Assessing driving behaviour using acceleration thresholds

David Stefan

Assessing driving behaviour using acceleration thresholds

By

D.A. Stefan

in fulfilment of the requirements for the degree of

Master of Science

in Mechanical Engineering

at the Delft University of Technology,

1 November 2022.

Cover image: Transfăgărășan, Romania (Viator, 2016).

Supervisors:	Dr. D.D. Heikoop	CBR
	Ir. T. Driessen	TU Delft
	Dr. Ir. J.C.F. de Winter	TU Delft
	Dr. D. Dodou	TU Delft

Abstract

Different types of sensors that monitor the driver, the vehicle, and the surroundings are increasingly being implemented in vehicles. These developments are relevant to formal driving test organizations for the use of sensors in the assessment of driving behaviour. However, no guidelines exist about the exact use of sensor-generated data for driving assessment purposes, such as the driving exam. Relatively low-cost and easily implementable sensors such as accelerometers (g-force sensors) could be used to assess driving behaviour during the driving exam to distinguish between desired and undesired driving behaviour. Earlier studies have already investigated the use of accelerometers to distinguish between driving styles using threshold values, but do not agree on the most optimal threshold to do this.

In this study, thresholds were created based on scripted driving exams driven by professional driving examiners, who portrayed driving styles commonly observed at the driving exam. These rides were driven over a period of three weeks at the Dutch driving license organisation (CBR) in Leusden, The Netherlands, where scripted exam rides are part of the training for new examiners. Using the accelerometer of an iPhone X, the accelerations generated during 21 rides were measured supported by a dashcam recording the road ahead. The experimenter drove along all rides and registered driving events, including turns, speed increases/decreases and lane changes to relate the measured accelerations to these specific events. The rides included in the dataset were divided into four different driving styles: aggressive, desired, overcautious and negligent, from which the first three were analysed due to their direct link with accelerations. The obtained acceleration data was analysed using 63 initial features which describe the amplitude of the acceleration signal through the various events, speed zones and acceleration directions.

It was found that accelerations differ significantly between aggressive and desired driving styles, whilst the difference between overcautious and desired driving is less apparent. The thresholds determined in this study were able to classify the created dataset between desired and undesired driving behaviour with accuracies varying from 50% to 83% ($M = 69.1, SD = 11.4$).

The results show that using acceleration data obtained during the driving exam can aid the examiner during the driving exam by giving insight into how well a candidate performed in terms of smooth driving through traffic. Concomitantly, a recent report from the Dutch government (Roemer, 2020) addressed that the CBR should investigate the possibilities for using sensor-collected driving data to give more insight to a driving exam candidate as to how to improve his/her driving behaviour after failing the driving exam, giving further rise to researching the use of driving data for practical use.

Acknowledgements

This thesis concludes my master's degree in mechanical engineering at Delft University of Technology. Writing a master's thesis is difficult and necessitates a great deal of self-discipline, perseverance, and resilience. I would like to take this opportunity to thank those individuals without whose assistance I would not have been able to complete this thesis, and my degree in mechanical engineering in general

First and foremost, I would like to thank Daniël, my daily supervisor. You showed genuine interest in the project from the start and provided me with critical feedback and new insights throughout. Furthermore, when we worked at the CBR office in Rijswijk, you were always available for discussions, which led to some of the most significant breakthroughs during the project. I would also like to thank the people from the CBR examiner training centre in Leusden. Driving along their training sessions was not only unique and interesting for me as an engineer, but it was also critical to the project's success.

Next, I would like to express my gratitude to Tom Driessen, Dr. Joost de Winter, and Dr. Dimitra Dodou. Throughout the project, you always allowed me to look at every detail from a scientific and academic standpoint. You were always quickly (online) available for a short discussion or even an hour-long meeting to discuss project concerns. You gave me the opportunity to truly learn and understand various aspects that come into play when writing a scientific piece of literature.

And of course, I could not have done this without the eternal support of my family and friends. Especially my girlfriend Nienke, who always supported me during these sometimes stressful periods. Last but not least, I am grateful to my friend Stan, who gave me discipline and strength throughout all my years of study.

David Stefan

The Hague, September 2022

Table of contents

1 Introduction

Enhancing driving performance and automated driving has been the subject of extensive research over the past years, covering topics such as in-vehicle sensors (Marti et al., 2019; Schoettle, 2017), computer vision in traffic (Ranft & Stiller, 2016; Rangesh & Trivedi, 2019), car path planning (Gonzalez et al., 2015; Marin-Plaza et al., 2018), and vehicle control (Farag, 2020; Lima et al., 2018). The realisation that fully autonomous driving might not be achieved within the next three to five decades is also becoming more and more apparent (Shladover, 2016; Tabone et al., 2021; Wei et al., 2021). Even the most advanced prototypes of automated vehicles require human intervention or behave in unexpected ways (Goodall, 2021; Boggs et al., 2020), suggesting that drivers will still need to be trained and licensed for driving purposes in the coming decades.

With the growing number of vehicular monitoring systems, the opportunity arises to use the earlier mentioned systems to aid in the assessment of driving behaviour. The current literature already contains a plethora of promising results in using intelligent systems to inspect driving behaviour and increase driver safety, for instance, by using driver-facing cameras to detect drowsiness and distraction (Chowdhury et al., 2018; Zhang et al., 2021) and more general unsafe driving behaviours (Figueredo et al., 2019; Hickman & Hanowski, 2011). In-vehicle data recorders (Shimshoni et al., 2015) and smartphones are commonly used to gather the data which is used to assess driving behaviour (Kalra et al., 2021; Zhang et al., 2019). Nowadays, some driving insurance companies monitor features such as speeding, hard braking, or similar undesired events to reward people for safe driving behaviour with reduced insurance fees (Arumugam & Bhargavi, 2019; Handel et al., 2013). In most of these cases, acceleration and speed measurements are used, as these measurements require relatively low-end/low-cost equipment and can, therefore, be implemented easily (Zhang et al., 2019). Sensors specifically used for these types of measurements are Inertial Measurement Units (IMU; Engelbrecht et al., 2014), accelerometers (Miyajima et al., 2007), or sensors of on-board diagnostics (OBD) systems (Malik & Nandal, 2021).

Driving licensing organisations may profit from the previously described vehicular monitoring systems in assessing driving behaviour, as their decision about desired and undesired driving is based on the human-intensive interpretation of driving behaviour conveyed by a driving examiner. Moreover, these organisations are sometimes challenged by candidates in terms of the examiner's verdict's reliability and validity, such as test-retest reliability (i.e., when the driving exam is performed twice with different examiners) and inter-examiner reliability (i.e., where two examiners take place in the car during the test). Although there are only a few studies in this area, those that exist show a test-retest reliability (pass/fail congruence of 64% in Baughan & Simpson, 1999, and 63% in Alger & Sundstrom, 2013), as well as inter-examiner reliability of the same test (72% in Bjørnskau, 2003, and 93% in Alger & Sundström, 2013). These studies show that there is still room for improvement regarding the validity and reliability of the driving exam.

Until now, the use of data such as in-vehicle accelerations in the driving exam is very limited in both the research field and practice. One of the challenges with the use of data for assessing driving behaviour, in the driving exam, and in general, is the absence of a data standard that sets guidelines on how to use the acceleration data to distinguish between desired and undesired driving. In fact, there are multiple implementation options for data processing within driving behaviour models (e.g. anomaly detection, machine learning). In addition, studies that investigated driving behaviour mostly have their own subjective interpretation in terms of desired driving behaviour (Osafune et al., 2017; Ylizaliturii-Salcedo et al., 2015), which they base on their perception of desired driving behaviour. These perceptions are often denoted in thresholds, which are used throughout different models for driving behaviour and create a boundary between two driving styles (i.e. discriminating between aggressive and desired driving behaviour). Other studies rely mostly on thresholds set by earlier works (e.g., Girbes et al., 2019). It has been found that overall, studies adopting a threshold for specific acceleration events adhere to thresholds ranging from 0.1 g to 0.5 g (see Table 1).

Table 1.

Acceleration thresholds for various events concerning the assessment of driving behaviour.

Threshold	Aggressive acceleration	Aggressive deceleration	Aggressive turn
0.1 g	Paefgen et al. (2012)	Paefgen et al. (2012)	Baldwin et al. (2004)
0.2 g	Baldwin et al. (2004)	Malik and Nandal, (2021); Osafune et al. (2017)	
0.3 g	Osafune et al. (2017); Ylizaliturri-Salcedo et al. (2015); Chigurupati et al. (2012)	Chigurupati et al. (2012)	Chigurupati et al. (2012); Miyajima et al. (2007)
0.4 g		Zeeman and Booysen (2013); Bergasa et al. (2014)	Dai et al. (2010); Ylizaliturri-Salcedo et al. (2015)
0.5 g	Fazeen et al. (2012)	Fazeen et al. (2012); Miyajima et al. (2007)	

Table 1 shows disagreement between studies in terms of the threshold value for when a person drives aggressively in terms of accelerations. This is in contrast to speed, where the speed limit on any road could always be used as threshold for classifying between desired and undesired driving behaviour, where the use of this threshold can vary from applying it as a strict upper bound (e.g. exceeding the speed limit means failing the driving exam) or using it via a severity score (e.g. calculating the deviation from the speed limit during the whole driving exam).

In order to determine an objective threshold for acceleration data, establishing a threshold based on professional judgement, such as that of driving examiners, could be a way forward. Driving examiners assess driving behaviour based on their experience during the exam, including the perceived accelerations. Moreover, driving examiners are trained intensively including a training course where only 10% of applicants actually pass all the requirements before they are eligible to assess driving behaviour. In this study, thresholds for desired accelerations in cars will be determined with the help of driving examiners from the Dutch driving exam organisation (CBR). The main research question to be answered in this study is:

"*How can in-vehicle acceleration measurements indicate (un)desired driving behaviour during the driving exam?"*

The following sub-questions will be answered:

1. What acceleration features can be used to determine differences between desired and undesired driving behaviour?
2. What acceleration threshold values measured during the exam can be used to determine differences between desired and undesired driving behaviour?
3. What accuracy can be achieved by classifying between desired and undesired driving using accelerations?

This study focused on the identification of thresholds for perceived acceleration (or g-forces) during driving. A dataset was created containing scripted driving exam rides which were used to analyse the accelerations perceived within the car. All rides were driven with specific driving behaviours through written scenarios, which will be used to compare the acceleration amplitude between different driving styles. The difference in acceleration amplitude between the driving styles was determined, and an acceleration threshold was defined.

2 Method

2.1 Creating the dataset

The data were collected in cooperation with the driving examiner training centre of the CBR. At this training facility, located in Leusden, The Netherlands, driving examiner trainees are trained to become licensed driving examiners. A part of their training consists of on-road training sessions, in which qualified coaches mimic exam candidates and driving styles commonly found (e.g., the nervous candidate, the sportive candidate) at the driving exam. The trainee takes the role of the examiner and is expected to make a pass or fail verdict based on the acted driving style.

The acceleration measurements were conducted during 21 of these training sessions from 28 March 2022 to 16 April 2022. All coaches involved were asked for consent before the start of the experiment, see Appendix A. The rides themselves all mimicked a standard driving exam conducted by the CBR, having a duration of 30 to 35 minutes and including the seven exam parts of the *Rijprocedure B* (desired driving procedure; CBR, 2020) listed in Appendix B.

2.1.1 Scenarios

During each of these 21 rides, the coach behind the steering wheel acted out a prescribed scenario. For each scenario, the driving behaviour acted out during the ride is defined with respect to the seven exam parts listed in the *Rijprocedure B* (CBR, 2020; i.e., road area, traffic situation or driving manoeuvre). In Appendix C, the scenarios driven during the 21 rides are listed.

For analysis purposes, the 21 scenarios driven have been categorized into four distinct driving style categories: aggressive driving behaviour (AGG), desired driving behaviour (DES), overcautious driving behaviour (CAU) and negligent driving behaviour (NGL). From these four driving style categories, only Category DES contains rides that should result in passing the driving exam. A detailed description of these four driving styles is listed in Appendix D. The four driving style categories were created as follows: After each ride, the coach driving and the experimenter together performed the categorization, which was based on the traits listed per scenario (i.e., taking speed bumps too harsh, being too careful at cross-sections) and the subjective interpretation of the scenario by the experimenter and the coach acting out the driving style. In order to validate the categorization, two colleague researchers were given the written descriptions of the 21 scenarios and asked to classify them into the same four driving style categories, resulting in an inter-rater agreement of κ = 76%, indicating a substantial agreement (McHugh, 2012).

2.1.2 Experimental setup

In total, four coaches and four cars were used during the experiment: a 2015 Skoda Octavia, a 2014 Peugeot 308 SW, a 2016 Seat Ateca and a 2015 Volkswagen T-Roc. During each of the 21 rides, three persons were present in the car: the coach acting out the scenario was driving the car, the examiner trainee was in the passenger seat, and the experimenter was sitting in the backseat, behind the examiner trainee (see Figure 1).

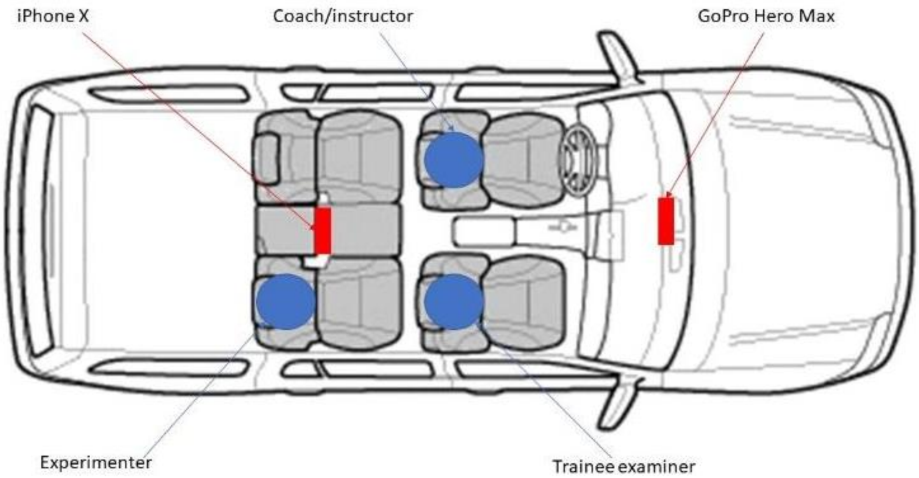

Figure 1. Measurement system (red) and person (blue) placement within the car during the experiment.

2.1.3 Measurements

During all 21 rides, the built-in accelerometer and GPS of an iPhone X were used to measure three-axis accelerations and position, respectively. The iPhone was mounted on the backseat of the car, aligned with the orientation frame of the car. Any misalignment between the orientation of the iPhone and the car was corrected using an alignment function as described in section 2.1.4. Using MATLAB mobile (version 9.1.2), the accelerations and GPS coordinates during each ride were logged to separate .mat files. The acceleration and location signals were sampled at 10 Hz and 1 Hz, respectively. In addition to the acceleration and location measurements, a GoPro Hero Max was used to film the road ahead during the rides. The video data from the GoPro was used for illustration purposes and to facilitate the analysis of anomalies from the accelerometer (i.e., excessive spikes in the acceleration signal).

The experimenter also used an open-source event-recorder app to manually register four types of events:

- Turns; driving over a curved road with an angle between about 60 and 120 degrees.
- Speed increase; noticeably increasing the speed of the vehicle by pressing the gas pedal; subjectively determined by experimenter.
- Speed decrease; noticeably decreasing the speed of the vehicle by pressing the brake pedal; subjectively determined by experimenter.
- Lane changes; switching lanes. Also includes exiting and merging the highway.

Note that speed increases and decreases were registered on all road types, including turns, curvy roads and straight roads. In addition to events, the maximum speed allowed in the zone where the event was driven was also registered (i.e. 30 km/h, 50 km/h, 80 km/h and 100 km/h). The event registration was based on the experimenter's own interpretation of the traffic situation from the backseat of the vehicle. In Figure 2, a snapshot from the GoPro video is shown together with the respective location, corresponding to the exact moment of the (turn) event registration by the experimenter. In Appendix E, similar snapshots are provided for other types of events.

Figure 2. Snapshot of an event registration (turn in 30 km/h speed zone). The red dot marks the GPS location where the event was registered. The driving direction is given by the red arrows.

2.1.4 Acceleration signal pre-processing

The acceleration signal was pre-processed by (1) removing the noise from the acceleration signal and (2) aligning any offset between the iPhone's orientation frame and the orientation frame of the vehicle (see Figure 3).

To remove the noise from the acceleration signal, a low pass filter with a cut-off frequency of 0.66 Hz was utilized. By determining the frequency of the acceleration signal whilst the vehicle was stationary (i.e., running engine in front of a traffic light), the frequency of the noise was determined at 0.66 Hz and higher, and henceforth eliminated using the low pass filter.

In order to ensure the accelerations of the phone were in line with the orientation frame of the vehicle, the offset between the two orientation frames was calculated and removed using an alignment function. For all three acceleration directions (lateral (x), longitudinal (y), and gravitational (z)), the equilibrium position was placed at 0 g. The offset from the equilibrium position was calculated by taking the mean of the acceleration signal and subsequently subtracting the mean of every acceleration data point.

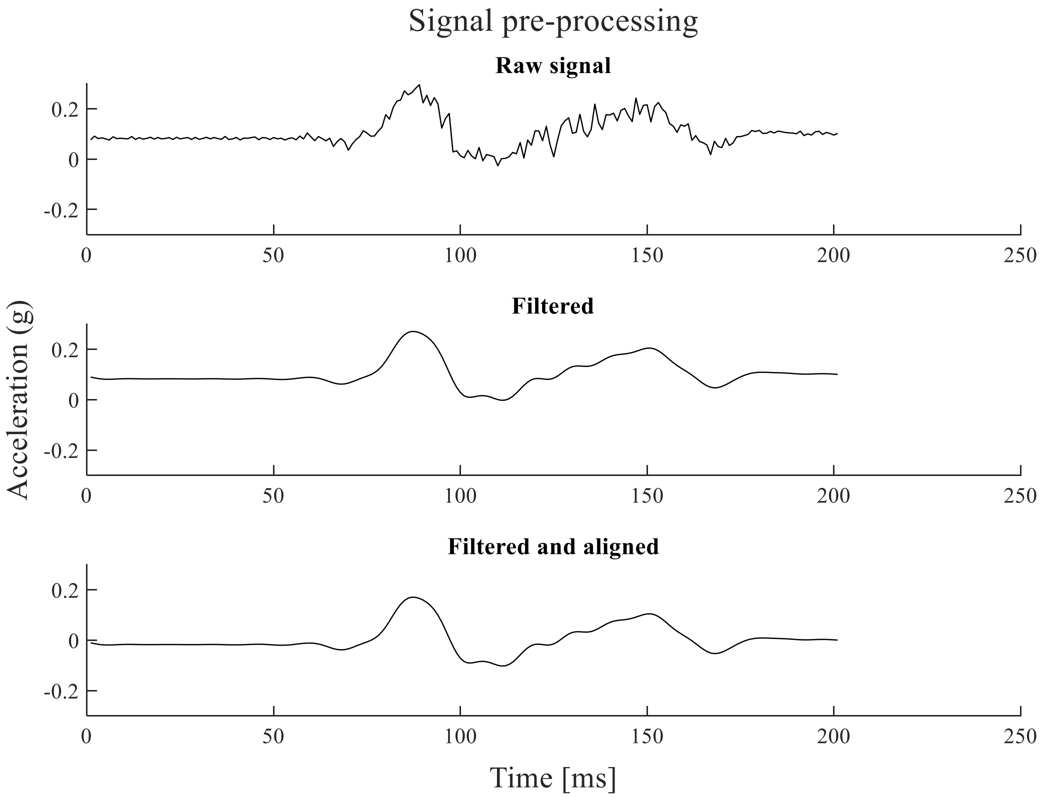

Figure 3. Acceleration pre-processing. The top plot shows a signal which is misaligned and contains noise. In the middle plot, the noise has been removed, and in the lower plot, the alignment is done.

