

# Understanding driver planning behaviour when overtaking a bicyclist: Time to collision estimations from naturalistic driving data

Master's thesis in Mechanical Engineering

SABINO MASTRANDREA



MASTER'S THESIS IN MECHANICAL ENGINEERING

Understanding driver planning behaviour when  
overtaking a bicyclist: Time to collision estimations from  
naturalistic driving data

SABINO MASTRANDREA

Department of Mechanics and Maritime Sciences  
Division of Vehicle safety  
Road-user Vehicle Interaction group  
CHALMERS UNIVERSITY OF TECHNOLOGY  
Göteborg, Sweden 2019

Understanding driver planning behaviour when overtaking a bicyclist: Time to collision estimations from naturalistic driving data  
SABINO MASTRANDREA

© SABINO MASTRANDREA, 2019-06-15

Master's Thesis 2019:64  
Department of Mechanics and Maritime Sciences  
Division of Vehicle safety  
Road-user Vehicle Interaction group  
Chalmers University of Technology  
SE-412 96 Göteborg  
Sweden  
Telephone: + 46 (0)31-772 1000

Chalmers reproservice/ Department of Mechanics and Maritime Sciences  
Göteborg, Sweden 2019-06-15

Understanding driver planning behaviour when overtaking a bicyclist: Time to collision estimations from naturalistic driving data  
Master's thesis in Mechanical Engineering  
SABINO MASTRANDREA  
Department of Mechanics and Maritime Sciences  
Division of Vehicle safety  
Road-user Vehicle Interaction group  
Chalmers University of Technology

## Abstract

In Europe, bicyclist road fatalities have increased for the last ten years. Active safety systems such as Automated Emergency Braking can give a considerable benefit in reducing bicyclist road fatalities, but knowledge on when they should intervene is still limited. Estimation of Time to Collision (TTC) when the driver starts planning to overtake (planning point) could help in improving active safety systems, since TTC is a good estimate of the available mitigation time for the algorithms to intervene. This thesis, carried out using naturalistic driving data from the UDRIVE project, mainly consisted in extracting data from bicyclist overtaking scenarios on rural road and modelling TTC and longitudinal distance to quantify how they are influenced by different factors. The result of the estimation is that the presence of oncoming traffic and an increase in bicyclist lateral distance caused a decrease of both TTC and longitudinal distance at the planning point. Moreover, male drivers showed higher TTC at the planning point than female drivers. Interestingly, the planning point was not affected by the overtaking strategy.

## Key words:

Bicyclist, traffic safety, naturalistic driving data, decision-making

# Contents

Abstract .....	I
Contents .....	II
Preface.....	V
Aknowledgements.....	V
Notations .....	VI
1 Introduction.....	1
1.1 Problem definition.....	1
2 Literature review .....	2
2.1 Definition of overtaking .....	2
2.2 Planning to overtake.....	3
3 Methodology .....	4
3.1 Naturalistic Driving Data .....	4
3.2 Raw data used.....	5
3.3 Data Reduction.....	5
3.3.1 Segments creation .....	7
3.3.2 Segments validation .....	7
3.4 Manual annotation of validated segments .....	8
3.4.1 Overtaking details .....	9
3.4.2 Planning point .....	9
3.4.3 Bicyclist details and road inclination .....	10
3.4.4 Quantification of the delay between video and signals .....	11
3.5 Data extraction .....	11
3.5.1 Driver information and country .....	12
3.5.2 Rotation of MobilEye data.....	12
3.5.3 Bicyclist speed .....	14
3.5.4 Bicyclist longitudinal distance .....	14
3.5.5 Bicyclist lateral distance at the planning point .....	15
3.5.6 Time to Collision .....	16
3.6 Generalised Linear Mixed Model (GLMM) .....	17
3.6.1 Factors and responses .....	18
3.6.2 Implementation in MATLAB® .....	19
4 Results.....	20
4.1 Analysed overtaking manoeuvres .....	20
4.2 TTC at the planning point .....	21

4.3	Longitudinal distance at the planning point .....	26
4.4	TTC comparison between planning point and steer away point .....	32
4.5	GLMM results .....	33
5	Discussion .....	37
5.1	When is the planning point? .....	37
5.2	Qualitative analysis .....	37
5.3	GLMM .....	38
5.3.1	Lateral distance .....	39
5.3.2	Bicyclist speed .....	39
5.3.3	Ego vehicle speed .....	39
5.3.4	Presence of oncoming traffic .....	40
5.3.5	Overtaking strategy .....	40
5.3.6	Gender .....	40
5.3.7	Age .....	40
5.4	Limitations .....	41
5.4.1	Manual annotation .....	41
5.4.2	Data extraction .....	41
6	Conclusion and future works .....	43
6.1	Future works .....	43
7	Bibliography .....	44
	Appendices .....	46
A	Time to Collision at the planning point – additional figures .....	47
B	Longitudinal distance at the planning point – additional figures .....	49



## **Preface**

This thesis is the final project for the Master's Degree in Mechanical Engineering at Politecnico di Milano, Italy, performed during an Erasmus+ exchange programme at Chalmers University of Technology, Göteborg, Sweden. The thesis took place in SAFER - Vehicle and Traffic Safety Centre at Chalmers, located in Göteborg, and partially in Veoneer Research, located in Vårgårda.

