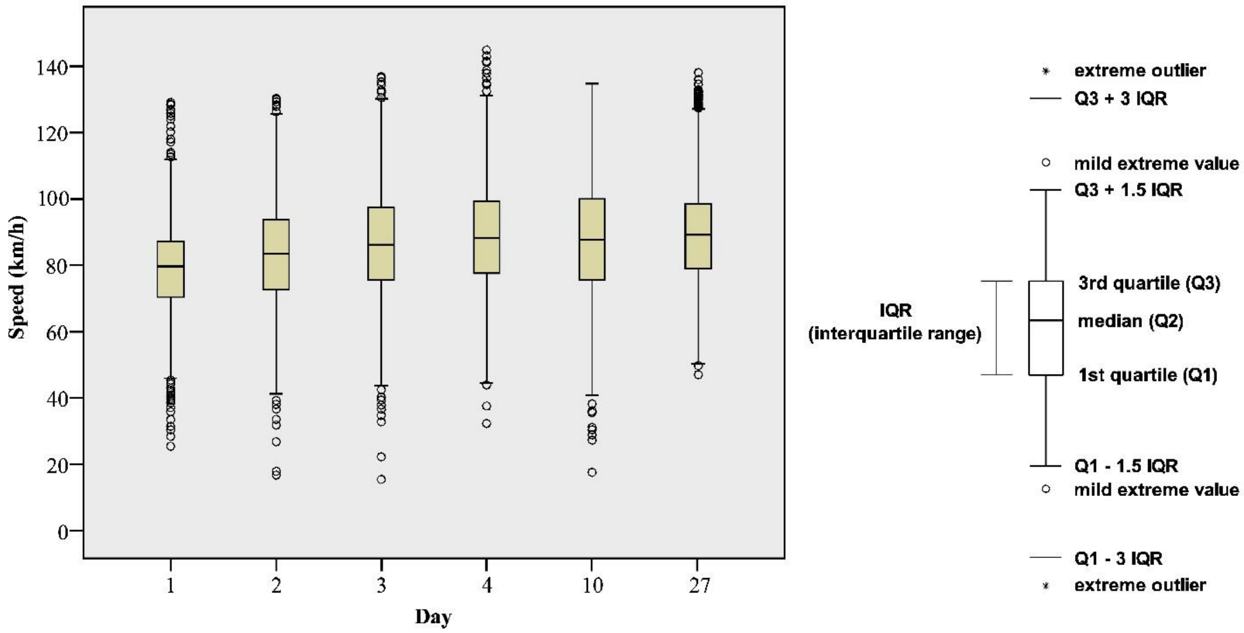


Figure 7

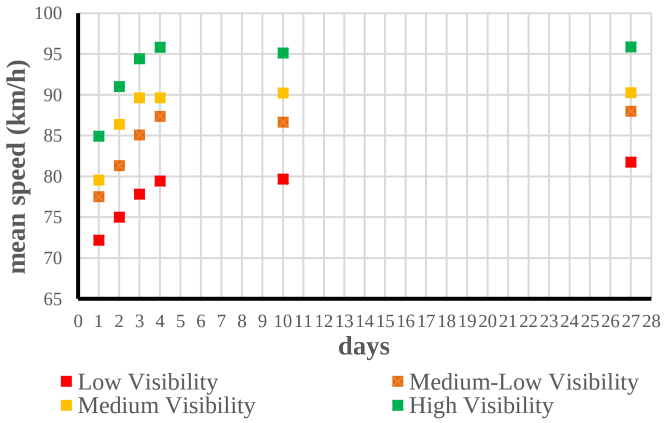


Figure 8

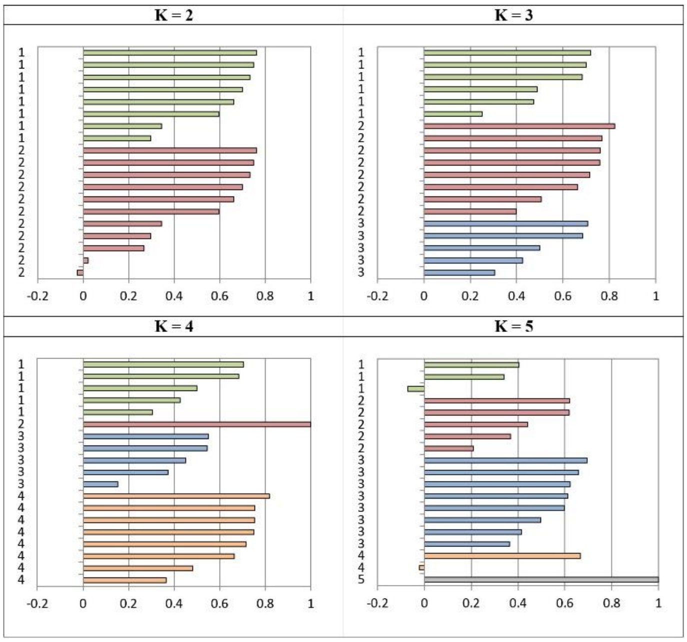


Figure 9

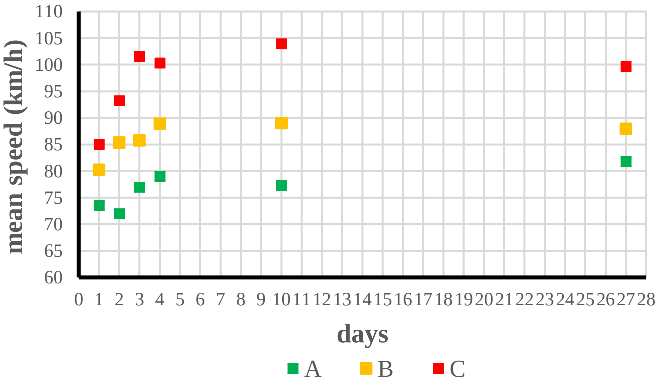


Figure 10

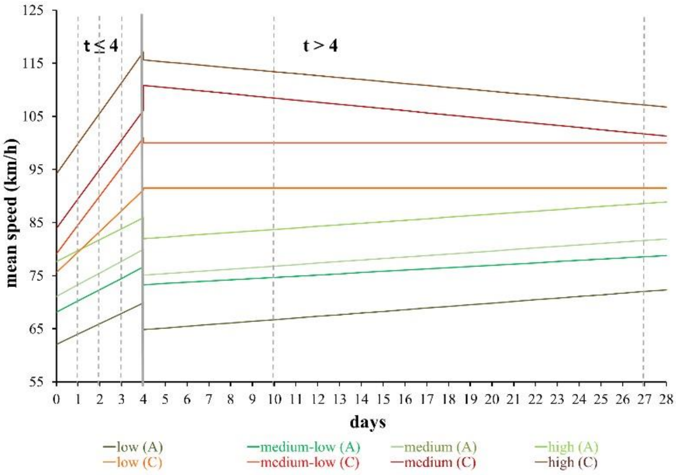


Figure 11