2.2 Data analysis

Together, 63 features (e.g., mean acceleration amplitude through turns in 30 km/h speed zone) were estimated from the acceleration measurements included in the dataset. Specifically, in the full-ride analysis, three features were analysed: full-ride mean amplitude in lateral, longitudinal, and gravitational directions. In the event analysis, acceleration was analysed for four event types (i.e., turns, speed increases, speed decreases, and lane changes), five speed zones (i.e., 30 km/h, 50 km/h, 80 km/h, 100 km/h, and all speed zones combined), and three acceleration directions (i.e., lateral, longitudinal, and gravitational), totalling 60 features. Note that some features are dependent on each other: for example, the event registered in all speed zones includes the events registered in the 30 km/h, 50 km/h, 80 km/h and 100 km/h speed zones.

The rides categorized as NGL contain traits in the driving behaviour such as improper looking behaviour (i.e. overviewing traffic situations) and road misplacement (i.e. improper positioning on cross-sections). Because these traits are not directly associated with acceleration behaviour, as also stated by the coach acting out the respective scenario (see Appendix D), the NGL rides were excluded from further analysis. Eventually, of the 21 rides, 10 were analysed since these make up the correctly measured rides measured in styles AGG (n = 3), DES (n = 5), and CAU (n = 2), see Figure 5.

2.2.1 Full-ride analysis

The first type of analysis looked into the acceleration data over the entirety of a single ride. The mean, standard deviation, and maximum and minimum of the acceleration amplitude signal in all three directions were calculated for each ride. Each ride was treated as an independent sample in its corresponding driving style (AGG, DES, or CAU).

2.2.2 Event analysis

Lateral, longitudinal, and gravitational acceleration during each of the four event types (i.e., turns, speed increases/decreases, and lane changes) were compared between the three driving styles. For each of the four events, first, an analysis was performed across speed zones, followed by an analysis per speed zone. The latter analysis was only conducted when the sample size for an event/feature in a specific speed zone was equal to or higher than five, to avoid overfitting (see Figure 5).

The main variable used in this analysis was the amplitude of the acceleration signal during an event. The duration of a driving event and the moment of registering an event varied. In order to minimize the effect of these two factors, the highest amplitude of the ten amplitudes before and after the event was taken (i.e., the maximum value of 20 acceleration amplitudes). Even though a number of these events were part of the same ride and driven by the same coach, each registered event was treated as an independent sample because the environment (i.e. location, speed, traffic) differed per event.

A Kolmogorov-Smirnov (KS) test revealed that the data used to compare events between driving style categories was not normally distributed, as shown in Appendix G. As a result, a non-parametric Kruskal-Wallis test was used. Effect size Cohen's *d* is also reported.

2.2.3 Thresholds

To discriminate between driving style categories (AGG, DES, and CAU), a lower bound and upper bound acceleration threshold was determined. Not all acceleration features analysed in sections 2.2.1 and 2.2.2 were used to develop thresholds since a lower and upper bound threshold can only be created if the accelerations for any feature in category AGG are higher than the accelerations in category DES and CAU, respectively. Therefore, for a threshold to be created for any feature, the mean of category AGG must be highest, followed by category DES and category CAU, respectively, as stated in criterion 1. This criterion is designed to operationalize the theoretically anticipated hypothesis that the more assertive and aggressively a driver drives, the higher the in-vehicle accelerations will be (Chhabra et al., 2019; Wang et al., 2022). This criterium also ensures consistency between the thresholds determined for the features, where the upper bound always discriminates between AGG and DES, and the lower bound always discriminates between CAU and DES.

$$\bar{x}_{CAU} < \bar{x}_{DES} < \bar{x}_{AGG} \qquad \text{Crit. (1)}$$

The baseline driving style was DES, since this driving style includes rides which passed the driving exam. If the feature (e.g., acceleration during turns in 30 km/h zones) complied with criterion 1, the threshold was calculated using the difference in means between style CAU and DES for the lower bound and between style DES and AGG for the upper bound, see equations 1 and 2.

$$Upper\ bound\ threshold = \frac{|\bar{x}_{AGG} - \bar{x}_{DES}|}{2} + \bar{x}_{DES} \qquad \text{Eq. (1)}$$

$$Lower\ bound\ threshold = \frac{|\bar{x}_{DES} - \bar{x}_{CAU}|}{2} + \bar{x}_{CAU} \qquad \text{Eq. (2)}$$

The thresholds determined using equations 1 and 2 do not take statistical differences or effect size into account. In order to determine the meaningfulness of a threshold with regard to discriminating between driving styles, a second criterion is stated where the Cohen's *d* effect size of a feature between two driving style categories must indicate a large effect size (i.e., >0.8).

$$Cohen's\ d > 0.8 \qquad \text{Crit. (2)}$$

2.2.4 Classification

A classification between driving styles and pass/fail verdict was done to investigate how well the features and their associated thresholds were able to discriminate between the three different driving styles and, subsequently, desired and undesired driving behaviour. Due to the possibility that some features did not comply with criterion 1, the classification using these features was not performed, since no thresholds were created for these features (see section 2.2.3).

The first classification was done for the ten rides included in categories AGG, DES, and CAU. Based on the thresholds determined using equations 1 and 2 and the features calculated from the acceleration measurements, the rides were classified into the three categories (see Figure 4: driving style level) using only one individual feature and their associated upper and lower bound threshold. The accuracy is defined as the percentage of the cases where the prediction of a driving style by means of the thresholds is equal to the actual driving style as described in the scenario.

After that, the rides listed in the earlier neglected category, NGL, were also included for classification purposes, totalling 18 rides. These rides cannot be classified over the three analysed driving styles (i.e. AGG, DES, and CAU), and were therefore classified over the exam result associated with the ride (see Figure 4: exam result level) based on the accelerations. Here, the accuracy is defined as the percentage of the cases where the prediction of the exam verdict is equal to the actual exam verdict.

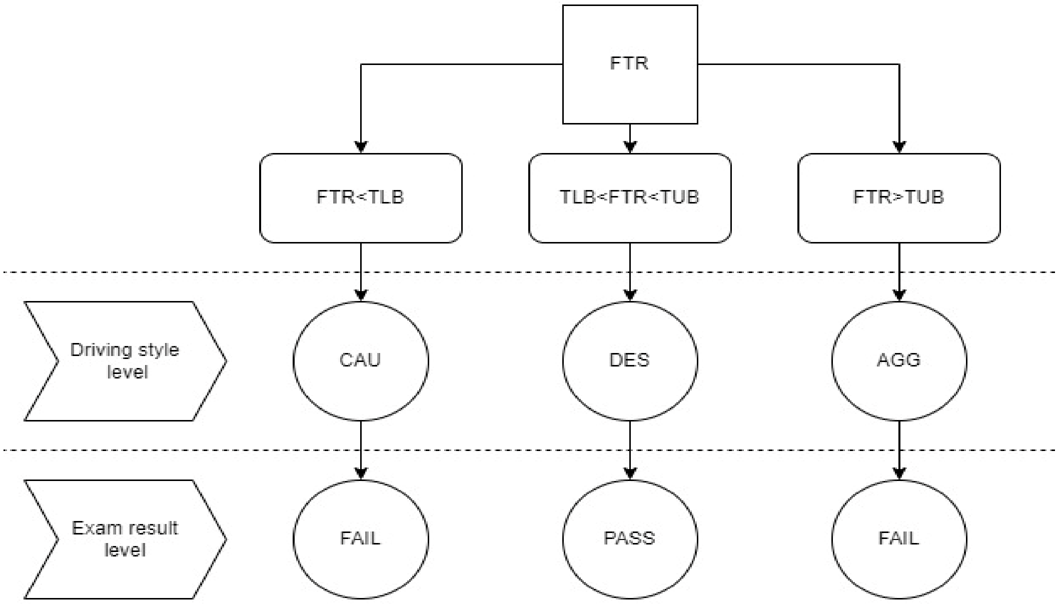

Figure 4. Classification tree. The feature (FTR) is classified using the lower bound threshold (TLB) and the upper bound threshold (TUB). The outcome variables are CAU/DES/AGG at the driving style level, and PASS/FAIL at the exam result level.

3 Results

An overview of the ride setup for each of the 21 rides is provided in Table 2. In total, the 21 rides took 670 min (*M = 32.07, SD = 4.94*) of driving, and covered a total distance of 567 km (*M = 27.22, SD = 7.32*). In Figure 5, three exemplary routes driven during the experiments are shown.

Table 2.

All 21 rides driven during the experiment with their environmental factors. The ride ID is based on test day (i.e., Ride 3.2 was the second ride on the third day of the experiment). Note that rides 1.1, 1.2 and 2.1 were excluded due to incorrect placement of the iPhone between the backseats of the car.

Ride	Coach ID	Car	Day	Time	Weather conditions
1.1	1	Skoda Octavia	30/03/2022	10:12-10:45	Cloudy
1.2	1	Skoda Octavia	30/03/2022	11:20-11:53	Cloudy
1.3	1	Skoda Octavia	30/03/2022	13:13-13:47	Cloudy
1.4	1	Skoda Octavia	30/03/2022	14:26-14:55	Cloudy
2.1	1	Peugeot 308 sw	31/03/2022	10:18-10:55	Rainy
2.2	1	Peugeot 308 sw	31/03/2022	11:27-11:58	Rainy
2.3	1	Peugeot 308 sw	31/03/2022	13:22-13:56	Rainy
3.1	2	Peugeot 308 sw	04/04/2022	09:23-09:50	Rainy
3.2	2	Peugeot 308 sw	04/04/2022	10:13-10:46	Rainy
3.3	2	Peugeot 308 sw	04/04/2022	11:35-12:04	Rainy
3.4	2	Peugeot 308 sw	04/04/2022	13:11-13:40	Rainy
3.5	2	Peugeot 308 sw	04/04/2022	13:58-14:31	Rainy
4.1	2	Peugeot 308 sw	07/04/2022	09:21-09:55	Sunny
4.2	2	Peugeot 308 sw	07/04/2022	10:10-11:07	Sunny
5.1	3	Volkswagen T-Roc	11/04/2022	10:50-11:20	Sunny
5.2	3	Volkswagen T-Roc	11/04/2022	13:06-13:31	Sunny
5.3	3	Volkswagen T-Roc	11/04/2022	14:01-14:30	Sunny
6.1	4	Seat Ateca	13/04/2022	09:21-09:54	Cloudy
6.2	4	Seat Ateca	13/04/2022	10:06-10:33	Cloudy
6.3	4	Seat Ateca	13/04/2022	11:00-11:30	Cloudy
6.4	4	Seat Ateca	13/04/2022	13:01-13:58	Cloudy

Due to incorrect placement of the iPhone between the backseats of the car, rides 1.1, 1.2, and 2.1 showed corrupted acceleration data. These rides were therefore excluded from further analyses, thus resulting in a total of 10 rides with valid acceleration recordings (Category NGL also being excluded; see Figure 5).

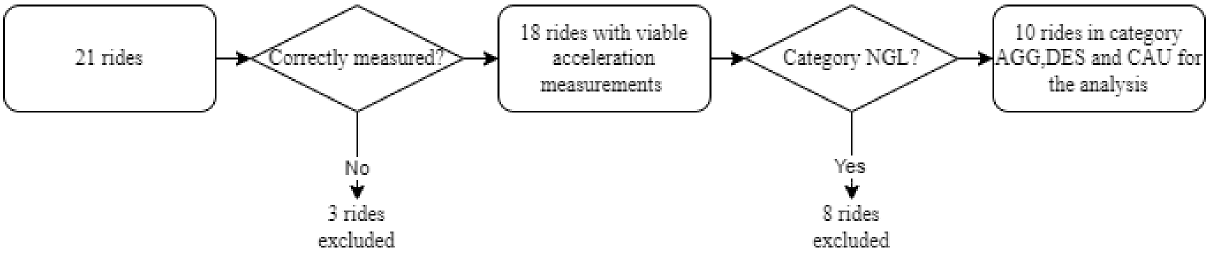

Figure 5. Flowchart indicating how some of the rides included in the dataset were excluded.

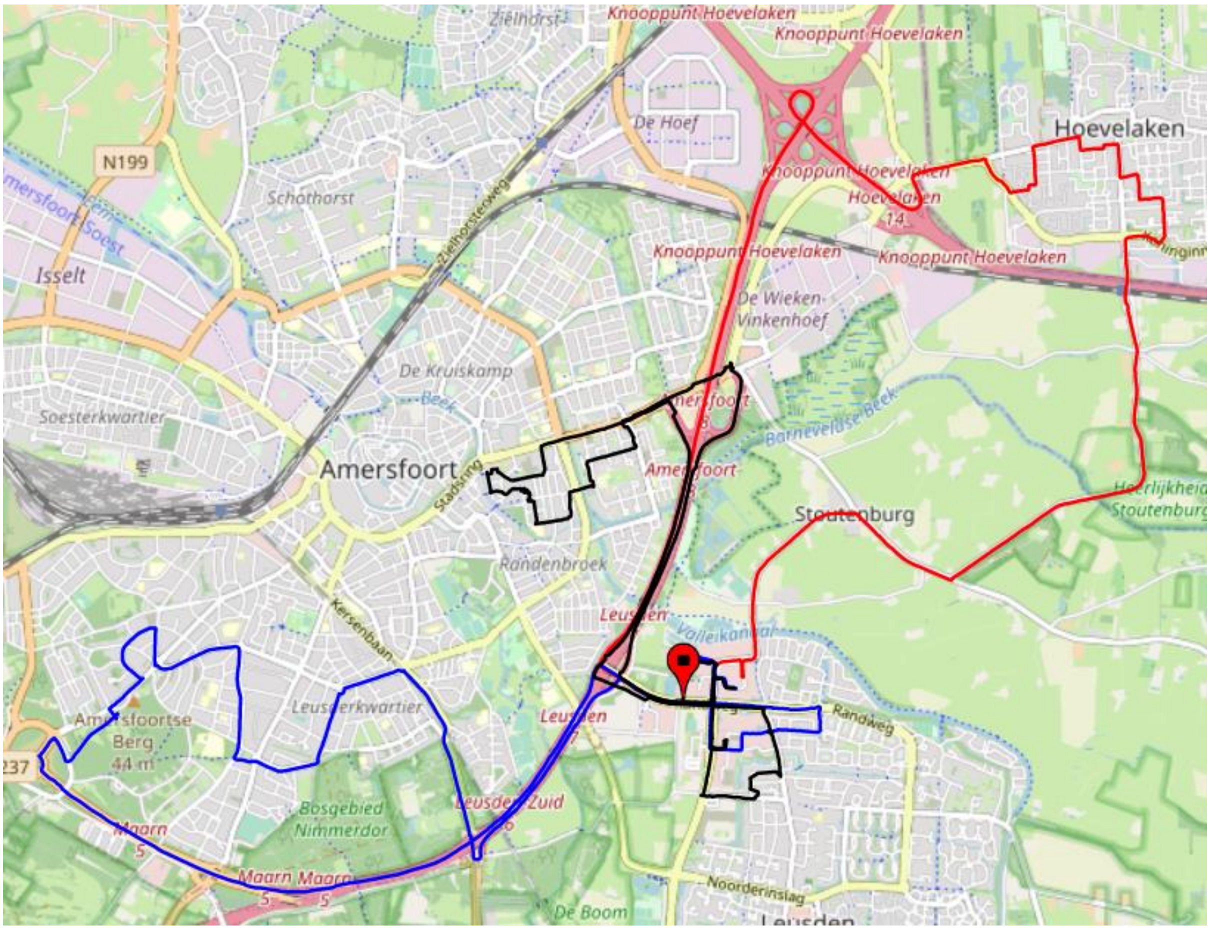

Figure 6. The routes of three rides driven during the experiment (i.e., rides 1.3, 2.3, and 3.3), colour coded per driving style: Category AGG = red, Category DES = black, Category CAU = blue.

3.1 Full-ride analysis

Table 3 shows the lateral, longitudinal and gravitational acceleration per driving style. For the corresponding values per ride, see Appendix F.

Table 3.

Descriptive statistics of the acceleration amplitudes per driving style measured during the entire driving exam.

Direction	**Category**	***M***	***SD***	**Max**	**Min**	***n***
Lateral						
	AGG	0.0474	0.0795	0.0524	0.0433	3
	DES	0.0415	0.0690	0.0444	0.0389	5
	CAU	0.0364	0.1039	0.0398	0.0329	2
Longitudinal						
	AGG	0.0638	0.0965	0.0752	0.0502	3
	DES	0.0513	0.0732	0.0636	0.0432	5
	CAU	0.0538	0.1137	0.0697	0.0377	2
Gravitational						
	AGG	0.0607	0.0983	0.0618	0.0400	3
	DES	0.0554	0.0796	0.0684	0.0387	5
	CAU	0.0400	0.1141	0.0493	0.0308	2

The mean amplitude in g-force of the acceleration signal is highest for the rides in category AGG for all three acceleration directions. The lowest mean amplitude for lateral and gravitational accelerations was for Category CAU, whereas for longitudinal acceleration, category DES portrayed the lowest values on average.

3.2 Event analysis

In Figure 7, three snapshots (of ride 3.1) are given of a heatmap in which acceleration data has been plotted for the three acceleration directions (lateral, longitudinal, and gravitational). From the snapshots in this example, it can be seen that the lateral acceleration is the highest through corners. In addition, no clear signs of discrimination between straight roads or corners can be seen based on longitudinal and gravitational accelerations.

Figure 7. Snapshots of heat maps of lateral (left), longitudinal (middle), and gravitational acceleration (right). The scale ranges from 0 g (red) to 0.5 g (yellow).