## **Aknowledgements**

This thesis was carried out with Marco Dozza (Chalmers) as examiner and Jordanka Kovaceva (Chalmers), Alessia Knauss (Veoneer) and Prateek Thalya (Veoneer) as supervisors, to whom I am extremely thankful for giving me the opportunity to work with them and for supporting me with their precious help and their positive attitude. Thank you to Erik Svanberg for helping me with the SALSA interface, to Alexander Rasch for the useful inputs about the statistical analysis and to all the personal in SAFER, for making it an extremely nice and pleasant place where to learn and work. I am thankful to professor Federico Cheli and to Stefano Arrigoni for being my supervisors at Politecnico di Milano and to professor Mauro Filippini and the International Mobility Office at Politecnico di Milano for giving me the great opportunity to spend one year in Sweden.

I am very thankful to my family and to all my friends.

Göteborg June 2019-06-15

SABINO MASTRANDREA

# Notations

## Abbreviations

AEB	Automated Emergency Braking
CAN	Controller Area Network
FoV	Field of View
GLMM	Generalised Linear Mixed Model
GPS	Global Positioning System
NDD	Naturalistic Driving Data
TTC	Time To Collision
UDRIVE	eUropean naturalistic Driving and Riding for Infrastructure & Vehicle safety and Environment

## Symbols – Manual annotation and data extraction

$d$	Bicyclist longitudinal distance
$d_{PP}$	Longitudinal distance at the planning point
$dL_b$	Bicyclist lane position
$dL_{ego}$	Ego vehicle lane position
$RV$	Relative Speed
$T$	Time axis
$T_{begin}$	Beginning of the segment
$T_{end}$	End of the segment
$T_{PP}$	Planning point
$T_{SP}$	Steer away point
$T_s$	Time stamp in which ego vehicle and bicyclist are travelling at the same speed
$T_0$	Time stamp in which the bicyclist disappears from MobilEye
$T_x$	Time stamp in which the bicyclist appears from MobilEye
$TTC$	Time to collision
$TTC_{PP}$	Time to collision at the planning point
$TTC_{SP}$	Time to collision at the steer away point
$V_b$	Bicyclist speed
$V_{ego}$	Ego vehicle speed
$\tilde{x}$	Bicyclist longitudinal distance before bicyclist appears
$x_{ME}$	Bicyclist longitudinal distance from Mobileye
$y_{ME}$	Bicyclist lateral distance from MobilEye
$y_{PP}$	Bicyclist lateral distance at the planning point
$y_{SP}$	Bicyclist lateral distance at the steer away point
$\tau$	Auxiliary integration variable
$\psi$	Heading angle

## Symbols – Coordinate systems

$p_{ME}$	Generic position vector
$\mathbf{R}$	Rotation matrix
$x$	Car-fixed longitudinal coordinate
$y$	Car-fixed lateral coordinate
$x'$	Local longitudinal coordinate
$y'$	Local lateral coordinate

## Symbols – Generalised Linear Mixed Model

$AIC$	Akaike Information Criterion
$g(\cdot)$	Link function
$R_{adj}^2$	Adjusted coefficient of correlation
$R_{ord}^2$	Ordinary coefficient of correlation
$u$	Random effects vector
$w$	Weight
$x$	Observable factor
$\mathbf{X}$	Fixed effects design matrix
$y$	Response
$\mathbf{Z}$	Random effects design matrix
$\beta$	Fixed effects vector
$\epsilon$	Model offset
$\mu$	Mean
$\eta$	Linear predictor
$\sigma$	Standard deviation



# 1 Introduction

Bicycles are a common mode of transportation, either for commuting or for sports (Pucher, J., et al. , 2011). Bicyclists are classified as Vulnerable Road Users (VRUs) because of the limited amount of protection they have, compared to drivers in motor vehicles (SWOV, 2012). Bicyclists are also bounded to look straight not to fall and do not have rear mirrors (Nero, 2017). In the United States, the number of bicyclist road fatalities is estimated to be twelve times higher than the road fatalities involving drivers in motor vehicles (Pucher, J., Dijkstra, L. , 2003). A recent study in Europe has shown that bicyclist road fatalities increased from 6% (2007) to 8% (2016) of all road fatalities in Europe (European Commission, June 2018). Accidents in which the driver approaches the bicyclist from behind are the most likely to be fatal or to result in severe injuries (Feng, F., Bao, S., 2018) and are more likely to occur on rural roads rather than in urban areas (Boufous, S., et al., 2012). Moreover, the high speed of motor vehicles compared to bicyclists make overtaking manoeuvres particularly dangerous (Dozza, M., et al., 2015). More details about how an overtaking is defined is given in Section 2.1.

Active safety systems such as Automated Emergency Braking (AEB) can have a strong influence in bicyclist road injuries or fatalities reduction, given that most of the collisions between bicyclists and motor vehicles is due to driver or bicyclist inattention (Räsänen, M., Summala, H., 1998) (The Royal Society for the Prevention of Accidents, 2017). To date, Bicyclist-AEB efficiency is limited due to its high conservativeness, which implies activation only when the scenario is considerably dangerous (Duan, J., et al., 2017). Moreover, the higher speed of bicyclists compared to pedestrians makes them difficult VRUs to recognize: the collision is not always avoided in EuroNCAP tests (EuroNCAP, 2018).