In total, 270 turns ($M = 28.94$, $SD = 6.82$), 59 speed increases ($M = 7.00$, $SD = 2.50$), 129 speed decreases ($M = 14.33$, $SD = 6.30$), and 51 lane changes ($M = 5.22$, $SD = 3.12$) were registered by the experimenter during the 10 rides included in this analysis. An overview of the total number of events per driving style is listed in Table 4.

Table 4.

Total events registered for each driving style, with *n* representing the number of rides listed in each driving style category.

Driving style	**Turns**	**Speed increases**	**Speed decreases**	**Lane changes**	***n***
AGG	70	29	58	13	3
DES	143	18	47	27	5
CAU	57	12	24	11	2
Total	270	59	129	51	10

3.2.1 Turns

Figure 8 shows the amplitude of the acceleration signal during the turns per driving style. In Appendix G, Table 14, descriptive statistics are given for the turns in each category.

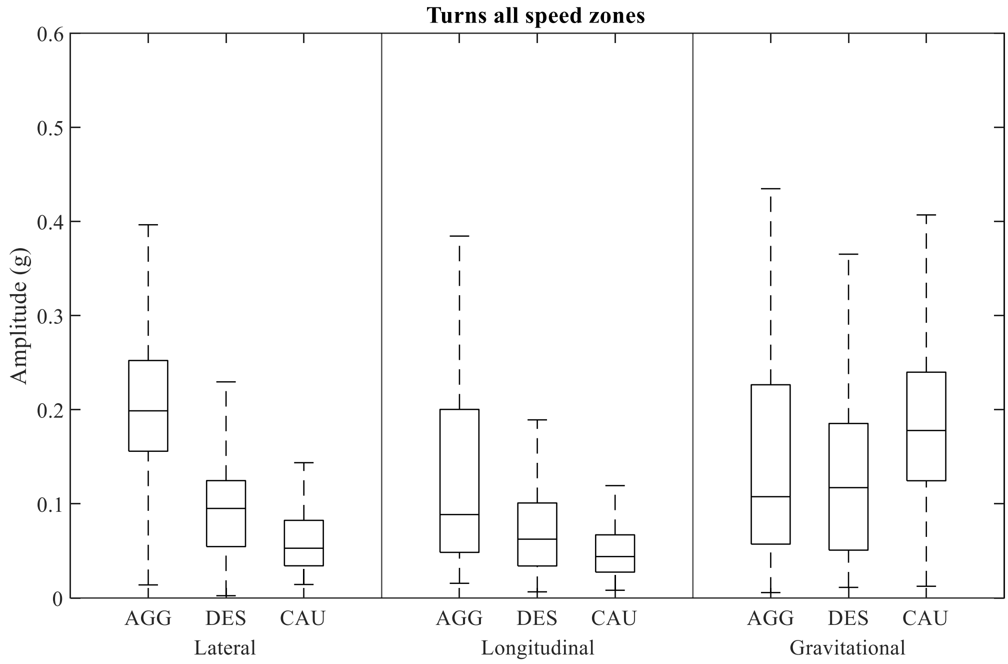

Figure 8. Box plots of the acceleration amplitude measured during each turn event per driving style category and acceleration direction.

The lateral and longitudinal accelerations during turns show the highest amplitude for the AGG driving style. The difference in the average amplitude between the three driving styles is statistically significant for lateral ($X^2(269) = 74.19$, $p < 0.001$) and longitudinal ($X^2(269) = 25.56$, $p < 0.001$) directions. The gravitational direction did not show a significant difference between the driving styles ($X^2(269) = 12.05$, $p = 0.109$).

Figure 9 and Figure 10 show the amplitude of the acceleration per driving style for turns in 30 km/h and 50 km/h zones, respectively.

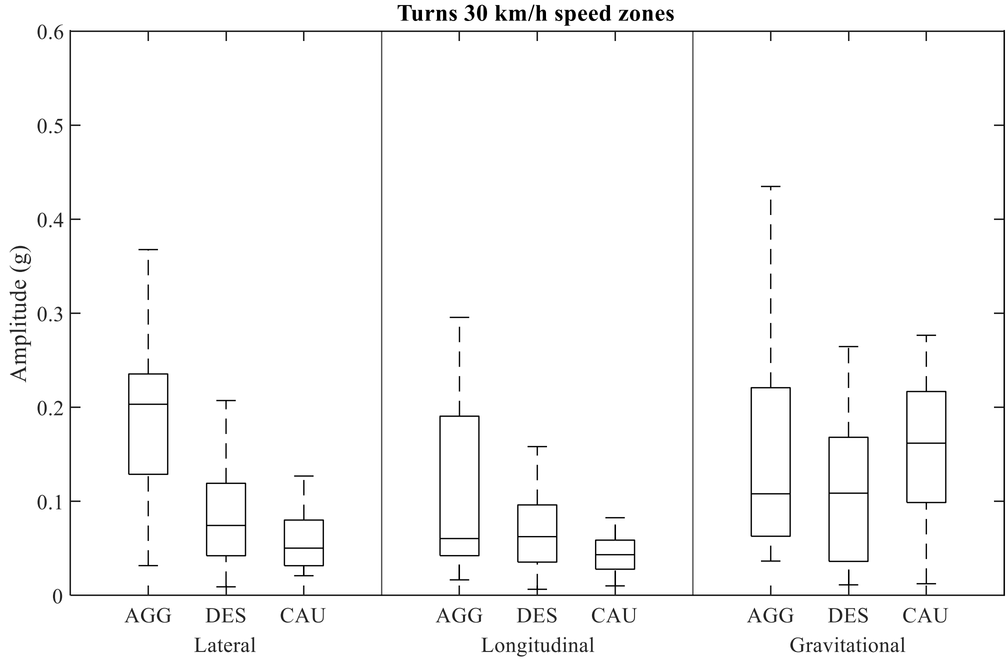

Figure 9. Box plots of the amplitude measured during each event per driving style category and acceleration direction.

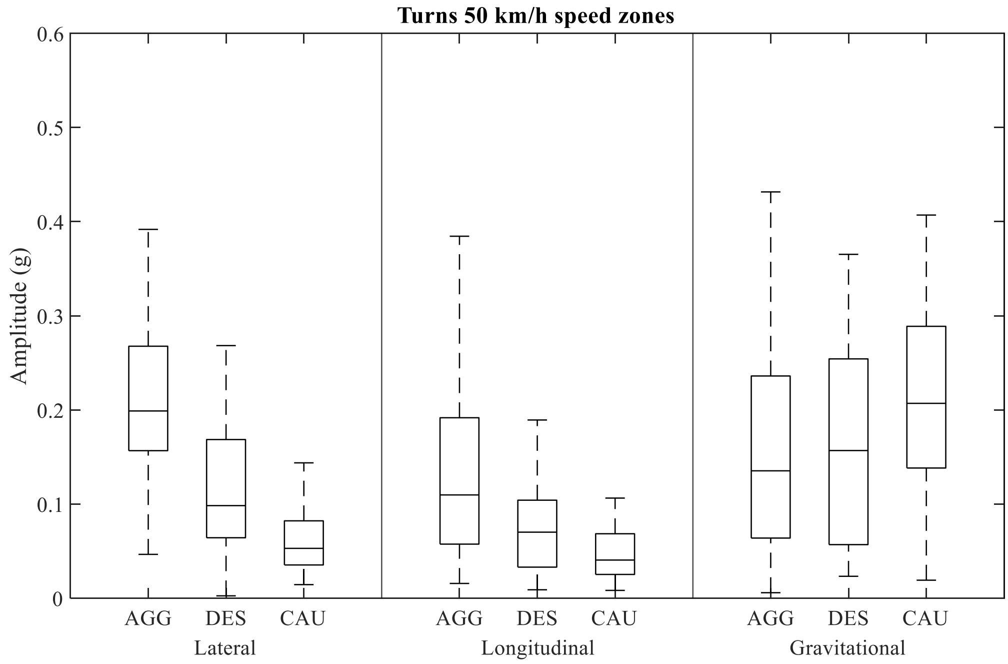

Figure 10. Box plots of the amplitude measured during each event per driving style category and acceleration direction.

The average amplitude for all three acceleration directions in these 30 km/h zones was significantly different between styles (lateral: $X^2(119) = 33$, $p < 0.001$; longitudinal: $X^2(119) = 12.41$, $p < 0.001$; gravitational: $X^2(119) = 10.03$, $p < 0.001$). For the turns in 50 km/h zones, the lateral and longitudinal accelerations were highest in Category AGG, followed by Category DES. A significant difference between categories was found for lateral ($X^2(112)$ =

37.47, $p < 0.001$) and longitudinal acceleration ($X^2(112) = 12.73$, $p = 0.002$). For gravitational acceleration, no significant difference between the three categories was found ($X^2(112) = 4.44$, $p < 0.001$).

The Cohen's *d* effect sizes between the three categories for all three acceleration directions are listed in Table 5.

Table 5.

Overview of all Cohen's *d* effect sizes between the three driving style categories for turns through different speed zones (i.e. all, 30 km/h and 50 km/h zones). In the driving style column (1st column), the two compared categories are listed for each table row.

Driving styles	**Lateral**			**Longitudinal**			**Gravitational**		
	All	**30**	**50**	**All**	**30**	**50**	**All**	**30**	**50**
AGG DES	1.13	1.29	1.01	0.72	1.16	0.61	0.46	0.72	0.30
DES CAU	0.59	0.56	0.22	0.24	0.34	0.15	0.45	0.31	0.58
AGG CAU	1.99	1.81	1.79	0.87	1.07	0.80	0.03	0.53	0.19

From Table 5, it can be observed that the largest difference among the compared groups is found between the lateral accelerations of categories AGG and CAU. From the three speed zone groups, the 30 km/h zone shows the largest effect size for all three acceleration directions when comparing categories AGG and DES. In general, the effect sizes between categories AGG and DES are larger than the effect sizes between categories DES and CAU.

3.2.2 Speed increases

Figure 11 shows the mean amplitude of the acceleration signal during the speed increase events per driving style. In Appendix G, Table 15, descriptive statistics are given for the speed increases in each category.

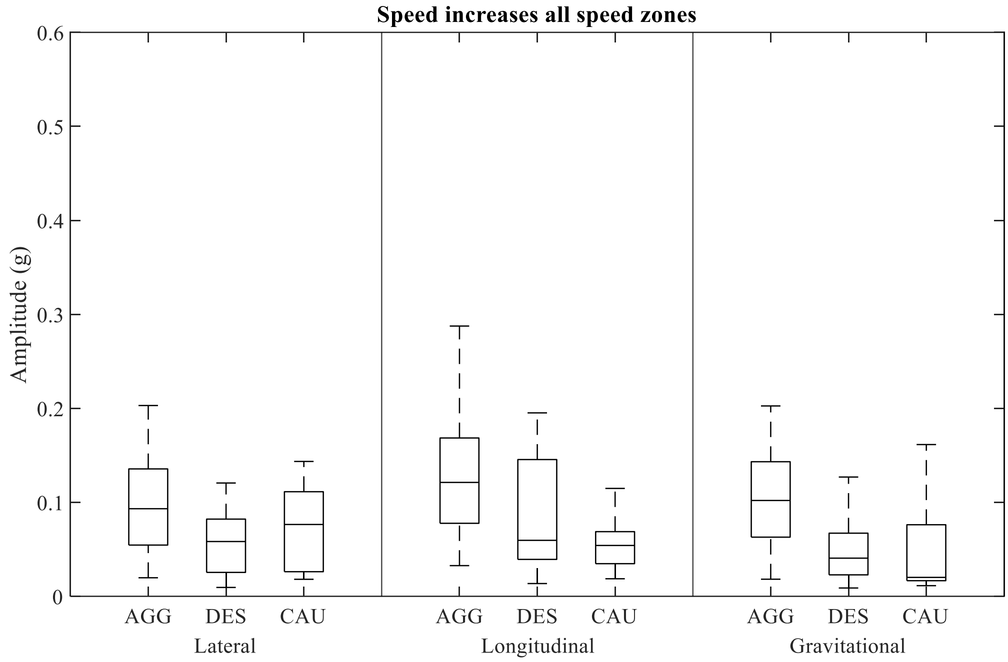

Figure 11. Box plots of the amplitude measured during each event per driving style category and acceleration direction.

Overall, the acceleration amplitudes of the events in Category AGG were the highest for all three directions. There is a significant difference between the acceleration events for lateral accelerations ($X^2(67) = 6.83$, $p = 0.033$), longitudinal accelerations ($X^2(67) = 8.22$, $p = 0.016$) and gravitational accelerations ($X^2(67) = 21.13$, $p < 0.001$).

Figure 12 and Figure 13 show the amplitude of the acceleration per driving style for speed increase events in 30 km/h and 50 km/h zones, respectively.

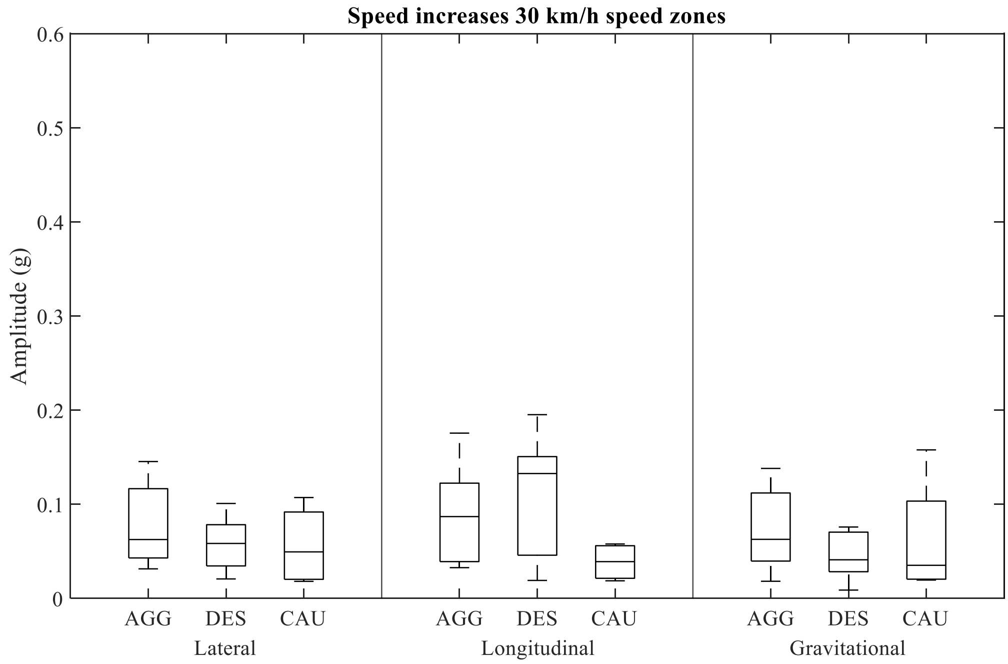

Figure 12. Box plots of the amplitude measured during each event per driving style category and acceleration direction.

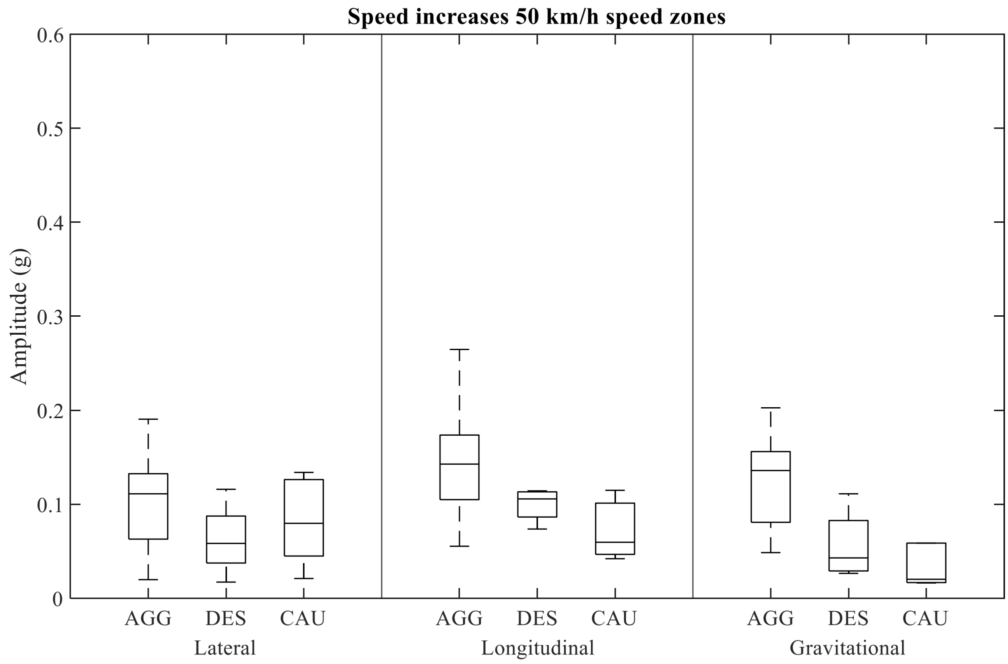

Figure 13. Box plots of the amplitude measured during each event per driving style category and acceleration direction.

In the 30 km/h speed zones, no clear distinction is visible in terms of amplitude between categories or acceleration direction. Comparing the categories, no significant differences were found for speed increase events in this speed zone (lateral: $X^2(16) = 1.87$, $p = 0.393$; longitudinal: $X^2(16) = 1.47$, $p = 0.479$; gravitational: $X^2(16) = 2.2$, $p = 0.333$). From the 50 km/h speed increase events listed in Figure 13, it can be seen that the acceleration events are highest in Category AGG for both longitudinal and gravitational directions. No significant difference was found comparing the three categories in terms of their lateral acceleration amplitude ($X^2(23) = 2.03$, $p = 0.363$), whilst the other two directions did show significant differences (longitudinal: $X^2(23) = 8.02$, $p = 0.018$; gravitational: $X^2(23) = 8.48$, $p = 0.014$).

The effect sizes between the three categories for all three acceleration directions are listed in Table 6.

Table 6.

Overview of all Cohen's *d* effect sizes between each of the three driving style categories for speed increases turns through different speed zones (i.e. all, 30 km/h and 50 km/h zones). In the driving style column (1st column), the two compared categories are listed for each table row.

Driving styles	**Lateral**			**Longitudinal**			**Gravitational**		
	All	**30**	**50**	**All**	**30**	**50**	**All**	**30**	**50**
AGG DES	0.64	0.06	0.20	0.39	0.38	1.04	0.21	0.34	1.21
DES CAU	0.18	0.54	0.54	0.48	0.76	0.03	0.00	0.37	0.30
AGG CAU	0.71	0.85	0.64	0.93	0.66	1.05	0.90	0.86	1.58

Overall, the largest effect sizes are present in the 50 km/h speed zones, with Cohen's *d* values larger than 1. In terms of acceleration directions, lateral and gravitational show the largest effect sizes.

3.2.3 Speed decreases

Figure 14 shows the mean amplitude during the speed decrease events per driving style, regardless of the speed zone they were driven in. In Appendix G, Table 16, descriptive statistics are given for the speed decreases in each category.

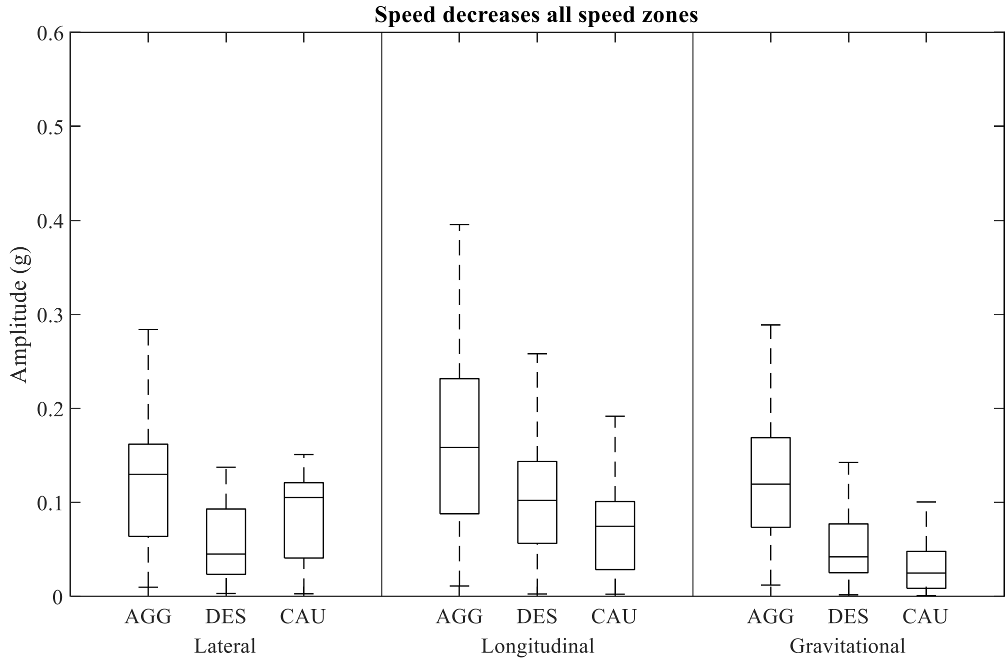

Figure 14. Box plots of the amplitude measured during each event per driving style category and per acceleration direction.

A clear pattern is visible from Figure 14, where for longitudinal and gravitational accelerations Category AGG shows the highest amplitudes, followed by Category DES. Looking at the lateral accelerations, Category CAU shows a higher amplitude than Category DES. All three acceleration directions show significant differences between the categories (lateral: $X^2(143) = 16.94$, $p < 0.001$; longitudinal: $X^2(143) = 21.41$, $p < 0.001$; gravitational: $X^2(143) = 58.45$, $p < 0.001$).

Figure 15 and Figure 16 show the amplitude of the acceleration per driving style for speed decrease events in 30 km/h and 50 km/h zones, respectively.

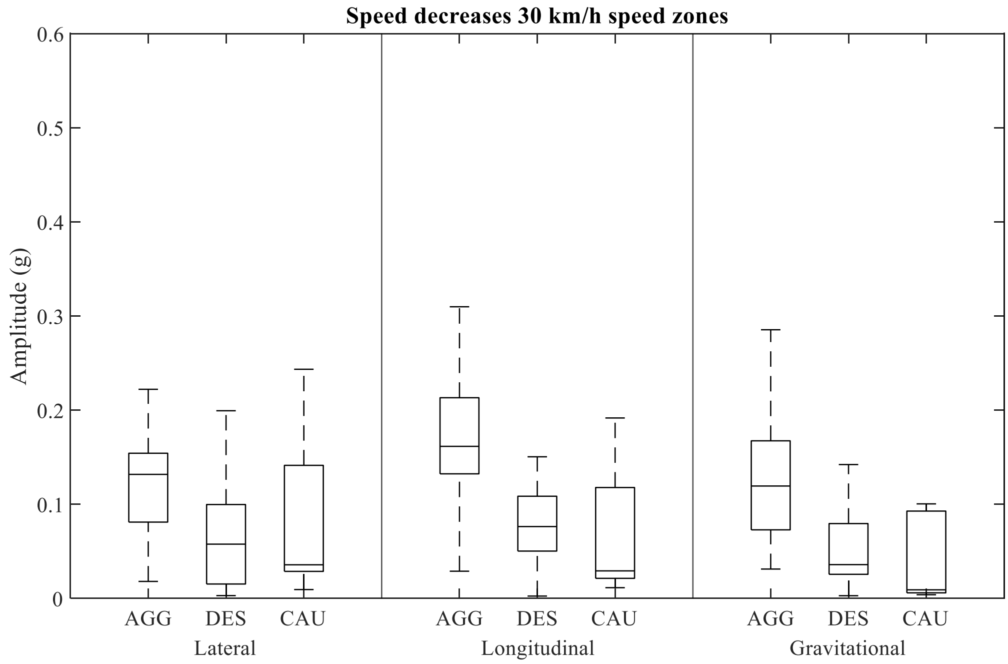

Figure 15. Box plots of the amplitude measured during each event per driving style category and acceleration direction.

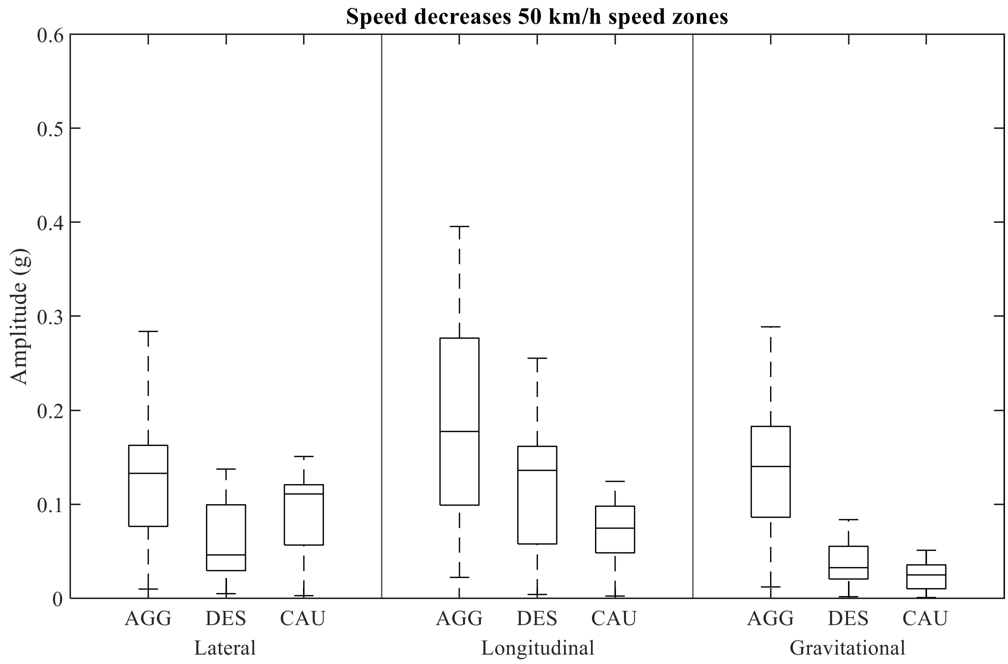

Figure 16. Box plots of the amplitude measured during each event per driving style category and acceleration direction.

In both Figure 15 and Figure 16, it can be seen that Category AGG shows the highest amplitudes during the speed decrease events in 30 km/h and 50 km/h speed zones respectively. For lateral acceleration in the 30 km/h speed zones, no significant differences were found ($X^2(35) = 2.14$, $p = 0.344$). For longitudinal and gravitational acceleration during the speed decrease events, there was a significant difference between the three categories (longitudinal: $X^2(35) = 9.75$, $p < 0.001$; gravitational: $X^2(35)= 35$, $p < 0.001$).

The effect sizes between the three categories for all three acceleration directions are listed in Table 7.

Table 7.

Overview of all Cohen's *d* effect sizes between each of the three driving style categories for speed decreases turns through different speed zones (i.e. all, 30 km/h and 50 km/h zones). In the driving style column (1st column), the two compared categories are listed for each table row.

Driving styles	**Lateral**			**Longitudinal**			**Gravitational**		
	All	**30**	**50**	**All**	**30**	**50**	**All**	**30**	**50**
AGG DES	0.76	0.40	0.98	0.83	0.92	0.64	1.16	1.25	1.48
DES CAU	0.11	0.14	0.29	0.36	0.35	0.46	0.74	0.31	0.46
AGG CAU	0.61	0.54	0.83	0.87	0.98	0.74	1.74	1.30	1.01

Overall, the effect sizes are largest for gravitational acceleration for all speed zones, followed by longitudinal and lateral subsequently.

3.2.4 Lane changes

Figure 17 shows the mean amplitude during the lane changes per driving style, regardless of the speed zone they were driven in. In Appendix G, Table 17, descriptive statistics are given for the lane changes in each category.

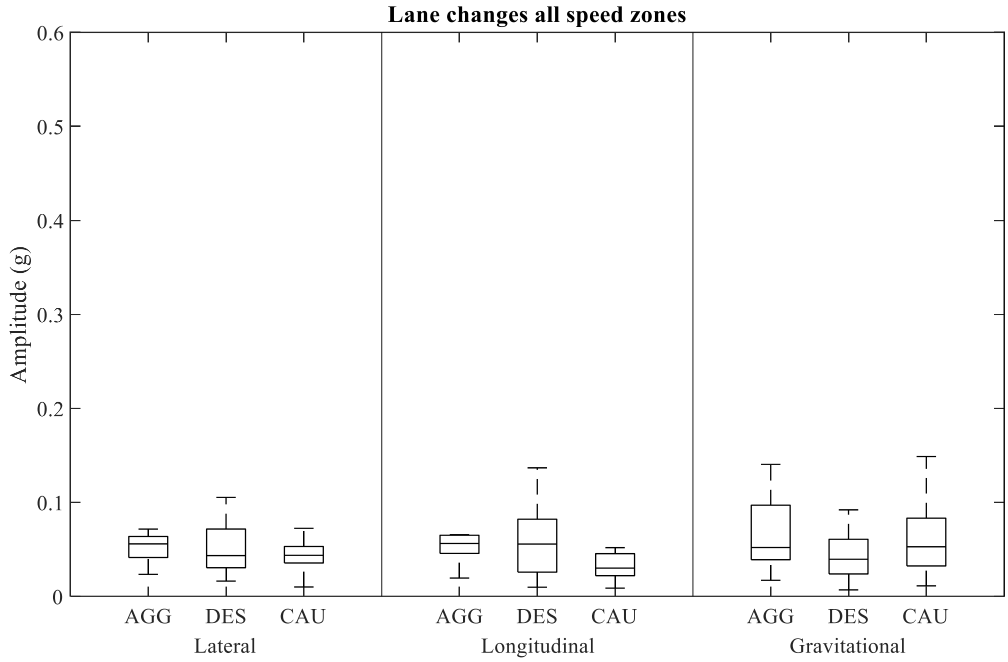

Figure 17. Box plots of the amplitude measured during each event per driving style category and acceleration direction.

For all three acceleration directions, the amplitudes in Category AGG are highest followed respectively by the amplitudes in category DES and CAU. The differences between the three categories are statistically significant for all three acceleration directions (lateral: $X^2(54) = 11.08$, $p < 0.001$; longitudinal: $X^2(54) = 7.44$, $p = 0.024$; gravitational: $X^2(54) = 13.45$, $p < 0.001$).

In Figure 18, the mean amplitude is given of the lane changes driven in 50 km/h speed zones. In contrast to the significant differences found whilst disregarding speed zones, the lane changes within 50 km/h speed zones do not show significant differences for any acceleration direction (lateral: $X^2(24) = 2.96$, $p = 0.228$; longitudinal: $X^2(24) = 0.68$, $p = 0.713$; gravitational: $X^2(24) = 3.52$, $p = 0.172$).

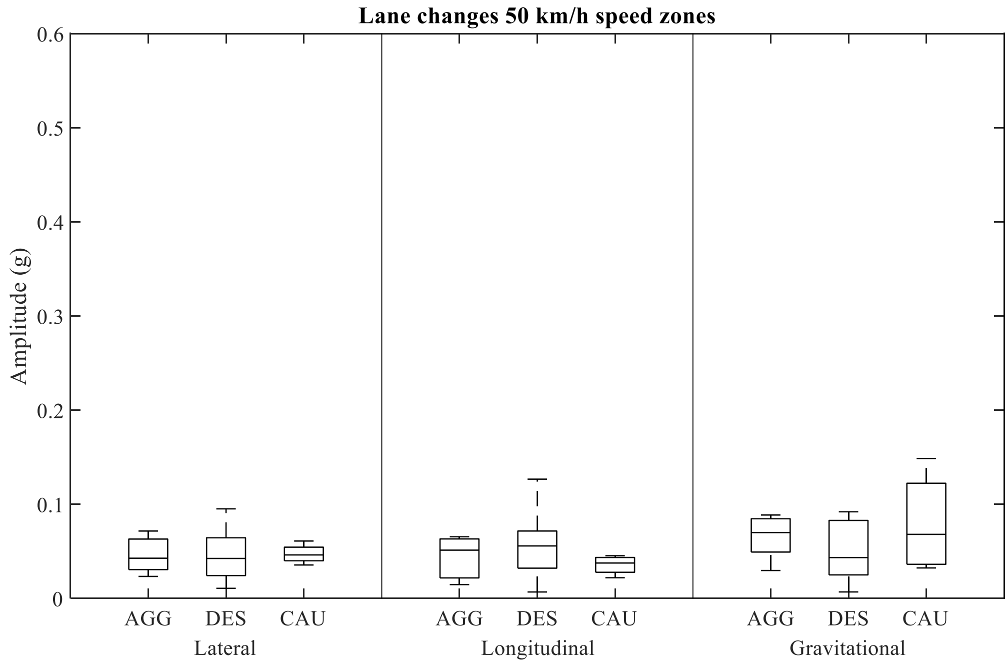

Figure 18. Box plots of the amplitude measured during each event per driving style category and acceleration direction.

The effect sizes between the three categories for all three acceleration directions are listed in Table 8.

Table 8.

Overview of all Cohen's *d* effect sizes between each of the three driving style categories for lane changes turns through different speed zones (i.e. all, 30 km/h and 50 km/h zones). In the driving style column (1st column), the two compared categories are listed for each table row.

Driving styles	Lateral		Longitudinal		Gravitational	
	All	50	All	50	All	50
AGG DES	0.99	0.60	0.43	0.14	1.23	0.88
DES CAU	0.01	0.75	0.42	0.09	0.00	0.05
AGG CAU	0.97	0.61	0.89	0.01	1.15	0.79

From Table 8, it can be observed that the effect sizes are larger between AGG and DES opposed to DES and CAU. As for the speed zones, the effect sizes are larger for all speed zones, as opposed to only the effect sizes measured at the 50 km/h zone lane changes.

3.3 Thresholds

In Table 9, the thresholds for the features analysed through sections 3.1 and 3.2 features are listed. The threshold for each feature consists of an upper bound for discriminating AGG from DES and a lower bound to discriminate between DES and CAU. Some features did not comply with criterion 1 and therefore do not have thresholds (see Figure 19).

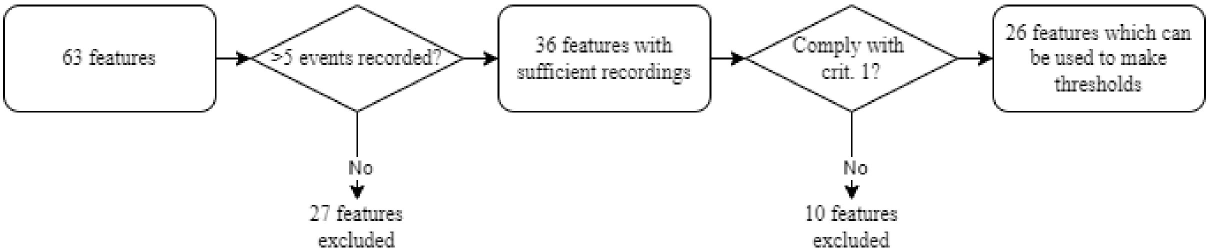

Figure 19. Flowchart indicating why some features were excluded from the data analyses and why some features were not included when making thresholds.

Table 9.

Thresholds determined for all features analyses in section 3.1 and 3.2. The asterisk notes a threshold for which the effect size is larger than 0.8 (i.e. complying with criterion 2, see 2.2.3).

Feature	Speed zone (km/h)	Dir.	Lower bound threshold (g)	Upper bound threshold (g)
Full ride mean	All	Lat	0.039	0.044
		Lon	0.053	0.058
		Gra	0.048	0.058
Turns	All	Lat	0.084	0.148*
		Lon	0.060	0.092
		Gra	N.A.	N.A.
	30	Lat	0.075	0.134*
		Lon	0.060	0.088
		Gra	N.A.	N.A.
	50	Lat	0.092	0.162*
		Lon	0.064	0.104*
		Gra	N.A.	N.A.
Speed increases	All	Lat	N.A.	N.A.
		Lon	0.075	0.113
		Gra	0.055	0.078
	30	Lat	0.059	0.070
		Lon	0.068	0.093
		Gra	0.052	0.058
	50	Lat	N.A.	N.A.
		Lon	0.056	0.118*

		Gra	0.050	0.097*
Speed decreases	All	Lat	N.A.	N.A.
		Lon	0.094	0.147*
		Gra	0.044	0.091*
	30	Lat	0.065	0.091
		Lon	0.079	0.134*
		Gra	0.039	0.074*
	50	Lat	N.A.	N.A.
		Lon	0.100	0.163*
		Gra	0.034	0.091*
Lane changes	All	Lat	0.041	0.047
		Lon	0.045	0.056
		Gra	N.A.	N.A.
	50	Lat	N.A.	N.A.
		Lon	0.051	0.052
		Gra	N.A.	N.A.

The ten features that did not comply with criterium 1 include gravitational acceleration during turns (n = 3; 30 km/h, 50 km/h, all speed zones), speed decreases (n = 2; 30 km/h, 50 km/h) and lane changes (n = 2; 50 km/h, all speed zones) and lateral acceleration during speed increases (n = 2; 50 km/h, all speed zones), speed decreases (n = 1; all speed zones).

The effect sizes between categories DES and CAU were not large (i.e. Cohen's $d < 0.8$) for any of the features, and did thus not comply with criterium 2. The features which complied with both criteria 1 and 2 are marked with an asterisk in Table 9

.

3.4 Classification

The thresholds listed in Table 9 were used to classify the rides between the three different driving style categories AGG, DES, and CAU. Table 10 shows the results of the classification.

Table 10.

Accuracies for all thresholds listed in Table 9. The number of rides included in the classification is annotated with ‘n = ’.