Estimation of Time to Collision (TTC) at the planning point during an overtaking of a bicyclist can contribute in improving the existing active safety systems, since it is an estimation of the time available for the AEB algorithm to intervene and mitigate or avoid a collision.

The aim of this thesis is to provide a method to identify bicyclist overtaking manoeuvres on rural road, to define and identify the planning point, to calculate TTC and longitudinal distance at the planning point and to create a model of the extracted TTC and longitudinal distance to quantify the influence of external factors. The long-term objective is to contribute in the improvement of the existing active safety systems in terms of bicyclist road injuries and fatalities reduction. The thesis is based on Naturalistic Driving Data (NDD) from the eUropean naturalistic Driving and Riding for Infrastructure & Vehicle safety and Environment (UDRIVE) project.

## 1.1 Problem definition

- When is the planning point?
- Which signals and information are needed for the estimation of TTC at the planning point?
- Which factors significantly influence the TTC at the planning point?

## 2 Literature review

In Section 2.1, the definition of overtaking is provided. Section 2.2 provides a description of how drivers plan to overtake.

### 2.1 Definition of overtaking

An overtaking is defined in the Collins dictionary as “an act or the process of moving past another vehicle or person travelling in the same direction” (Collins Dictionary, 2019). An overtaking can be described with higher level of detail by defining a strategy and by dividing it into different phases.

Firstly, an overtaking must be characterized by a certain strategy, that defines how the overtaking is performed. In a previous study (Hegeman, G., et al., 2005), the overtaking strategy was defined as follows:

- Accelerative - the driver slows down and follows the leading road user for a while before passing.
- Flying - the driver overtakes the leading road user keeping a relatively constant speed.
- Piggy backing - the driver is following its leading vehicle in a row, while the leading vehicle is performing an overtaking.
- 2+ - the driver overtakes more than one vehicle in the same manoeuvre.

Secondly, an overtaking is not a punctual event, but it has a certain duration that allows to divide it in different phases and to potentially analyse each of them separately. The definition used in this thesis is the one provided by Marco Dozza et al. (Dozza, M., et al., 2015), to be consistent with the definition adopted in previous works (Schindler, R., Bast, V., 2015) (Kovaceva, J., et al., 2018) (Nero, 2017) (Rasch, 2018) (Panero, 2018).

- Phase 1, also called *approaching phase*, is the phase in which the driver approaches the road user from behind.
- Phase 2, also called *steering away phase*, is the phase in which the driver diverges from its current lane position, to avoid the collision. The *steer away point* is the point in time in which the driver starts diverging by steering.
- Phase 3, also called *passing phase*, is the phase in which the driver is inside the *passing zone*: this zone is defined from two meters behind to two meters ahead of the passed road user.
- Phase 4, also called *returning phase*, is the phase in which the driver exits the *passing zone* and returns the same lane position as before performing the overtaking.

The phases are illustrated in Figure 1.

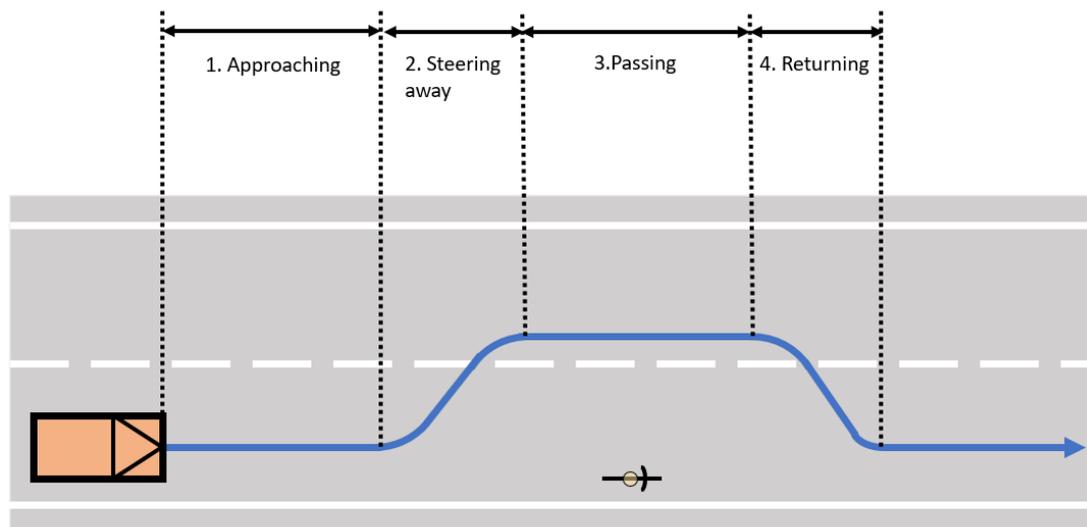


Figure 1 Different phases of the overtaking (M. Dozza et al., 2015).

## 2.2 Planning to overtake

When any moving object is perceived (*perception*), the driver starts *planning* and, then, decides an *action* to perform (Zago, M., et al., 2008). Specifically, in describing the decision-making process in bicyclist overtaking manoeuvres, *perception* indicates the moment in which the bicyclist is perceived by the driver, *planning* consists in deciding if the overtaking should be performed or not and *action* consists in preparing to overtake by, for instance, pressing the gas pedal, activating the turn indicator or steering away. The focus in this thesis is *planning*: according to previous studies (Hegeman, G., et al., 2005), when the driver is planning to overtake, the following questions need to be answered:

- Is it needed to overtake? This mainly depends on behavioural factors, as well as on the ego vehicle speed.
- Is it allowed to overtake? The driver checks the surroundings, looks the road signals or the lane markings to understand if overtaking is allowed by the regulations.
- Is there the opportunity to overtake? The driver wants to know if there is any possible danger in performing the overtaking. In particular, the driver predicts the path to travel during the manoeuvre and assesses whether it is dangerous or not.