Feature	Speed zone (km/h)	Dir.	Accuracy based on driving style (n = 10) (%)	Accuracy based on pass/fail (n = 18) (%)
Full ride mean	All	Lat	50.00	83.33
		Lon	40.00	77.78
		Gra	20.00	66.67
Turns	All	Lat	90.00	77.78
		Lon	70.00	66.67
		Gra	N.A.	N.A.
	30	Lat	70.00	72.22
		Lon	70.00	83.33
		Gra	N.A.	N.A.
	50	Lat	70.00	72.22
		Lon	60.00	72.22
		Gra	N.A.	N.A.
Speed increases	All	Lat	N.A.	N.A.
		Lon	60.00	72.22
		Gra	70.00	72.22
	30	Lat	40.00	66.67
		Lon	70.00	55.56
		Gra	60.00	72.22
	50	Lat	N.A.	N.A.
		Lon	70.00	55.56
		Gra	30.00	50.00
Speed decreases	All	Lat	N.A.	N.A.
		Lon	60.00	77.78
		Gra	80.00	50.00
	30	Lat	50.00	61.11
		Lon	50.00	66.67
		Gra	30.00	44.44
	50	Lat	N.A.	N.A.
		Lon	60.00	72.22
		Gra	60.00	44.44
Lane changes	All	Lat	70.00	66.67
		Lon	50.00	66.67
		Gra	N.A.	N.A.
	50	Lat	N.A.	N.A.
		Lon	50.00	77.78
		Gra	N.A.	N.A.

The individual features classified with an average accuracy of 55.4% (*SD = 15.9)* between the three driving styles. The highest accuracy was found with the mean of the lateral acceleration during turns in all speed zones (90%). Three features had an accuracy lower than a random guess (33.3%): Full ride gravitational, speed increases 50 km/h gravitational and speed decreases 30 km/h gravitational.

For the classification between pass and fail, the average accuracy was 69.1% (*SD = 11.4*). The feature with the highest accuracy was the mean longitudinal acceleration during lane changes, with an accuracy of 88.9%. Furthermore, it can be seen that all lateral and longitudinal features outperform a random guess (50%) whilst four gravitational features (i.e. speed increase in 50 km/h zone and speed decrease in all, 50 and 30 km/h zones) do not.

4 Discussion

This study aimed to investigate the use of in-vehicle acceleration data to distinguish between desired and undesired driving behaviour with thresholds. Scripted exam rides were driven whilst an accelerometer measured all perceived accelerations during the ride. The data from the accelerometer was analysed to investigate differences between different driving styles. Ultimately, thresholds for accelerations were determined and used to classify the rides over different driving style categories and through a pass/fail verdict. In this discussion, the results of the study will be reviewed and the use of thresholds for assessing driving behaviour will be discussed. Lastly, limitations of the study will be addressed whereafter recommendations for further reach are proposed.

4.1 Assessing driving behaviour using accelerations

In order to assess driving behaviour in terms of desired and undesired driving behaviour, one has to determine when it becomes (un)desired, for instance by means of determining thresholds. Commonly, this is left to a researcher’s own interpretation (see e.g., Fazeen et al., 2012; Handel et al., 2014; Malik & Nandal, 2021). The research presented here is unique in using professional driving examiners of the Dutch driving test organisation (CBR) who were following a professional driving examiner training programme for the assessment of driving behaviour. Due to the help of the driving examiners (i.e. coaches), who combined had over a century of experience in the examination of driving behaviour, and their professional driving examiner training programme, the thresholds created in this study have a professional foundation, in terms of assessing driving behaviour. The scenarios used in their training allowed for clear-cut, professionally distinguished differences between desired and undesired behaviour, which served as reliable input to this study.

Since the thresholds were created with the help of examiners, it is important to discuss how assessing driving behaviour using data would be applied in practice. Driessen et al. (2021) investigated the view of examiners on using data to aid during the driving exam. Among other things, they asked examiners what kind of data would be beneficial to aid during the driving exam, rated from not useful at all (1) to very useful (5). Although distance to objects scored highest in terms of usefulness (4.35 out 5), information concerning de-/acceleration scored third highest, with 3.54 out 5. Overall, it was concluded that “examiners are positive about receiving data in the driving test, especially if the data could help them explain their verdict to the candidate" (Driessen et al., 2021, p. 74). The findings from Driessen et al. (2021) and the present study are pertinent in light of recently published recommendations suggesting that the Dutch driving education system might need to consider a shift from a test-led system to a test-and-education-driven approach (Roemer, 2021; see Helman et al., 2017 for comparable European proposals). The report from Roemer (2021) also suggests performing experiments with instrumented vehicles in order to move toward a more competency-based assessment.

Using data to enhance the driving exam verdict can be related to the levels of automation as proposed by Sheridan (2003, p. 358). Driessen et al. (2021) also asked examiners for their opinion on how far they would like the automation of the driving exam to go. On a scale from 1, *"the computer offers no assistance, human must do it all"*, to 10, *"the computer decides everything and acts autonomously, ignoring the human"*, the examiners would accept Level 2 or 3 at maximum: *"the computer offers a complete set of action alternatives"* and *"narrows the selection down to a few"* (Driessen et al., 2021). In other words, the examiners were open to having access to computer-generated material like graphs or scores, but they did not want higher levels of assistance from a computer. The findings of this study, particularly the developed thresholds, could be used here to allow the computer to make a classification, stating the pass/fail verdict based on accelerations measured during the exam. This classification would not be seen as final outcome of the exam, but may aid the examiner in his decision and argumentation.

Although the thresholds would not be used for a direct final exam verdict (e.g. level 10 proposed by Sheridan, 2003), it is relevant to discuss how well the thresholds perform individually. Overall, the thresholds show a wide array in terms of the accuracies from the thresholds, with some performing with accuracies between 80 and 90 percent and some features underperforming a random guess (i.e. lower than 33 or 50 percent for classifying between driving style and exam result, respectively). It is difficult to compare these results with earlier studies since other studies mainly focused on event classification (Bergasa et al., 2014; Chen et al., 2019) or event detection (Daptardar et al., 2015; Fazeen et al., 2011; Sun et al., 2019), whilst this study focused on classifying entire rides (i.e., driving exams). However, Pramunanto et al. (2019) used a Naive Bayes model to classify drivers as defensive, normal, or aggressive using thresholds, where each ride was driven with all three driving styles, and the model classified when each driving style was driven within that specific ride. Their model performed with accuracies varying from 87% to 95% over the different rides. Other approaches in the literature use clustering and/or machine learning techniques, with which higher accuracies can be obtained in the classification between desired and undesired driving (Stefanowski, 2007: 95%; Qin et al., 2018: 97%). Some studies do not classify between driving styles but rather calculate a score based on acceleration measured during the ride (Liao et al., 2020; Warren et al., 2019; Chen et al., 2019) for a more continuous driving behaviour assessment.

For all events investigated in this research, all three acceleration directions (i.e., lateral, longitudinal and gravitational) have been considered to encompass all features possibly obtainable from the accelerometer. The classification accuracies for these directions show that longitudinal accelerations were highest (M = 72.22; SD = 10.64), followed by lateral accelerations (M = 69.23; SD = 9.52) and gravitational acceleration (M = 65.81; SD = 13.58). In terms of events, it was determined that lateral accelerations were highest in turns, whilst for speed increase/decrease events, the longitudinal accelerations showed the largest amplitude. In other studies, only a single acceleration direction was noted as a meaningful threshold (e.g., only lateral acceleration for turns; Fazeen et al., 2011), without looking at how other acceleration directions could influence the classification results. In terms of effect size, not encompassing all three acceleration directions could lead to missing capable features such as gravitational acceleration during speed decrease events. Even though the effect size of the longitudinal acceleration was larger for speed decreases, the gravitational acceleration was still able to classify with an accuracy of 80% between driving styles using the current dataset.

Although most features and acceleration directions complied with criterium 1, ten features did not. This was caused by the fact that for all events, all acceleration directions were taken into account without avoiding any mismatches. Such a mismatch is for example looking at gravitational accelerations during turn events. Most turns did not feature any type of vertical displacement and therefor have limited gravitational accelerations. In addition, most speed increases/decreases took place on straight roads and therefor mostly have higher longitudinal acceleration amplitudes and lower lateral acceleration amplitudes. For the other 26 features which did comply with criterium 1, the relation implied with the criterium might seem trivial as it is expected that aggressive rides contain higher acceleration amplitudes as opposed to the amplitudes from overcautious rides. However, it is to be reminded that the coaches acted out driving behaviour commonly observed during the exam without exaggerating certain aberrant behaviours. For example, when acting out an aggressive scenario (i.e., ride 1.3), the goal of the coach was not to drive as aggressively as possible for the entire ride but rather to act as a driving exam candidate who drives more aggressively than desired. The coaches drove these scripted exam rides based on their years of experience in examining driving behaviour. Because most features satisfied criterium 1, it was possible to actually make upper and lower bound thresholds that could differentiate between the three analysed driving styles and classify a proper pass/fail verdict.

4.2 Thresholds for ride classification

In this study, the method to create upper and lower bound thresholds was based on the means of the three analysed driving styles (see section 2.2.3, equation 1 and 2). Reflecting on this method, there are more combinations of upper and lower bound thresholds, which would result in different receiver operator characteristics (ROC). Moreover, by placing the threshold in the middle of two means, as done in this study, the specificity and sensitivity of the classification were both set at 50%. Increasing the upper bound threshold or decreasing the lower bound threshold would result in an increased specificity, meaning that more rides would be predicted as “pass”, but also decreasing its sensitivity (i.e. increasing the false positives). In Figure 20, two ROC curves are shown which show the ROC of both the threshold created in this study (see section 3.3) and a fluctuating threshold, where the upper bound increases and lower bound decreases, starting with 100% sensitivity (i.e. small threshold range) and ending with 100% specificity (i.e. large threshold range). A commonly used performance metric for such comparisons is area under curve (AUC), which is higher for the fluctuating threshold (AUC = 0.64) opposed the threshold created in section 3.3 (AUC = 0.60).

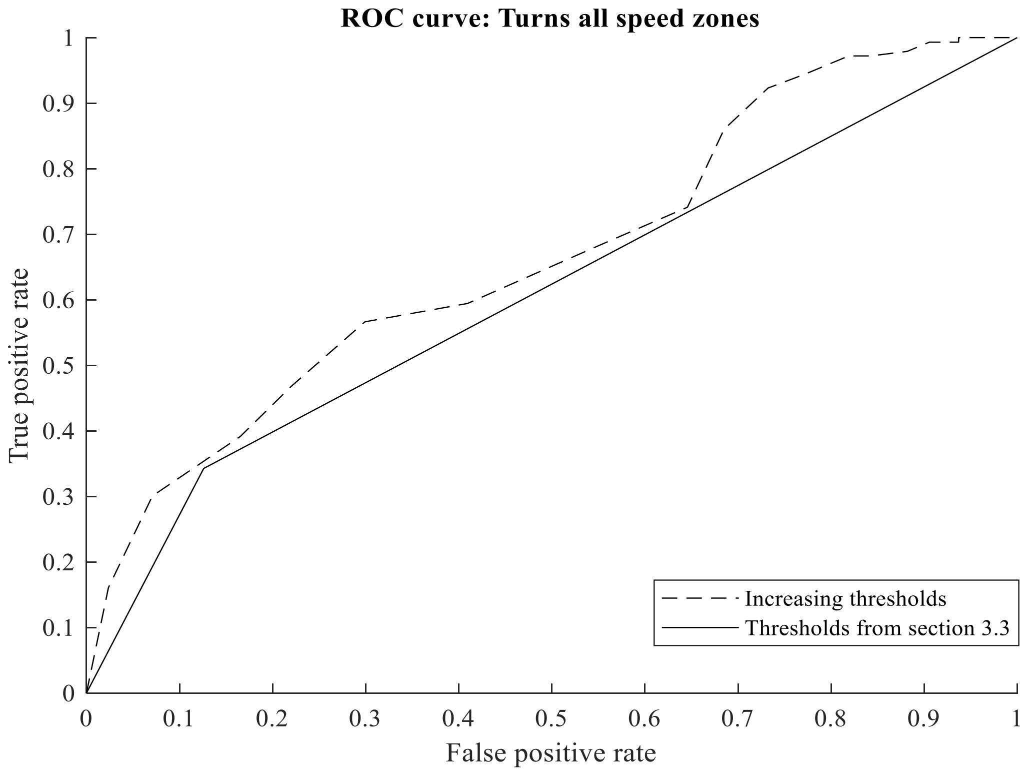

Figure 20. ROC curve of the thresholds created in this study, compared to a threshold with increasing upper bound and decreasing lower bound in step sizes of 0.01g.

ROC optimization is a commonly used technique to optimize classification algorithms and thresholds (e.g. Carrington et al., 2020; Huang & Ling, 2005). However, the higher AUC for the fluctuating threshold and a higher AUC in general do not directly imply a better performance, due to classification-threshold invariance not always being desirable (Lobo et al., 2008). Looking at driving exams, it might be most optimal to have 100% sensitivity, where it is ensured that when a candidate is predicted as "pass", the true outcome also is "pass". Due to the overlap between the acceleration amplitudes, more false negatives (i.e. failing the exam, whilst should pass) will arise due the 100% sensitivity. In the end, the goal of using the thresholds is to aid the driving examiner in their verdict, therefore, a 50/50 balance between sensitivity and specificity (as used in this study) is argued to be the most optimal in giving the examiner a fair judgement from the computer. If the computer would be used for higher level automation, further research should be conducted to investigate the optimal ROC.

In addition to choosing other thresholds and fluctuating upper and lower bounds, the classification using the thresholds can also be performed differently. The introduction shows the wide variety of thresholds used to assess driving behaviour. Although these thresholds are listed alongside each other, it must be stated that these studies vary in the exact use and implementation of the thresholds. In the current study, classification was not based on a single event, but rather on the average of multiple events which occurred during the ride. Alternatively, classification could for instance be based on the number of exceedances within a certain time window of a given threshold (e.g. five times in 60 seconds). The driving exam as conducted by the CBR assesses the driving of the candidate based on the total package, meaning that although a candidate driver could make an error for one or two driving events (e.g. accelerating too harsh through turns), he/she could still pass the exam. The CBR uses a so-called AEX matrix, where the nature (‘Aard’: A), the severity (‘Ernst’: E) and the number of times (X) of a certain driving error are interpreted by the examiner and taken into account for the final verdict (CBR, 2020). Using a single threshold for a certain driving event would not align with the AEX matrix, since the nature (A) and the severity (E) of an event simply cannot be determined using a binary classifier.

An alternative approach to determine the threshold could be based on the largest discriminatory distance. This approach is visualized in Figure 21,where the threshold is varied from 0 to 0.5 to see where the discrimination between any of the three categories is the largest, when the rides are normalized for each minute. For an upper bound threshold (i.e. discriminating AGG from DES) the largest relative distance is found at 0.31 g for lateral acceleration, 0.19 g for longitudinal acceleration and 0.37 g for gravitational acceleration. The main pitfall with this approach is that it does not account for driving events or any contextual relation to the measured accelerations, making it less viable to use as an aid for an examiner.

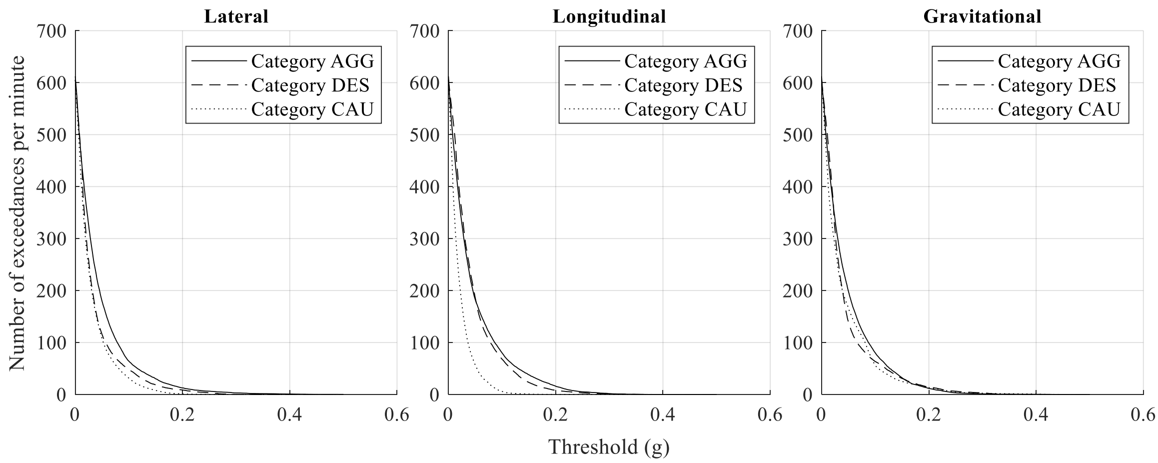

Figure 21. Fluctuating threshold for analyses with respect to the acceleration amplitude. All rides have been normalized for each minute, where in each minute, the number of exceedances is listed per threshold and per driving style.

4.3 Limitations

This study showed how in-vehicle acceleration measurements can be used to classify between desired and undesired driving behaviour. Although the results indicate that for some features the accuracy was as high as 90%, there are some limitations which need to be considered regarding the study.

With a coach acting out a scripted driving behaviour on the open road, it is inevitable to sometimes deviate from the script to avoid hazardous situations. Therefore, where the situation on the road was not suitable to show a certain driving behaviour trait (e.g. aggressive turning or overcautious braking), even though it was listed in the scenario, other than aggressive or overcautious driving has been displayed, and therefor recorded. This could explain the regular overlaps between the three different driving styles in terms of accelerations measured. More data (i.e., more rides) could create a more complete picture of the various driving styles investigated in this study, and therewith overcome this limitation.

A second limitation of this study is with the event registration. The experimenter personally interpreted every traffic situation and registered an event if it was in line with the rules for event registering (section 2.1.3). Due to the fact that detecting events automatically through accelerometer measurements (Bennajeh & Said, 2022; Ylizaliturri-Salcedo et al., 2015) or using GPS and map matching (Zhao et al., 2019) did not show promising results (i.e. accuracies under 90%), the choice for manual even registering was made. Registering turns and lane changes was done objectively, since both events do not require any other metric (such as speed or acceleration) to qualify as an event. However, both speed increases and speed decreases are registered subjectively, since the trigger for registering either of these events would be different between experimenters. In this study, the mean longitudinal acceleration amplitude was 0.1 g ($SD = 0.06$) for all speed increase events and 0.12 g ($SD = 0.09$) for speed decrease events. However, for the experimenter, it was the actual speed increase/decrease which was the main trigger for registering these events instead of taking the severity (i.e. high acceleration amplitude) into account. The experimenter mostly registered an event when the driver increased/decreased their speed with respect to traffic situations such as roundabouts, traffic lights or traffic jams, where speed increases/decreases were required. Although the manual event registering was based on a subjective interpretation of the experimenter in these cases, the experimenter was responsible for the event registering during all of the rides, in order to preserve consistency.