The planning point is defined as the moment when *planning* starts. The decision to overtake is made when all the three questions above have an affirmative answer, and is followed by an *action*, e.g. activating the turn indicator or steering away.

## 3 Methodology

In this chapter, the methodology is provided. Section 3.1 is a brief introduction to NDD and the tools used in this thesis. Section 3.2 is a list of the raw data used. Section 3.3 provides the procedure used to reduce data before performing manual annotation, explained in Section 3.4, and data extraction, explained in Section 3.5. Finally, Generalised Linear Mixed Models are used to model the TTC and the longitudinal distance at the planning point, as explained in Section 3.6.

### 3.1 Naturalistic Driving Data

The study is performed using NDD, which are data collected during real traffic conditions, using instrumented vehicles. The strength of NDD is that drivers are performing their daily activities while data is collected, and the influence of the measurement systems installed in the vehicle (already made less obtrusive as possible) should not affect the driver's behaviour. UDRIVE is a project that involved 120 cars, developed in seven European countries, that counts more than 45000 total car driving hours (Castermans, J., 2017).

Vehicles were instrumented with a Data Acquisition System (DAS) that collects data from the ego vehicle's CAN bus, a GPS sensor, video cameras mounted in the driver compartment, and MobilEye. Mobileye is a smart camera that provides data describing up to four road users, when they are in its Field of View (FoV), based on computer vision. The four sources have different sampling frequencies. The interface that allowed to handle NDD was SALSA, Smart Automation for Large data Sets Analysis.

The following is a summary of how the database is structured:

- Full record - full trip done by a driver, characterized by attributes and time series.
- Segment - portion of record defined from a begin time to an end time, characterized by attributes and time series.
- Attribute - property defined by the user via manual annotation, applicable to either a full record or a segment.
- Time series – collection of all the signals synchronized on the same time vector. For instance, all the signals from the CAN bus are grouped in one time series, signals from MobilEye are collected in another time series (different source implies different sampling frequency and, consequently, different time vector). A time series can also be defined by the user.
- Signals – vectors of time dependent data.
- Node – scripts used to handle data in records or in segments. Nodes can be used on the full records to generate segments, or to create new user-defined time series and signals.

## 3.2 Raw data used

Table 1 provides a list of the raw data used in this thesis and the corresponding source (either CAN bus, MobilEye, GPS sensor or video feed, as in Section 3.1).

*Table 1 Raw data used, with corresponding source.*

Source	Variable
CAN bus	Accelerator pedal position [arbitrary]
	Brake pedal status [categorical]
	Ego vehicle speed [ <i>km/h</i> ]
	Steering wheel angle [ <i>deg</i> ]
	Turn indicator status [categorical]
	Yaw Rate [ <i>deg/s</i> ]
MobilEye	Lane position – left marking/edge distance [ <i>m</i> ]
	Lane position – right marking/edge distance [ <i>m</i> ]
	Road user lateral distance [ <i>m</i> ]
	Road user longitudinal distance [ <i>m</i> ]
	Road user relative speed [ <i>m/s</i> ]
	Road user type [categorical]
GPS	Road type [categorical]
Video feed	Frontal and lateral views; driver's face, cabin and pedals views

## 3.3 Data Reduction

The events of interest for this study are bicyclist overtaking manoeuvres on rural road, either accelerative or flying (see Section 2.1), which are a minor part of the extensive amount of driving hours available within the UDRIVE project. Hence, data reduction was needed first. A filtering algorithm was applied to all the records to create overtaking segments, as described in Section 3.3.1. The algorithm created 8623 segments. An ideal filter should be able to detect all the overtaking manoeuvres (no false negatives) with 100% hit rate (no false positives). Since the perfect filter cannot be coded, the segments created by the filtering algorithm were subsequently validated via manual annotation as in Section 3.3.2. The segments reviewed for validation were roughly 500 out of the 8623 total segments. The rest of the analysis (starting from Section 3.4) was performed on the validated segments only. After manual annotation of validated segments and data extraction, a quality check of the extracted data allowed to remove segments before the statistical analysis. The procedure is illustrated in Figure 2.

Figure 3 qualitatively illustrates the relation among all the created segments, the validated segments and the real subset of accelerative or flying bicyclist overtaking manoeuvres on rural roads within the full database.