The created dataset included several different information streams (i.e. accelerometer, video, scenarios, event logs), while the dataset was limited in terms of the number of samples and the variety in the driven scenarios. The 21 rides driven during the experiment resulted in only 10 rides which were usable to create the thresholds to classify between the three driving style categories AGG, DES and CAU. Due to this fact, the sample size for the full ride analysis also did not allow for testing for significant differences, since this sample size was simply too small for a proper statistical test (Field, 2017, p. 871). Relatedly, the reason there was an uneven distribution over the driven driving styles (i.e. AGG: n = 3, DES: n = 5, CAU: n = 2 and NGL: n = 11) was due to the fact that the scenarios driven were part of the examiner training programme and not the experimenter's choice. More rides would allow for a more enhanced dataset, while specific (pre-)selection of the rides subject to the research would be necessary to guarantee an even driving style distribution, which inevitably calls for a large time window of the research execution.

The event analysis had a substantially larger number of samples with 509 registered driving events. Even though this made for the possibility to perform statistical tests, there was a sign of violation of independence. For proper functionality, statistical tests such as the Kruskal-Wallis test used in this study require the samples to be independent of one another (Field, 2017). In this study, each event registered was treated independently, although some were registered in the same ride. In such a ride, environmental factors such as the driver, the car and the driven scenario were all consistent for each registered event. Since other environmental factors such as traffic, event location, road type, and duration differed for each event, the assumption of independence for each event was justified because these factors violated the dependence between the events.

4.4 Recommendations

As discussed in section 4.1, many other studies rely on machine learning (ML) techniques for the classification and assessment of driving behaviour such as neural networks (Zhang et al., 2019), dynamic time warping (Chigurupati et al., 2012; Sun et al., 2021) and fuzzy logic (Bennajah & Said, 2020; Kalra et al., 2021). In this study, the classification (section 3.4) was based on a single node decision tree, without looking further into other classification methods. However, the features calculated for all 21 rides could be used with other classification methods, to investigate the performance of other classification techniques besides the technique used in this study. A pitfall of many ML algorithms is overfitting when small sample sizes are used (Subramanian & Simon, 2013). With the dataset created in this study, a ML algorithm appeared unreliable, since the sample size (i.e. 21 rides) is too small (Figueroa et al., 2012; Gowen et al., 2019). For further investigation of the possibilities of using ML on this dataset, it is recommended that the sample size in terms of rides must be significantly expanded (i.e. include more rides/driving exams).

The events being compared still differ in terms of environmental factors such as location, road type, and traffic rules. A more in-depth look could be obtained by comparing the driving styles over the same event, such as a specific cross section commonly taken during a driving exam. Although each ride took a slightly different route in the current dataset, each ride exited the CBR parking area via the same turn. As an example, in Figure 22, the lateral acceleration of three rides from three different driving styles is given. For additional research, it may be preferable to drive over a specific cross section hundreds of times with three different driving styles where external factors such as traffic, vehicle, weather and driver are consistent over each sample. This ensures that the only main difference between samples is the driving style. Another discovery could be the path driven during a driving event. If GPS data were sampled at a higher frequency (50-100 Hz), it could provide extra insight to see the path driven by each driving style during a driving event using merely GPS data (see for example Driessen et al., 2022, where the GPS data of lane changes is investigated).

Another recommendation for further research would be to also include other vehicle data types such as speed (e.g. full-ride average speed, speed through turns; e.g., in Eboli et al., 2019), following distance (e.g. average following distance on highways; e.g., Yimer et al., 2020) or looking behaviour (e.g. average gaze time on the road ahead; e.g., Navarro et al., 2016). Also jerk, the first time derivative of acceleration, is commonly used in other studies to detect undesired driving behaviour (see e.g. Zaki et al., 2014). Whereas acceleration can indicate the severity of a speed decrease event (e.g. abrupt braking), jerk can indicate the smoothness of the speed decrease event (e.g. smooth braking). Including more vehicle data measured during driving exams within the dataset could create a better performance for the classification of driving exams and also give driving examiners more data to explain their verdict to a driving exam candidate.

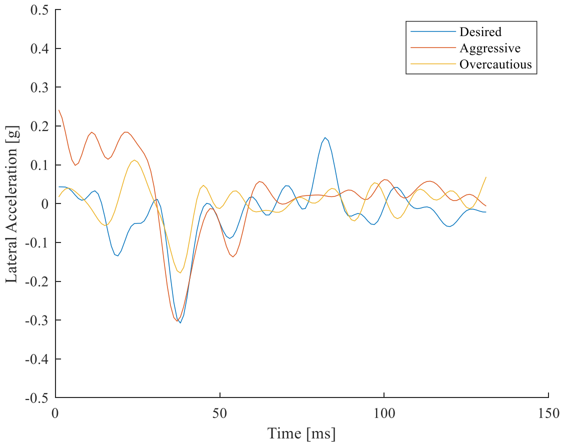

Figure 22. Comparison of the lateral acceleration same turn event (i.e. first turn after leaving CBR parking area). The left image shows the acceleration data, whilst the right image shows the path of the car through extracted using GPS data. This turn is made in a 30 km/h speed zone on cobbled road with traffic inbound from one direction. For this example, three rides from three different driving style categories are shown.

5 Conclusion

With the increasing number of mobile devices that can generate driving-related data, the possibility for using data-driven assessments for driving exams rises. One of the possible data streams, accelerations, was used in this study to classify driving exam rides between driving styles and pass/fail verdicts. An experiment was conducted where 21 recreated driving exams were driven with different driving styles, in order to investigate the differences between driving styles based on in-vehicle acceleration measurements taken during each ride.

It is concluded that to discriminate between desired and undesired driving behaviour, more data streams are needed to achieve perfect classification accuracy. However, there are features which do show significant differences, such as the lateral acceleration through turns in 30 km/h speed zones and longitudinal acceleration during speed decreases in 50 km/h speed zones. Using the thresholds, a classification between the driving style categories (i.e. aggressive, desired, overcautious) can be performed with an accuracy up to 90.0% and classification between pass/fail with accuracies up to 88.9%, all based on the existing dataset. In the end, from the 63 acceleration features derived from the iPhone’s accelerometer, only 15 features showed large effect sizes when compared between the three different driving styles.

This study shows results for solely relying on data to classify between desired and undesired driving behaviour. From the results, it can be concluded that a 100% classification accuracy is not yet achievable using this type of data. In addition, a computer may find it difficult to assess a candidate's higher-order driving skills (e.g. traffic anticipation and intentions) for the same reason that automated vehicles struggle to understand traffic context and predict what other road users will do (Rudenko et al., 2020; Vinkhuyzen & Cefkin, 2016). Altogether, the study emphasizes the fact examiners will still be the main factor responsible for subjecting a pass/fail verdict over a driving exam, aided by data. Not only the examiner could benefit from these developments since candidates could also use it to better understand why they passed/failed their exam.

References

Alger, S., & Sundström, A. (2013). Agreement of driving examiners' assessments – Evaluating the reliability of the Swedish driving test. *Transportation Research Part F: Traffic Psychology and Behaviour*, *19*, 22–30. https://doi.org/10.1016/j.trf.2013.02.004

Arumugam, S., & Bhargavi, R. (2019). A survey on driving behavior analysis in usage based insurance using big data. *Journal of Big Data*, *6*(1). https://doi.org/10.1186/s40537-019-0249-5

Baldwin, K. C., Duncan, D. D., & West, S. K. (2004). The driver monitor system: A means of assessing driver performance. *Johns Hopkins APL Technical Digest*, *25*(3), 269–277.

Baughan, C., & Simpson, B. (1999). Consistency of driving performance at the time of the L-test, and implications for driver testing. *Behavioral Research in Road Safety, Transport Research Laboratory*, *IX*, 206–214.

Bennajeh, A., & Said, L. B. (2022). Autonomous agent adaptive driving control based on fuzzy logic theory and normative behavior. *Journal of Intelligent & Fuzzy Systems*, 1–11. https://doi.org/10.3233/jifs-213498

Bergasa, L. M., Almeria, D., Almazan, J., Yebes, J. J., & Arroyo, R. (2014). Drivesafe: An app for alerting inattentive drivers and scoring driving behaviors. *2014 IEEE Intelligent Vehicles Symposium Proceedings*, 240–245.

Bjørnskau, T. (2003). Stryk eller stå. En undersøkelse av faktorer som påvirker resultatene av praktisk førerprøve [pass or fail. A study of factors that affect the results of the driving test]. *TØI*, *2003*(662).

Boggs, A. M., Arvin, R., & Khattak, A. J. (2020). Exploring the who, what, when, where, and why of automated vehicle disengagements. *Accident Analysis & Prevention*, *136*, 105406. https://doi.org/10.1016/j.aap.2019.105406

Carrington, A. M., Fieguth, P. W., Qazi, H., Holzinger, A., Chen, H. H., Mayr, F., & Manuel, D. G. (2020). A new concordant partial AUC and partial c statistic for imbalanced data in the evaluation of machine learning algorithms. BMC medical informatics and decision making, 20(1), 1-12.

CBR. (2020). *Rijprocedure B*. Retrieved 6 September 2022, from https://www.cbr.nl/nl/voor-rijscholen/nl/partner-in-verkeersveiligheid/rijprocedure.htm

Chen, C., Zhao, X., Zhang, Y., Rong, J., & Liu, X. (2019). A graphical modeling method for individual driving behavior and its application in driving safety analysis using GPS data. Transportation research part F: traffic psychology and behaviour, 63, 118-134.

Chhabra, R., Verma, S., & Rama Krishna, C. (2019). Detecting aggressive driving behavior using mobile smartphone. In Proceedings of 2nd International Conference on Communication, Computing and Networking (pp. 513-521). Springer, Singapore.

Chigurupati, S., Polavarapu, S., Kancherla, Y., & Nikhath, A. K. (2012). Integrated computing system for measuring driver safety index. *Int. J. Emerg. Technol. Adv. Eng.*, *2*(6), 384–388.

Chowdhury, A., Shankaran, R., Kavakli, M., & Haque, M. M. (2018). Sensor Applications and Physiological Features in Drivers' Drowsiness Detection: A Review. *IEEE Sensors Journal*, *18*(8), 3055–3067. https://doi.org/10.1109/jsen.2018.2807245

Dai, J., Teng, J., Bai, X., Shen, Z., & Xuan, D. (2010, March). Mobile phone based drunk driving detection. In 2010 4th International Conference on Pervasive Computing Technologies for Healthcare (pp. 1-8). IEEE.

Daptardar, S., Lakshminarayanan, V., Reddy, S., Nair, S., Sahoo, S., & Sinha, P. (2015, November). Hidden Markov model based driving event detection and driver profiling from mobile inertial sensor data. In 2015 IEEE SENSORS (pp. 1-4). IEEE.

Driessen, T., Picco, A., Dodou, D., de Waard, D., & de Winter, J. (2021, November). Driving examiners' views on data-driven assessment of test candidates: An interview study. *Transportation Research Part F: Traffic Psychology and Behaviour*, *83*, 60–79. https://doi.org/10.1016/j.trf.2021.09.021

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. https://doi.org/10.1016/j.trc.2016.04.002

Engelbrecht, J., Booysen, M. J., Rooyen, G., & Bruwer, F. J. (2015). Survey of smartphone-based sensing in vehicles for intelligent transportation system applications. *IET Intelligent Transport Systems*, *9*(10), 924–935. https://doi.org/10.1049/iet-its.2014.0248

Farag, W. A. (2020). Model-predictive-control complex-path tracking for self-driving cars. *International Journal of Modelling, Identification and Control*, *34*(3), 265. https://doi.org/10.1504/ijmic.2020.111624

Fazeen, M., Gozick, B., Dantu, R., Bhukhiya, M., & González, M. C. (2012). Safe Driving Using Mobile Phones. *IEEE Transactions on Intelligent Transportation Systems*, *13*(3), 1462–1468. https://doi.org/10.1109/tits.2012.2187640

Field, A. (2017). Discovering Statistics Using SPSS. In *No Title*. SAGE Publications.

Figueredo, G. P., Agrawal, U., Mase, J. M. M., Mesgarpour, M., Wagner, C., Soria, D., Garibaldi, J. M., Siebers, P. O., & John, R. I. (2019). Identifying Heavy Goods Vehicle Driving Styles in the United Kingdom. *IEEE Transactions on Intelligent Transportation Systems*, *20*(9), 3324–3336. https://doi.org/10.1109/tits.2018.2875343

Figueroa, R. L., Zeng-Treitler, Q., Kandula, S., & Ngo, L. H. (2012). Predicting sample size required for classification performance. *BMC Medical Informatics and Decision Making* , *12*, Article 8.

Girbés, V., Hernández, D., Armesto, L., Dols, J., & Sala, A. (2019). Drive Force and Longitudinal Dynamics Estimation in Heavy-Duty Vehicles. *Sensors*, *19*(16), 3515. https://doi.org/10.3390/s19163515

Gonzalez, D., Perez, J., Milanes, V., & Nashashibi, F. (2016). A Review of Motion Planning Techniques for Automated Vehicles. *IEEE Transactions on Intelligent Transportation Systems*, *17*(4), 1135–1145. https://doi.org/10.1109/tits.2015.2498841

Goodall, N. J. (2021). Comparison of automated vehicle struck-from-behind crash rates with national rates using naturalistic data. *Accident Analysis &Amp; Prevention*, *154*, 106056. https://doi.org/10.1016/j.aap.2021.106056

Gowen, E., Poliakoff, E., & Casson, A. J. (2019, November 7). Machine learning algorithm validation with a limited sample size. *PLOS ONE*, *14*(11), e0224365. https://doi.org/10.1371/journal.pone.0224365

Handel, P., Ohlsson, J., Ohlsson, M., Skog, I., & Nygren, E. (2014). Smartphone-Based Measurement Systems for Road Vehicle Traffic Monitoring and Usage-Based Insurance. *IEEE Systems Journal*, *8*(4), 1238–1248. https://doi.org/10.1109/jsyst.2013.2292721

Helman, S., Vlakveld, W., Fildes, B., Oxley, J., Fernandez-Medina, K., & Weekley, J. (2017). *Study on driver training, testing and medical fitness*. European Commission - Mobility & Transport - Road Safety. Retrieved 6 September 2022, from https://ec.europa.eu/transport/road_safety/sites/default/files/dl_study_on_training_testing_med_fitness.pdf

Hickman, J. S., & Hanowski, R. J. (2012). An Assessment of Commercial Motor Vehicle Driver Distraction Using Naturalistic Driving Data. *Traffic Injury Prevention*, *13*(6), 612–619. https://doi.org/10.1080/15389588.2012.683841

Huang, J., & Ling, C. X. (2005). Using AUC and accuracy in evaluating learning algorithms. IEEE Transactions on knowledge and Data Engineering, 17(3), 299-310.

Kalra, N., Kumar Goyal, R., Parashar, A., Singh, J., & Singla, G. (2021). Driving Style Recognition System Using Smartphone Sensors Based on Fuzzy Logic. *Computers, Materials &Amp; Continua*, *69*(2), 1967–1978. https://doi.org/10.32604/cmc.2021.018732

Lima, P. F., Collares Pereira, G., Mårtensson, J., & Wahlberg, B. (2018). Experimental validation of model predictive control stability for autonomous driving. *Control Engineering Practice*, *81*, 244–255. https://doi.org/10.1016/j.conengprac.2018.09.021

Lobo, J. M., Jiménez-Valverde, A., & Real, R. (2008). AUC: a misleading measure of the performance of predictive distribution models. Global ecology and Biogeography, 17(2), 145-151.

Malik, M., & Nandal, R. (2021). A framework on driving behavior and pattern using On-Board diagnostics (OBD-II) tool. *Materials Today: Proceedings*. https://doi.org/10.1016/j.matpr.2021.07.376

Marin-Plaza, P., Hussein, A., Martin, D., & Escalera, A. D. L. (2018). Global and Local Path Planning Study in a ROS-Based Research Platform for Autonomous Vehicles. *Journal of Advanced Transportation*, *2018*, 1–10. https://doi.org/10.1155/2018/6392697

Marti, E., de Miguel, M. A., Garcia, F., & Perez, J. (2019). A Review of Sensor Technologies for Perception in Automated Driving. *IEEE Intelligent Transportation Systems Magazine*, *11*(4), 94–108. https://doi.org/10.1109/mits.2019.2907630

MATLAB Mobile (8.9.1). (2017). [IOS application]. Mathworks. https://nl.mathworks.com/products/matlab-mobile.html

McHugh, M. L. (2012). Interrater reliability: the kappa statistic. *Biochemica Media, 22*(3), 276–282.

Miyajima, C., Nishiwaki, Y., Ozawa, K., Wakita, T., Itou, K., Takeda, K., & Itakura, F. (2007). Driver Modeling Based on Driving Behavior and Its Evaluation in Driver Identification. *Proceedings of the IEEE, 95*(2), 427–437. https://doi.org/10.1109/jproc.2006.888405

Navarro, J., François, M., & Mars, F. (2016). Obstacle avoidance under automated steering: Impact on driving and gaze behaviours. Transportation research part F: traffic psychology and behaviour, 43, 315-324.

Osafune, T., Takahashi, T., Kiyama, N., Sobue, T., Yamaguchi, H., & Higashimo, T. (2017). Analysis of accident risks from driving behaviors. *International Journal of Intelligent Transportation Systems Research, 15*(3), 192–202.

Paefgen, J., Kehr, F., Zhai, Y., & Michahelles, F. (2012). Driving behavior analysis with smartphones: Insights from a controlled field study. *Proceedings of the 11th International Conference on Mobile and Ubiquitous Multimedia*, 1–8.

Pramunanto, E., Zaini, A., & Rizkiana, V. (2019, November). Prototype of Driving Behavior Monitoring System Using Naïve Bayes Classification Method. In 2019 International Conference on Computer Engineering, Network, and Intelligent Multimedia (CENIM) (pp. 1-5). IEEE.

Qin, S., Zhang, S., & Wu, M. (2018, December). Clustering-Based Driving Behavior Recognition Using Inertial Sensors. *2018 11th International Symposium on Computational Intelligence and Design (ISCID)*. https://doi.org/10.1109/iscid.2018.00088

Ranft, B., & Stiller, C. (2016). The Role of Machine Vision for Intelligent Vehicles. *IEEE Transactions on Intelligent Vehicles, 1*(1), 8–19. https://doi.org/10.1109/tiv.2016.2551553

Rangesh, A., & Trivedi, M. M. (2019). No Blind Spots: Full-Surround Multi-Object Tracking for Autonomous Vehicles Using Cameras and LiDARs. *IEEE Transactions on Intelligent Vehicles, 4*(4), 588–599. https://doi.org/10.1109/tiv.2019.2938110

Roemer, E. (2021). Van rijles naar rijonderwijs. Advies verbetering autorijscholenbranche [from driving lessons to driving education. Advice for improving the driving school industry]. *The Hague, the Netherlands: Ministry of Infrastructure and Water*

Management. https://www.rijksoverheid.nl/documenten/rapporten/ 2021/04/14/bijlage-1-van-rijles-naar-rijonderwijs-advies-verbeteren-autorijscholenbranche

Rudenko, A., Palmieri, L., Herman, M., Kitani, K. M., Gavrilla, D. M., & Arras, K. (2020). Human motion trajectory prediction: A survey. *The International Journal of Robotics Research*, *39*(8), 895–935.