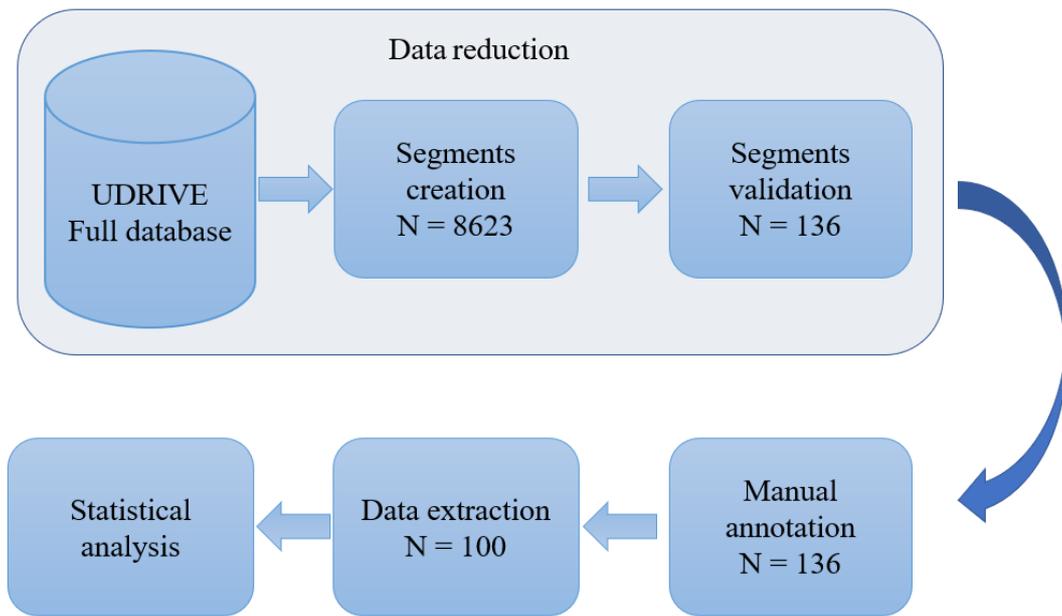


Figure 2 Flowchart showing the steps for data analysis.  $N$  is the number of selected segments after each step.

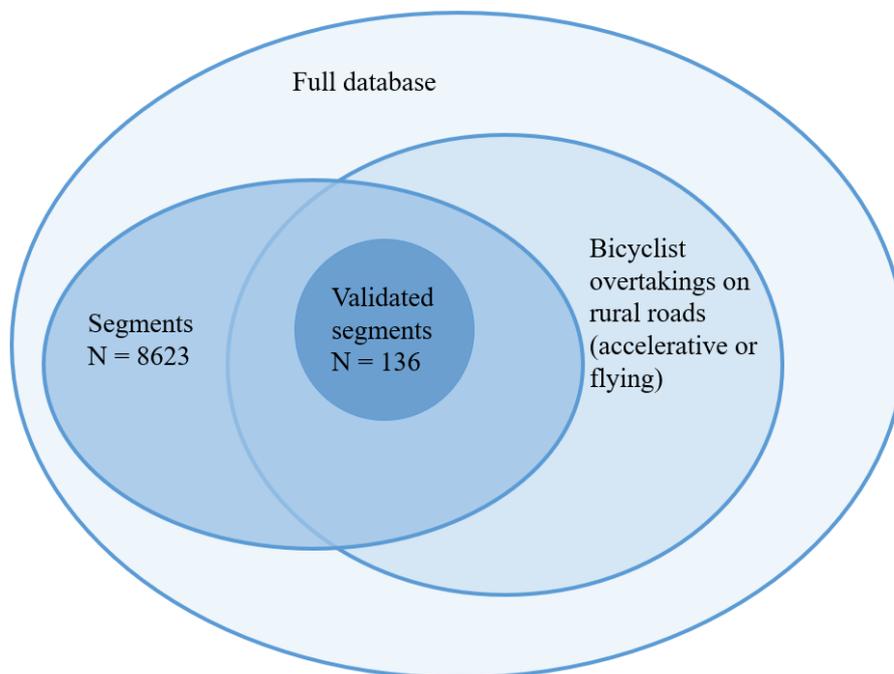


Figure 3 Data reduction: subsets in the database (the size of the Venn diagrams is just qualitative).

### 3.3.1 Segments creation

The road type from the GPS, the ego vehicle speed from the CAN bus, the road user type and relative speed from MobilEye were given as an input to a node (see Section 3.1). The node implemented a filtering algorithm in each record of the database to create segments. The following conditions were used for the creation of segments:

1. The road type from the GPS sensor was “rural road”.
2. One of the four road users identified by MobilEye was a bicyclist.
3. The absolute speed of the bicyclist, calculated as the sum of the ego vehicle speed and the bicyclist relative speed from MobilEye, was positive.

The first and second conditions were used to isolate identifications of a bicyclist on rural road, the third condition was added to exclude bicyclists travelling in opposite direction. A fourth condition of having a negative bicyclist speed relatively to the ego vehicle, even if compliant with the definition of overtaking, was not used. This choice was made knowing that bicyclists are intrinsically slower than motor vehicles and to avoid having a too selective filter.

The filtering algorithm created a Boolean signal, “false” by default, and “true” when all the conditions above were respected, with an addition of ten seconds before and after their occurrence to include all the phases of the overtaking (see Section 2.1): a segment was extracted from the record whenever the Boolean signal was true.

### 3.3.2 Segments validation

Manual annotation was firstly used to validate the segments created by the algorithm explained in Section 3.3.1. Specifically, segments validation consisted in looking at the video feed and the signals of each segment to assess if it should be included in the analysis provided starting from Section 3.4 or not. To be included in the analysis, a segment had to be compliant with the following requirements:

1. The segment included a bicyclist overtaking on a rural road, with a clearly distinguishable steer away point.
2. If the segment included an overtaking, the overtaking strategy was not piggy backing or 2+, as this thesis is focused on accelerative and flying overtaking manoeuvres only.

The majority of the segments discarded based on the first condition included bicyclists travelling in opposite direction (although the third condition given in Section 3.3.1, bicyclist travelling in opposite direction were present because of errors in measurements), as well as bicyclists travelling on the side walk, on a bicyclist lane or on another road, still close enough to be detected by MobilEye.

The criteria used to identify piggy backing and 2+ overtaking manoeuvres were the following:

- **Piggy backing** – Regardless of the speed variation of the ego vehicle, overtaking manoeuvres were annotated as “piggy backing” when a leading vehicle was followed in a row by the ego vehicle. This was assessed by ensuring that a leading vehicle was present within MobilEye’s FoV and that the longitudinal distance was kept relatively constant for the whole manoeuvre.
- **2+** – Multiple road users were overtaken during the same manoeuvre.

### 3.4 Manual annotation of validated segments

Manual annotation was used not only as a filtering tool, as in Section 3.3.2, but also as a tool to define the attributes of the validated segments. This section provides a description of how the manual annotation continued for the validated segments. The list of attributes defined via manual annotation is provided in Table 2.

Table 2 List of the annotated attributes for the validated segments.

Annotated attributes	Type	Based on:
Overtaking strategy (see Section 3.4.1)	Categorical	Ego vehicle speed
Oncoming traffic (see Section 3.4.1)	Boolean	Video feed, MobilEye
Steer away point (see Section 3.4.1)	Time stamp	Video feed, steering wheel angle
Planning point (see Section 3.4.2)	Time stamp	Video feed, pedals, turn indicator, steering wheel angle
Planning point indicator (see Section 3.4.2)	Categorical	Video feed, pedals, turn indicator, steering wheel angle
Bicyclist action type (see Section 3.4.3)	Categorical	Video feed
Bicyclist and ego vehicle are travelling at the same speed (see Section 3.4.3)	Time stamp	MobilEye, Video feed
Road Inclination (see Section 3.4.3)	Categorical	Video feed
Ego vehicle stops - signal (see Section 3.4.4)	Time stamp	Ego vehicle speed
Ego vehicle stops - video (see Section 3.4.4)	Time stamp	Video feed

### 3.4.1 Overtaking details

Each overtaking was classified as either accelerative or flying, using the definition given in Section 2.1, by means of the categorical variable “overtaking strategy”. During manual annotation, to be consistent during the whole analysis and to be able to classify also “intermediate” overtaking strategies, the following criteria based on the ego vehicle speed were used to assess the overtaking strategy:

- **Accelerative** - The ego vehicle reduced its speed during the approaching phase of at least 15% with respect to the speed before the manoeuvre started.
- **Flying** - The ego vehicle speed was relatively constant, in complementarity with the accelerative overtaking definition.

Presence of oncoming traffic was annotated using a Boolean, which was set to “true” if a road user moving on the opposite lane was detected by MobilEye during the manoeuvre.

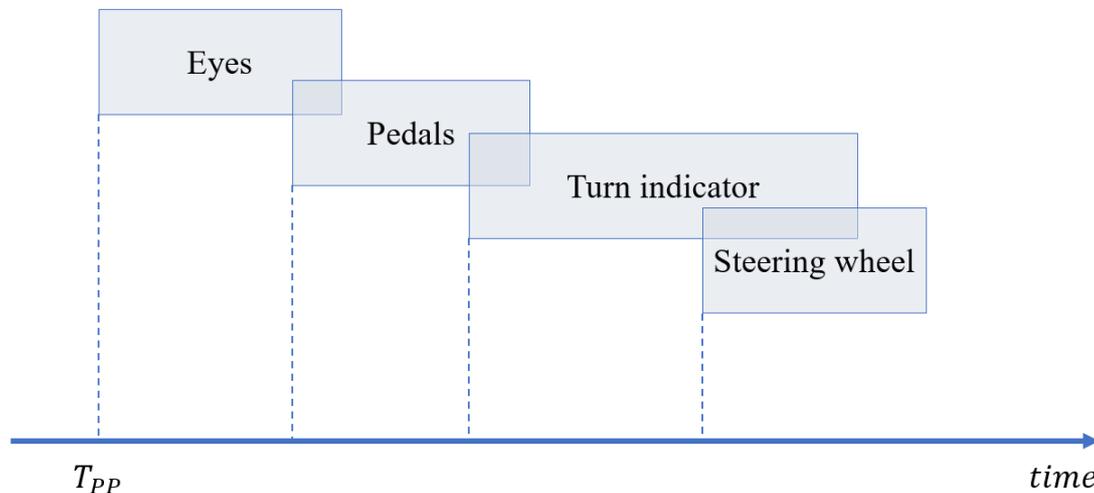
The steer away point,  $T_{SP}$ , was a time stamp annotated based on the video feed and the steering wheel angle. Specifically,  $T_{SP}$  was annotated as the time stamp in which the vehicle was visibly changing lane position, after moving the steering wheel, to perform the overtaking.