Schoettle, B. (2017). Sensor fusion: A comparison of sensing capabilities of human drivers and highly automated vehicles. *Report No. SWT-2017-12*.

Sheridan, T. B. (2003, January 1). *Telerobotics, Automation, and Human Supervisory Control (The MIT Press)*. The MIT Press.

Shimshoni, Y., Farah, H., Lotan, T., Grimberg, E., Dritter, O., Musicant, O., Toledo, T., & Omer, H. (2014). Effects of parental vigilant care and feedback on novice driver risk. *Journal of Adolescence*, *38*(1), 69–80. https://doi.org/10.1016/j.adolescence.2014.11.002

Shladover, S. E. (2016). The Truth about “Self-Driving” Cars. *Scientific American*, *314*(6), 52–57. https://doi.org/10.1038/scientificamerican0616-52

Stefanowski, J. (2007). On combined classifiers, rule induction and rough sets. *Transactions on Rough Sets Vi*, *4*.

Subramanian, J., & Simon, R. (2013). Overfitting in prediction models–is it a problem only in high dimensions?. Contemporary clinical trials, 36(2), 636-641.

Sun, R., Cheng, Q., Xie, F., Zhang, W., Lin, T., & Ochieng, W. Y. (2021, January). Combining Machine Learning and Dynamic Time Wrapping for Vehicle Driving Event Detection Using Smartphones. *IEEE Transactions on Intelligent Transportation Systems*, *22*(1), 194–207. https://doi.org/10.1109/tits.2019.2955760

Tabone, W., de Winter, J., Ackermann, C., Bärgman, J., Baumann, M., Deb, S., Emmenegger, C., Habibovic, A., Hagenzieker, M., Hancock, P., Happee, R., Krems, J., Lee, J. D., Martens, M., Merat, N., Norman, D., Sheridan, T. B., & Stanton, N. A. (2021). Vulnerable road users and the coming wave of automated vehicles: Expert perspectives. *Transportation Research Interdisciplinary Perspectives*, *9*, 100293. https://doi.org/10.1016/j.trip.2020.100293

Viator (2016, August 16). Viator. Retrieved 14 September 2022, from https://www.viator.com/nl-NL/tours/Bucharest/Day-Trip-to-Transfagarasan-Road-and-Draculas-Fortress-Poienari-from-Bucharest/d22134-7793P5

Vinkhuyzen, E., & Cefkin, M. (2016). Developing Socially Acceptable Autonomous Vehicles. *Ethnographic Praxis in Industry Conference Proceedings*, *2016*(1), 522–534. https://doi.org/10.1111/1559-8918.2016.01108

Wang, J., Xu, W., Fu, T., Gong, H., Shangguan, Q., & Sobhani, A. (2022). Modeling aggressive driving behavior based on graph construction. Transportation Research Part C: Emerging Technologies, 138, 103654.

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

Wei, W., Gao, F., Scherer, R., Damasevicius, R. & Polap, D. (2021). Design and Implementation of Autonomous Path Planning for Intelligent Vehicle. 網際網路技術學刊, *22*(5), 957–964. https://doi.org/10.53106/160792642021092205002

Yimer, T. H., Wen, C., Yu, X., & Jiang, C. (2020). A study of the minimum safe distance between human driven and driverless cars using safe distance model. arXiv preprint arXiv: 2006.07022.

Ylizaliturri-Salcedo, M. A., Tentori, M., & Garcia-Macias, J. A. (2015). Detecting aggressive driving behavior with participatory sensing. *International Conference on Ubiquitous Computing and Ambient Intelligence*, 249–261.

Zaki, M. H., Sayed, T., & Shaaban, K. (2014). Use of drivers' jerk profiles in computer vision–based traffic safety evaluations. Transportation Research Record, 2434(1), 103-112.

Zeeman, A. S., & Booysen, M. J. (2013). Combining speed and acceleration to detect reckless driving in the informal public transport industry. *16th International IEEE Conference on Intelligent Transportation Systems (Itsc 2013)*, 756–761.

Zhang, J., Wu, Z., Li, F., Luo, J., Ren, T., Hu, S., Li, W., & Li, W. (2019). Attention-Based Convolutional and Recurrent Neural Networks for Driving Behavior Recognition Using Smartphone Sensor Data. *IEEE Access*, *7*, 148031–148046. https://doi.org/10.1109/access.2019.2932434

Zhang, Y., Chen, Y., & Gao, C. (2021). Deep unsupervised multi-modal fusion network for detecting driver distraction. *Neurocomputing*, *421*, 26–38. https://doi.org/10.1016/j.neucom.2020.09.023

Zhao, W., Yin, J., Wang, X., Hu, J., Qi, B., & Runge, T. (2019). Real-Time Vehicle Motion Detection and Motion Altering for Connected Vehicle: Algorithm Design and Practical Applications. *Sensors*, *19*(19), 4108. https://doi.org/10.3390/s19194108

Appendix

A. Consent form

Driving performancebased on acceleration measurements
Consent form for participants

Researchers

MSc student: David Stefan (d.a.Stefan@student.tudelft.nl)
Supervisors: Dr.ir. J.C.F. de Winter (j.c.f.dewinter@tudelft.nl), Ir. Tom Driessen (t.driessen@tudelft.nl), Dr. Dimitra Dodou (d.dodou@tudelft.nl)

Location

Examiner training centre of the CBR, Leusden

This document describes the purpose of this study, the experimental procedure, the right to withdraw, and data handling. Read all sections carefully and answer the questions on page 2. The experiment is conducted by the TU Delft, in cooperation with CBR.

Aim of the research

Driving performance is normally assessed subjectively, with humans having their own interpretation of what desired driving behaviour is. To this end, a sample of experienced driving examiners (instructors), recruited through the Dutch Driving Authority (CBR), will drive in various styles and provide their interpretations about driving behaviour and perceived g-forces. In addition to the instructors' observations, measurements will be made in the car to extract the perceived g-forces. Together, the acceleration measurements and instructors' observations will be used to determine acceleration threshold values distinguishing between good/bad driving performance.

Experimental procedure

The experiment will be conducted at the examiner training centre of the CBR in Leusden, The Netherlands. At this facility, aspirant driving examiners are educated about how to properly asses the driving performance of people participating in the driving exam. During training, experienced driving examiners act as instructors, imitating various driving behaviours of participants in driving exams, associated with good/bad driving performance.

Prior to the drives, you (the instructor) will disclose to the experimenter the driving styles which will be employed during the drive. After that, you will conduct your driving activity as normally when training an aspirant driving examiner.

For the acceleration measurements, the sensors from a smartphone (iPhone X (2018)) will be used. The mobile application MATLAB Mobile will be used to extract the sensor data from the sensors and log them to enable further processing of the acquired sensor data.

To log the examiners' observations and the imitated driving behaviour, the mobile application of "buttons event recorder" will be used. With this app, the user can declare certain buttons, which resemble the events which might occur during the drive.

To enhance the data processing, a GoPro Max will film the public road whilst driving. The GoPro will be mounted on top of the dashboard of the vehicle and will only film the road ahead. As for audio, only the GoPro's frontal microphones will be switched on, since the device can't exclude audio recording entirely. Enabling only the frontal microphones ensures minimal recording of conversations within the vehicle. To ensure no audio will be gathered in the dataset, the experimenter erases all audio immediately after gathering the video through processing in the GoPro video editor. In the dataset itself, only videos will be listed without any audio.

Risks

You face no greater risk than what you would face during your regular working activity. Although highly unlikely, there is a possibility of a traffic accident. In such cases, we bear no liability for the consequences.

Right to withdraw

Your participation in this study is completely voluntary. You have the right to refuse or withdraw from the experiment at any time, without negative consequences and without having to provide any explanation.

Data handling

All data collected during the experiment will be stored anonymously and will only be used for scientific research. You will not be personally identifiable in future publications based on this work, in data files shared with other researchers, or in data repositories. This signed consent form will be kept in a secure locker.

Questions

If you have any questions, please contact one of the researchers.

Please answer the following questions:	**Yes**	**No**
I voluntarily participate in this study	○	○
I have read and understood the information in this document	○	○
I understand the risks involved and accept that I am solely responsible for any accidents that may arise	○	○
I understand that I can withdraw from the study at any time without negative consequences	○	○
I agree that the data collected during the experiment will be used for academic research and may be presented in a MSc thesis, PhD thesis, publication, and public data repository	○	○

Participant name:

Date:

Signature: ___________________

B. Seven exam parts CBR

3.1 Driving away

3.2 Driving on straight and curvy roads

3.3 Behaviour near and around cross sections

3.4 Merging and exiting

3.5 Overtaking and side way movements

3.6 Behaviour near and around special road sections (i.e. bus stop, parking area, roundabouts)

3.7 Special road manoeuvres (i.e. parallel parking, driving reverse).

C. Scenarios overview

Table 11.

Overview of all the scenarios driven during the 21 rides.

Ride number	Title	Description/traits	Result
Ride 1.1	The good driving candidate	Candidate drives carefully, seems a bit slow at times, but it is sufficient	Pass
Ride 1.2	Not adapted / indecisive candidate	3.2 and 3.4: Candidate does not dare to drive at speed on through roads, after warning nevertheless relapse	Fail
Ride 1.3	Candidate drives too fast	3.2 bumps too fast 3.4 being stuck 3.5 overtaking cyclists too fast 3.7 parking too fast	Fail
Ride 1.4	Over casual candidate	3.4 Inserting too early on highway 3.5 overtaking where it is not allowed	Fail
Ride 2.1	Not adapted / indecisive candidate	3.3 Candidate brakes on every cross-section 3.4 too late with looking 3.6 stop at every roundabout and does not show signs of decisiveness	Fail
Ride 2.2	Candidate drives too fast	3.2 Driving too fast 3.3 harsh braking and accelerating 3.6 approaching cross section too fast	Fail
Ride 2.3	The good driving candidate	Candidate drives well, but makes a mistake at 3.1, 3.3, 3.4, 3.5, 3.6 and 3.7 in terms of viewing behaviour. Normal accelerations and speeds driven.	Pass

Ride 3.1	Over casual candidate	3.3 Does not look left 3.6 multiple errors, looking behaviour and road placement 3.7 inappropriate looking behaviour	Fail
Ride 3.2	Candidate struggles with road placement	3.4 too far left on the driving lane 3.4 too close to block lines 3.5 too close around cyclists and parked cars	Fail
Ride 3.3	Inappropriate looking behaviour	3.3 no looking at all once 3.5 no looking at all once	Pass
Ride 3.4	Clumsy candidate	Inappropriate grip on steering wheel 3.3 driving with clutch activated through corners 3.4 Looking to far over shoulder, leading to wrong placement on road	Fail
Ride 3.5	Inappropriate timing, too late with car controls	3.3 Too late with looking on cross sections, too late with the gear shifts 3.4 Too late with mering on highway, too late with lane switch	Fail
Ride 4.1	Not adapted / indecisive candidate	3.3 stops at every busy cross section 3.4 too late with looking 3.6 stops at every roundabout	Fail
Ride 4.2	Clumsy candidate	3.2 Does not shift properly, too far right driving on the road 3.3 forget to use indicator, corners taken too wide	Fail
Ride 5.1	Doubting candidate	3.1 Does not look good, engine stops several times	Fail

		3.2 road placement too far right 3.3 corners too wide	
Ride 5.2	Dangerous driving	3.2 till 3.6 too fast on all sections. Multiple interferences	Fail
Ride 5.3	Candidate has only one large error	3.5 at switching lanes, the candidate does not look at all, happens only once	Pass
Ride 6.1	Doubtful candidate	3.2 too slow 3.4 too slow 3.7 too slow	Fail
Ride 6.2	Doubtful candidate	3.3, 3.4 and 3.6 inappropriate looking behaviour on all sections	Fail
Ride 6.3	Doubtful candidate	3.1 forget to look once 3.3 forget to look once 3.4 candidate does not look in dead corner once	Pass
Ride 6.4	Unpredictable candidate	Drives without proper intentions and misplaces road placement constantly	Fail

D. Scenario categorization

The 21 different rides each have their own scenario which differs from the others based on factors such as errors made and candidate behaviour. To analyse driving behaviour, the drives are categorised based on the scenarios explaining the driving behaviour driven during each of the 21 rides. In the scenarios, the errors made during the ride are listed and give an impression of what mistakes were made during the drive. Using these mistakes written up in the scenario (i.e. inadequate braking, driving too fast over speed bumps), the overall ride can be classified as belonging to one of the four categories.

A categorization in driving style is chosen over a categorization based on pass/fail. This is mainly because of the assumption that a driving exam candidate can fail due to low accelerations (i.e. driving overcautious) or high accelerations (i.e. aggressive). If opting for only discriminating between pass/fail opposed to driving style, the fail group would contain rides with all kinds of acceleration, not necessarily high or low.

The following four categories are used to categorise the driving styles: Aggressive driving behaviour (AGG), desired driving behaviour (DES, overcautious driving behaviour (CAU) and negligent driving behaviour (NGL). In order to properly categorize the rides, the experimenter provided a clear description of each category:

Category AGG aggressive driving

The aggressive driving style can be clearly identified by traits such as high speeds and accelerations. Common events found in this behaviour are harsh braking, harsh lane changing and exceeding the speed limit. The aggressive driver does not brake for road bumps and approaches cross sections with excessive speeds. In terms of the driving exam, the aggressive driver will fail the driving exam at CBR.

Category DES: desired driving

As stated in the title of this category, the desired driver shows good interactions with traffic and blends in with the everyday traffic. In addition, the desired driving also considers safety of other road users and the safety of the passengers within the car. Driving scenarios placed within this category will pass the driving exam at CBR. However, it does not mean that the ride was completely flawless. Errors could have been made during the ride, but the severity was not too much and the errors only occurred a few times.

Category CAU: overcautious driving

The rides placed within this Category CAU drove too slowly and showed errors which hindered other traffic users. Common events occurring within this driving style are stopping at every cross-section or roundabout without any traffic present. The driver from this category doesn't show any signs of insight into traffic, and isn't able to blend in with the other road users. At the driving exam, the person driving with an overcautious driving style will fail.

Category NGL: clumsy/negligent driving

The final category is different from the others, since it is expected that this driving style will not have specific characteristics as the other three in terms of acceleration data. Common traits found in this driving style are not being aware of the surrounding traffic, wrong placement on the road or inappropriate handling of the clutch and gas pedal. At the driving exam, a driver driving with this driving style will fail.

In Table 12, an overview is given of all the rides and the categories they have been placed in by the experimenter who drove along during all of the 21 rides. In the latter two columns, two other people from within the CBR organisation also categorised the rides over the same four categories. Note that the experimenter clustered the rides based on his experience during the ride and the scenarios and in cooperation with the driver (examiner), whilst the CBR staff members only read the prescribed scenario, as listed in Table 12.

Table 12

Driving style categorization based on the driving behaviour listed in each scenario.

Ride	Experimenter	CBR colleague	CBR colleague
1.1	Category DES	Category DES	Category DES
1.2	Category CAU	Category CAU	Category CAU
1.3	Category AGG	Category AGG	Category AGG
1.4	Category NGL	Category AGG	Category NGL
2.1	Category CAU	Category CAU	Category CAU
2.2	Category AGG	Category AGG	Category AGG
2.3	Category DES	Category DES	Category DES
3.1	Category NGL	Category NGL	Category NGL
3.2	Category NGL	Category NGL	Category NGL
3.3	Category DES	Category DES	Category DES
3.4	Category NGL	Category NGL	Category NGL
3.5	Category NGL	Category NGL	Category NGL
4.1	Category CAU	Category CAU	Category CAU
4.2	Category NGL	Category NGL	Category NGL
5.1	Category NGL	Category NGL	Category CAU
5.2	Category AGG	Category AGG	Category AGG
5.3	Category DES	Category DES	Category DES
6.1	Category CAU	Category CAU	Category CAU
6.2	Category NGL	Category NGL	Category CAU
6.3	Category DES	Category DES	Category DES
6.4	Category NGL	Category NGL	Category NGL

Overall, the three different participants were unanimous on 76.2% of all the scenarios and their corresponding category. Taking into account the experimenter's extra information, a 100% unanimity was identified with either of the two other CBR staff members. Based on this categorising, the overall categories would then contain the following rides:

- **Category AGG:** 1.3, 2.2 and 5.2
- **Category DES:** 1.1, 2.3, 3,3, 5.3 and 6.3
- **Category CAU:** 1.2, 2.1, 4.1 and 6.1
- **Category NGL:** 1.4, 3.1, 3.2, 3.4, 3.5, 4.2, 5.1, 6.2 and 6.4

Categorising the drives based on the text written in the scenario shows promising results, since all three people subjected all drives to the same category (76.2%). However, this categorising approach is purely based on the interpretation of the text written in the scenario and not on any of the acceleration data. On the other hand, the categorization done using the acceleration data as main input created very blurred and indecisive categories. This is mostly because the three acceleration directions measured all showed different properties per ride. In other words, the acceleration signal of for example ride 2.3 placed the ride in different categories, using different acceleration data as input (lateral and longitudinal).

E. Driving event registration

Using the event recorder, four types of events have been registered: turns, speed increases, speed decreases and lane changes. The event registration was based on the experimenter's own interpretation of the traffic situation from the backseat of the vehicle and the driving event. In this appendix, three snapshots have been given per event, where the video recording and location indicate the exact moment the experimenter registered the event.

E.1 Turns

Figure 23. Turn event registered. Ride 1.3 in 30 km/h speed zone in Leusden, NL.

Figure 24. Turn event registered. Ride 5.2 in 50 km/h speed zone in Achterberg, NL.

Figure 25. Turn event registered. Ride 3.1 in 30 km/h speed zone in Leusden, NL.

E.2 Speed increase

Figure 26. Speed increase event registered. Ride 2.3 in 50 km/h speed zone in Leusden, NL.

Figure 27. Speed increase event registered. Ride 3.4 in 30 km/h speed zone in Leusden, NL.

Figure 28. Speed increase event registered. Ride 3.4 in 30 km/h speed zone in Leusden, NL.

E.3 Speed decrease

Figure 29. Speed decrease event registered. Ride 2.3 in 30 km/h speed zone in Amersfoort, NL.

Figure 30. Speed decrease event registered. Ride 2.3 in 50 km/h speed zone in Amersfoort, NL.

Figure 31. Speed decrease event registered. Ride 3.4 in 50 km/h speed zone in Leusden, NL.

E.4 Lane changes

Figure 32. Lane change even registered. Ride 2.3 in 100 km/h speed zone in Leusden, NL.