### 3.4.2 Planning point

In this thesis, the planning point was defined as the first time stamp in which data proved for the first point in time that the driver was planning to overtake (see Section 2.2). Specifically, all the signals that could give hints about the fact that the driver was planning to overtake were analysed. The name of the signal, among the following, that allowed the identification of the planning point was annotated using the categorical variable “planning point indicator”. Follows a list of the possible planning point indicators:

- **Eyes** - The driver looks far away to verify if there is a bicyclist, looks the speedometer or looks the rear mirror to make sure no vehicle is performing an overtaking.
- **Pedals** - The driver might release the accelerator pedal when a bicyclist is visible and decelerate, for instance to have more time to check if there is oncoming traffic, to look the rear mirror, or because the decision to follow the bicyclist and perform an accelerative overtaking was taken. The instant when the accelerator pedal position sets to zero or to a value lower than the beginning of the segment is considered.
- **Turn indicator** - The turn indicator could be activated before the steer away point.
- **Steering wheel** - The steering wheel angle signal was analysed together with the video to annotate the steer away point. The steer away point could be coincident with the planning point in case of flying overtaking.

The time stamp in which there was evidence that the driver was planning to overtake was annotated as planning point and labelled with the corresponding indicator. If more than one indicator was providing a possible planning point, only the first occurrence was considered, as illustrated in Figure 4.



*Figure 4* An example of what a driver could do during the approaching phase. In this case more planning point indicators could be used to identify the planning point: the annotated planning point is the first point in time  $T_{PP}$  and the annotated planning point indicator is “Eyes”.

### 3.4.3 Bicyclist details and road inclination

Bicyclist details and road inclination were annotated to support the bicyclist speed calculations, performed as in Section 3.5.3, and consisted in:

- Bicyclist action type - categorical variable. It was “standing” if the bicyclist had a foot on the ground, “fast” if the bicyclist was wearing sportswear and/or moving with an aerodynamic body position, “normal” elsewhere. The nomenclature used is consistent with the one used in a previous work (Panero, 2018).
- $T_s$  - time stamp in which ego vehicle and bicyclist are travelling at the same speed
- Road inclination - categorical variable. It was either “flat”, “downhill” or “uphill”.

While the bicyclist action type and the road inclination were easy to annotate,  $T_s$  needs some further explanation. In some overtaking manoeuvres, the ego vehicle followed the bicyclist at its same speed for a while: this was assessed by checking if the raw signal from MobilEye gave a relatively constant longitudinal distance of the bicyclist from the ego vehicle for a few time stamps and by looking at the video feed. In such cases, one of the time stamps in which the bicyclist longitudinal distance was constant was annotated as  $T_s$ .

### 3.4.4 Quantification of the delay between video and signals

The videos and the signals have a time delay of decimals of second. This delay was quantified using the same procedure used in (Nero, 2017): the time stamps in which the vehicle stopped in both the signal and the video feed were annotated separately and used to calculate the delay. Continuing, the time stamps annotated based on the video feed (the planning point when the indicator was “Eyes” and the steer away point) were synchronized with the signals using the calculated delay. Notice that the time stamp in which bicyclist and ego vehicle are travelling at the same speed  $T_S$  was not corrected because it was based on the signal from MobilEye, while video feed was used only to check whether the bicyclist and the driver were really travelling at the same speed (see Section 3.4.3).

## 3.5 Data extraction

While data about the ego vehicle was always available from the CAN bus, the information about the bicyclist was only available when it was in the FoV of MobilEye. Hence, the bicyclist distance and speed at the planning point had to be extrapolated before extracting the data needed for the statistical analysis. The procedure for data extrapolation and extraction was inspired by previous works done for pedestrian overtaking manoeuvres (Rasch, 2018) (Panero, 2018) and bicyclist overtaking manoeuvres (Nero, 2017). Table 3 provides an overview of the data extracted from each segment.

*Table 3 Variables extracted from each segment*

Extracted variable	Type
Ego vehicle speed at the planning point	Continuous [ <i>km/h</i> ]
Driver’s ID (see Section 3.5.1)	Categorical
Driver’s gender - Male/Female (see Section 3.5.1)	Categorical
Driver’s age (see Section 3.5.1)	Continuous [ <i>years</i> ]
Country (see Section 3.5.1)	Categorical
Bicyclist speed (see Section 3.5.3)	Continuous [ <i>km/h</i> ]
Bicyclist longitudinal distance from the ego vehicle at the planning point (see Section 3.5.4)	Continuous [ <i>m</i> ]
Bicyclist lateral distance from the ego vehicle at the planning point (see Section 3.5.5)	Continuous [ <i>m</i> ]
TTC at the planning point (see Section 3.5.6)	Continuous [ <i>s</i> ]
TTC at the steer away point (see Section 3.5.6)	Continuous [ <i>s</i> ]

### 3.5.1 Driver information and country

The information about the driver was provided in the database. The variables of interest concerning the driver were the driver's age (in years), the driver's gender (male/female) and the driver's ID (a number associated to each driver, useful as a grouping variable for the observed data). The country where the overtaking was observed was extracted from the database as a categorical variable. The country was used to have an overview of where the overtaking manoeuvres were collected and was used in the calculation of the bicyclist lateral distance at the planning point (see Section 3.5.5) to use the same sign convention in the United Kingdom (only left-hand traffic country involved in the UDRIVE project) and in right-hand traffic countries.

### 3.5.2 Rotation of MobilEye data

Since data from MobilEye was given in a local coordinate system, translating and rotating together with the ego vehicle, there was the need to transform the provided values in a car-fixed coordinate system (only translating with the vehicle). The car-fixed coordinate system provided data about the relative position of the bicyclist keeping a constant orientation of longitudinal and lateral coordinate. The ego vehicle heading angle  $\psi$  was used to calculate a rotation matrix  $\mathbf{R}(\psi)$ , used to rotate the coordinate system.