Figure 33. Lane change event registered. Ride 3.4 in 50 km/h speed zone in Leusden, NL.

Figure 34. Lane change event registered. Ride 5.3 in 50 km/h speed zone in Amersfoort, NL.

F. Full ride analysis

In Table 13, an overview is given of the mean absolute amplitude (*M*) and the absolute amplitude standard deviation (*SD*) of each ride.

Table 13

Descriptive statistics of the amplitude measured during each ride including mean (*M*) and standard deviation (*SD*) for lateral (Lat), longitudinal (Lon) and gravitational (Gra) accelerations.

Ride ID	*M*			*SD*		
	Lat	Lon	Gra	Lat	Lon	Gra
Ride 1.1	0,044	0,046	0,039	0,063	0,067	0,063
Ride 1.2	0,184	0,085	0,097	0,245	0,120	0,140
Ride 1.3	0,087	0,087	0,087	0,079	0,097	0,096
Ride 1.4	0,062	0,061	0,048	0,047	0,059	0,061
Ride 2.1	0,472	0,113	0,729	0,536	0,150	0,782
Ride 2.2	0,063	0,070	0,062	0,084	0,084	0,106
Ride 2.3	0,050	0,070	0,058	0,066	0,077	0,080
Ride 3.1	0,057	0,047	0,040	0,088	0,093	0,079
Ride 3.2	0,169	0,175	0,139	0,229	0,212	0,170
Ride 3.3	0,039	0,063	0,068	0,068	0,062	0,095
Ride 3.4	0,040	0,037	0,068	0,068	0,052	0,096
Ride 3.5	0,037	0,047	0,075	0,064	0,068	0,083
Ride 4.1	0,040	0,070	0,031	0,069	0,061	0,054
Ride 4.2	0,007	0,024	0,011	0,009	0,025	0,013
Ride 5.1	0,043	0,067	0,051	0,076	0,093	0,074
Ride 5.2	0,046	0,075	0,060	0,073	0,109	0,091
Ride 5.3	0,042	0,064	0,049	0,073	0,089	0,069
Ride 6.1	0,033	0,038	0,049	0,057	0,053	0,068
Ride 6.2	0,038	0,042	0,058	0,067	0,061	0,082
Ride 6.3	0,042	0,048	0,064	0,076	0,069	0,088
Ride 6.4	0,048	0,045	0,065	0,074	0,062	0,087

G. Event analysis

G.1 Test for normality

In Figure 35, the absolute amplitude of all four different events is shown, together with a normal distribution fit based on the mean and standard deviation of the amplitudes per acceleration direction.

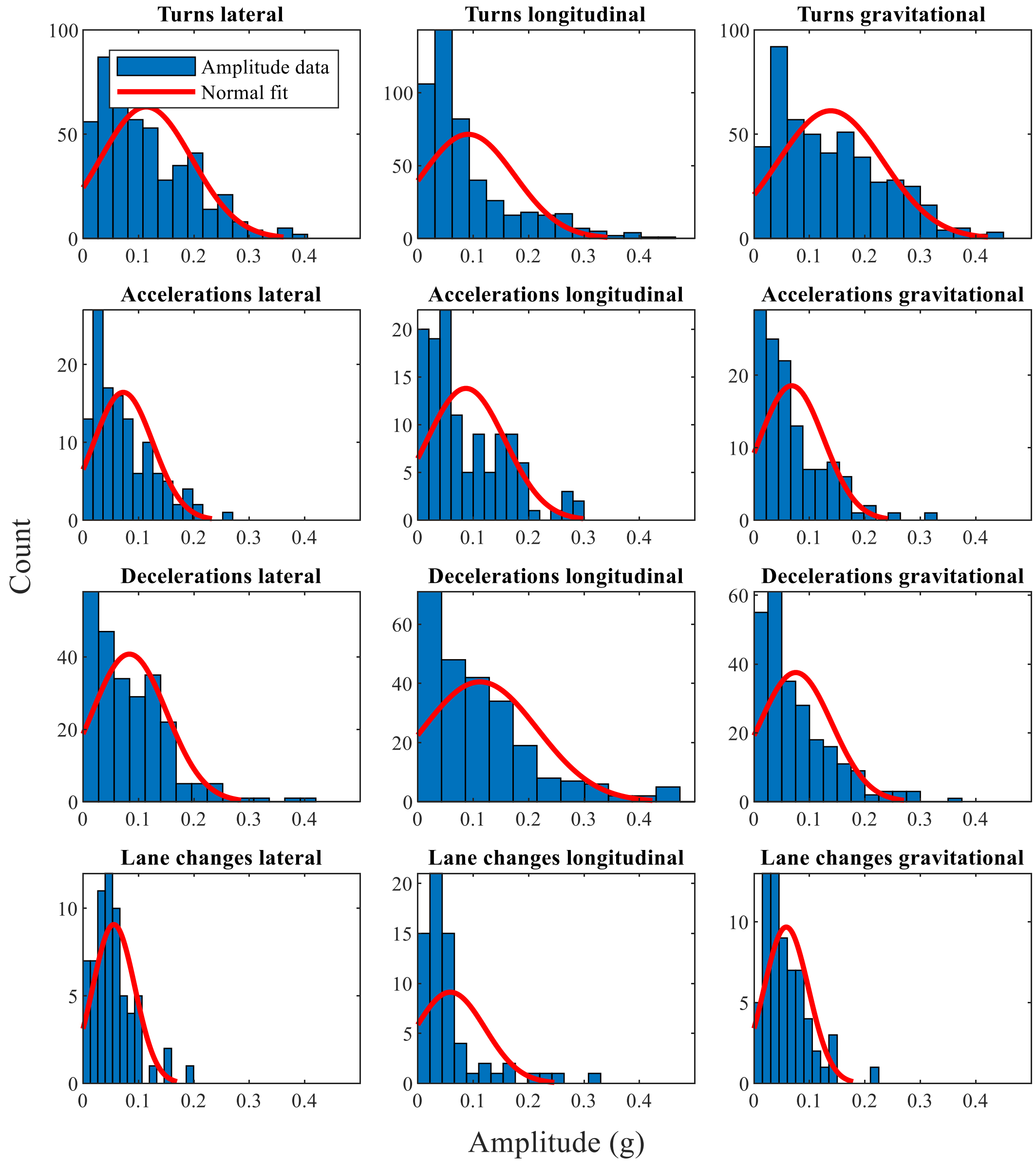

Figure 35. Histograms of the amplitudes for each event in all speed zones. Fitted with normal distribution based on mean and standard deviation of each group.

From Figure 35, it can be observed that the amplitude data of the events don't show a fit with the normal distribution. Using a Kolmogorov Smirnov test, it has been found that no samples available in the dataset come from a normal distribution.

G.2 Descriptive statistics

Table 14

Descriptive statistics of the amplitude for all turn events registered.

Turns		30 km/h					50 km/h					80 km/h					100 km/h				
		M	*SD*	*Max*	*Min*	*n*	*M*	*SD*	*Max*	*Min*	*n*	*M*	*SD*	*Max*	*Min*	*n*	*M*	*SD*	*Max*	*Min*	*n*
AGG	Lat	0,26	0,12	0,50	0,06	25	0,25	0,10	0,48	0,12	36	0,26	0,30	0,47	0,05	2	0,23	0,11	0,33	0,06	5
	Lon	0,17	0,09	0,33	0,05	25	0,18	0,11	0,48	0,04	36	0,17	0,16	0,28	0,06	2	0,19	0,11	0,28	0,04	5
	Gra	0,15	0,14	0,52	0,00	25	0,18	0,12	0,46	0,00	36	0,07	0,05	0,10	0,03	2	0,08	0,03	0,12	0,05	5
DES	Lat	0,14	0,08	0,51	0,01	67	0,16	0,09	0,39	0,03	50	0,17	0,16	0,35	0,06	3	0,17	0,11	0,31	0,04	4
	Lon	0,11	0,07	0,44	0,01	67	0,12	0,07	0,33	0,03	50	0,08	0,02	0,10	0,05	3	0,08	0,02	0,11	0,05	4
	Gra	0,10	0,08	0,34	0,01	67	0,13	0,10	0,38	0,00	50	0,07	0,05	0,13	0,03	3	0,17	0,16	0,35	0,03	4
CAU	Lat	0,10	0,05	0,21	0,03	28	0,10	0,06	0,30	0,05	27	-	-	-	-	0	0,21	0,00	0,21	0,21	1
	Lon	0,09	0,04	0,19	0,04	28	0,10	0,07	0,39	0,03	27	-	-	-	-	0	0,17	0,00	0,17	0,17	1
	Gra	0,11	0,06	0,31	0,02	28	0,17	0,11	0,44	0,02	27	-	-	-	-	0	0,22	0,00	0,22	0,22	1
NGL	Lat	0,10	0,07	0,32	0,01	108	0,14	0,11	0,79	0,00	154	0,11	0,12	0,31	0,02	5	0,12	0,10	0,25	0,03	6
	Lon	0,09	0,08	0,40	0,00	108	0,09	0,08	0,46	0,00	154	0,07	0,05	0,13	0,00	5	0,05	0,01	0,07	0,04	6
	Gra	0,07	0,06	0,30	0,00	108	0,09	0,10	0,44	0,00	154	0,06	0,06	0,16	0,02	5	0,10	0,09	0,25	0,02	6
All	Lat	0,13	0,09	0,51	0,01	228	0,15	0,11	0,79	0,00	267	0,16	0,16	0,47	0,02	10	0,17	0,11	0,33	0,03	16
	Lon	0,10	0,08	0,44	0,00	228	0,11	0,09	0,48	0,00	267	0,09	0,08	0,28	0,00	10	0,11	0,08	0,28	0,04	16
	Gra	0,06	0,11	0,52	0,00	228	0,12	0,11	0,46	0,00	267	0,07	0,05	0,16	0,02	10	0,12	0,10	0,36	0,02	16

Table 15

Descriptive statistics of the amplitude for all speed increase events registered.

Speed increases		30 km/h					50 km/h					80 km/h					100 km/h				
		Mu	**SD**	**Max**	**Min**	**n**	**M**	**SD**	**Max**	**Min**	**n**	**M**	**SD**	**Max**	**Min**	**n**	**M**	**SD**	**Max**	**Min**	**n**
AGG	Lat	0,11	0,03	0,15	0,08	5	0,15	0,08	0,33	0,05	15	0,23	0,10	0,41	0,15	6	0,12	0,06	0,19	0,07	3
	Lon	0,11	0,06	0,19	0,05	5	0,17	0,08	0,36	0,06	15	0,18	0,08	0,35	0,12	6	0,15	0,06	0,19	0,07	3
	Gra	0,12	0,05	0,20	0,07	5	0,11	0,09	0,26	0,01	15	0,18	0,12	0,34	0,05	6	0,14	0,02	0,16	0,12	3
DES	Lat	0,11	0,05	0,19	0,05	8	0,13	0,07	0,21	0,04	4	0,12	0,01	0,12	0,11	3	0,10	0,08	0,19	0,03	3
	Lon	0,15	0,10	0,36	0,02	8	0,09	0,04	0,13	0,04	4	0,16	0,09	0,24	0,07	3	0,13	0,08	0,22	0,06	3
	Gra	0,07	0,04	0,13	0,02	8	0,05	0,02	0,07	0,02	4	0,01	0,01	0,02	0,01	3	0,10	0,09	0,19	0,02	3
CAU	Lat	0,08	0,04	0,12	0,05	4	0,10	0,04	0,14	0,05	5	-	-	-	-	0	0,14	0,04	0,18	0,10	3
	Lon	0,08	0,02	0,10	0,06	4	0,09	0,04	0,14	0,06	5	-	-	-	-	0	0,13	0,07	0,21	0,07	3
	Gra	0,08	0,06	0,17	0,04	4	0,05	0,03	0,07	0,02	5	-	-	-	-	0	0,16	0,23	0,42	0,02	3
NGL	Lat	0,09	0,08	0,30	0,01	19	0,07	0,06	0,24	0,00	42	-	-	-	-	0	0,07	0,05	0,14	0,02	6
	Lon	0,09	0,07	0,22	0,01	19	0,06	0,05	0,25	0,00	42	-	-	-	-	0	0,07	0,04	0,13	0,04	6
	Gra	0,08	0,11	0,38	0,00	19	0,07	0,07	0,29	0,00	42	-	-	-	-	0	0,05	0,05	0,14	0,00	6
All	Lat	0,10	0,06	0,30	0,01	36	0,10	0,07	0,33	0,00	66	0,14	0,07	0,33	0,05	9	0,13	0,08	0,33	0,02	15
	Lon	0,10	0,08	0,36	0,01	36	0,09	0,07	0,36	0,00	66	0,17	0,08	0,35	0,07	9	0,11	0,06	0,22	0,04	15
	Gra	0,08	0,09	0,38	0,00	36	0,07	0,07	0,29	0,00	66	0,13	0,13	0,34	0,01	9	0,10	0,11	0,43	0,00	15

Table 16

Descriptive statistics of the amplitude for all speed decrease events registered.

Speed decreases		**30 km/h**					**50 km/h**					**80 km/h**					**100 km/h**				
		M	**SD**	**Max**	**Min**	**n**	**M**	**SD**	**Max**	**Min**	**n**	**M**	**SD**	**Max**	**Min**	**n**	**M**	**SD**	**Max**	**Min**	**n**
AGG	Lat	0,16	0,07	0,26	0,02	18	0,18	0,10	0,42	0,01	34	0,29	0,33	0,53	0,05	2	0,14	0,10	0,25	0,04	4
	Lon	0,20	0,11	0,45	0,03	18	0,23	0,14	0,71	0,01	34	0,12	0,11	0,20	0,05	2	0,11	0,10	0,25	0,02	4
	Gra	0,10	0,12	0,42	0,00	18	0,12	0,10	0,34	0,00	34	0,06	0,02	0,07	0,05	2	0,05	0,03	0,08	0,01	4
NGL	Lat	0,12	0,09	0,29	0,01	13	0,09	0,07	0,27	0,01	22	0,09	0,03	0,13	0,06	5	0,14	0,09	0,28	0,01	7
	Lon	0,11	0,05	0,20	0,01	13	0,15	0,10	0,45	0,01	22	0,18	0,07	0,27	0,10	5	0,14	0,09	0,26	0,03	7
	Gra	0,07	0,05	0,16	0,00	13	0,06	0,06	0,23	0,00	22	0,08	0,05	0,15	0,03	5	0,10	0,09	0,29	0,02	7
CAU	Lat	0,11	0,11	0,27	0,01	5	0,11	0,07	0,29	0,01	18	-	-	-	-	0	0,18	0,00	0,18	0,18	1
	Lon	0,09	0,09	0,22	0,02	5	0,10	0,08	0,38	0,01	18	-	-	-	-	0	0,20	0,00	0,20	0,20	1
	Gra	0,06	0,07	0,16	0,00	5	0,03	0,03	0,10	0,00	18	-	-	-	-	0	0,01	0,00	0,01	0,01	1
NGL	Lat	0,07	0,06	0,25	0,00	34	0,08	0,07	0,35	0,00	87	0,05	0,04	0,10	0,01	4	0,12	0,09	0,25	0,07	4
	Lon	0,07	0,06	0,26	0,01	34	0,08	0,09	0,40	0,00	87	0,03	0,02	0,06	0,02	4	0,08	0,02	0,11	0,06	4
	Gra	0,07	0,07	0,27	0,00	34	0,05	0,05	0,22	0,00	87	0,03	0,03	0,07	0,00	4	0,10	0,08	0,18	0,00	4
All	Lat	0,11	0,08	0,29	0,00	70	0,11	0,08	0,42	0,00	161	0,11	0,14	0,53	0,12	11	0,14	0,08	0,28	0,11	16
	Lon	0,11	0,09	0,45	0,01	70	0,12	0,12	0,71	0,00	161	0,12	0,89	0,27	0,02	11	0,12	0,08	0,26	0,17	16
	Gra	0,07	0,08	0,42	0,00	70	0,06	0,07	0,34	0,00	161	0,06	0,05	0,15	0,00	11	0,08	0,07	0,29	0,00	16

Table 17

Descriptive statistics of the amplitude for all lane change events registered.

Lane changes		**30 km/h**					**50 km/h**					**80 km/h**					**100 km/h**				
		M	**SD**	**Max**	**Min**	**n**	**M**	**SD**	**Max**	**Min**	**n**	**M**	**SD**	**Max**	**Min**	**n**	**M**	**SD**	**Max**	**Min**	**n**
AGG	Lat	0,10	0,02	0,13	0,08	4	0,12	0,05	0,20	0,07	7	0,17	0,06	0,21	0,12	2	-	-	-	-	0
	Lon	0,12	0,04	0,17	0,07	4	0,09	0,03	0,13	0,05	7	0,13	0,01	0,13	0,12	2	-	-	-	-	0
	Gra	0,14	0,09	0,27	0,07	4	0,08	0,06	0,16	0,01	7	0,14	0,12	0,22	0,06	2	-	-	-	-	0
DES	Lat	-	-	-	-	0	0,09	0,05	0,21	0,03	14	-	-	-	-	0	0,07	0,03	0,13	0,03	13
	Lon	-	-	-	-	0	0,08	0,04	0,16	0,03	14	-	-	-	-	0	0,08	0,05	0,17	0,03	13
	Gra	-	-	-	-	0	0,06	0,04	0,18	0,01	14	-	-	-	-	0	0,06	0,04	0,11	0,00	13
CAU	Lat	0,08	0,05	0,17	0,05	5	0,09	0,07	0,19	0,05	4	-	-	-	-	0	0,06	0,01	0,07	0,05	2
	Lon	0,06	0,04	0,13	0,04	5	0,09	0,07	0,19	0,05	4	-	-	-	-	0	0,05	0,00	0,05	0,05	2
	Gra	0,08	0,07	0,19	0,03	5	0,05	0,03	0,10	0,02	4	-	-	-	-	0	0,07	0,01	0,08	0,07	2
NGL	Lat	-	-	-	-	0	0,06	0,05	0,18	0,01	20	-	-	-	-	0	0,07	0,07	0,33	0,00	23
	Lon	-	-	-	-	0	0,07	0,07	0,34	0,00	20	-	-	-	-	0	0,06	0,04	0,13	0,01	23
	Gra	-	-	-	-	0	0,04	0,03	0,09	0,00	20	-	-	-	-	0	0,04	0,03	0,14	0,00	23
All	Lat	0,09	0,38	0,16	0,05	9	0,08	0,05	0,21	0,01	45	0,17	0,17	0,21	0,12	2	0,07	0,05	0,33	0,00	38
	Lon	0,08	0,04	0,17	0,04	9	0,08	0,06	0,34	0,00	45	0,13	0,13	0,13	0,12	2	0,69	0,04	0,17	0,01	38
	Gra	0,10	0,08	0,27	0,30	9	0,05	0,04	0,18	0,00	45	0,14	0,14	0,22	0,06	2	0,05	0,03	0,14	0,00	38