The ego vehicle heading angle was calculated as in equation (1).

$$\psi(T) = \int_{T_{SP}}^T \omega_z(\tau) d\tau \quad T_{begin} < T < T_{end} \quad (1)$$

where  $T$  is the time axis,  $\psi(T)$  is the time history of the heading angle,  $\omega_z(T)$  is the time history of the yaw rate from the CAN bus,  $T_{begin}$  is the beginning of the overtaking segment,  $T_{end}$  is the end of the overtaking segment,  $T_{SP}$  is the steer away point and  $\tau$  is an auxiliary variable to express the integral. The steer away point is chosen as a starting point for the integral to obtain null heading angle when the driver enters the *steering away phase* (see Section 2.1).

The rotation from a local coordinate system to a car-fixed coordinate system was performed as in equation (2).

$$\begin{bmatrix} x \\ y \end{bmatrix} = \underbrace{\begin{bmatrix} \cos(\psi) & -\sin(\psi) \\ \sin(\psi) & \cos(\psi) \end{bmatrix}}_{\mathbf{R}(\psi)} \cdot \begin{bmatrix} x' \\ y' \end{bmatrix} \quad (2)$$

where  $x$  and  $y$  are the car-fixed coordinates,  $x'$  and  $y'$  are the local coordinates,  $\psi$  is the heading angle and  $\mathbf{R}(\psi)$  is the rotation matrix. The two coordinate systems are illustrated on the ego vehicle in Figure 5. Figure 6 shows how the projections of a generic positive position vector,  $p_{me}$ , change from a local to a car-fixed coordinate system.

Starting from the next Section, the terms longitudinal distance and lateral distance from MobilEye refer to the values in car-fixed coordinates.

For some overtaking manoeuvres, the yaw rate from the CAN bus was not available. In this case, the values from MobileEye were assumed to be directly given in car-fixed coordinates.

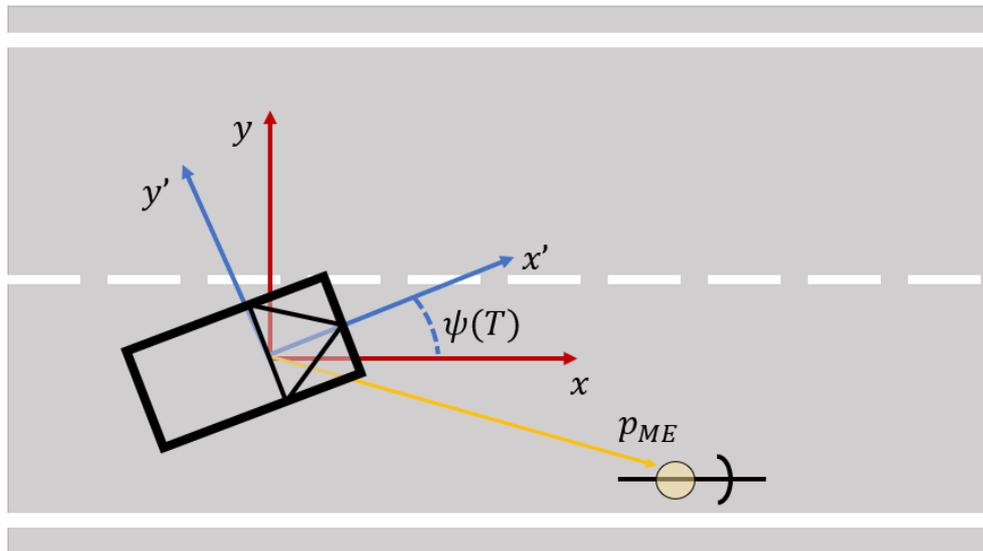


Figure 5 Illustration of the car-fixed coordinate system  $x$ - $y$ , the local coordinate system  $x'$ - $y'$  and the heading angle.

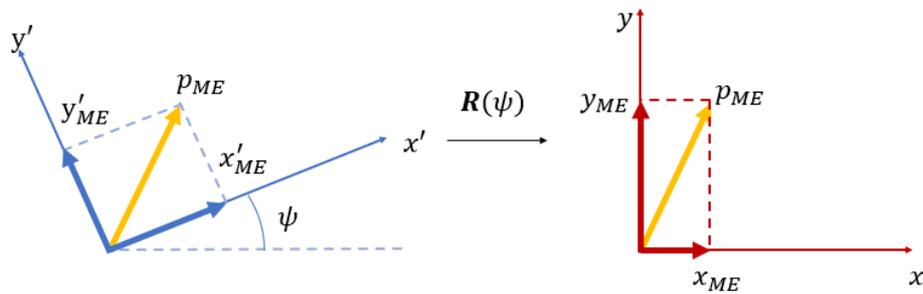


Figure 6 Rotation of the coordinate system using the rotation matrix  $\mathbf{R}(\psi)$ , for the same generic vector  $p_{ME}$ .

### 3.5.3 Bicyclist speed

If the time stamp in which bicyclist and ego vehicle are travelling at the same speed ( $T_s$ ) was annotated (see Section 3.4.3), the bicyclist speed was set equal to the ego vehicle speed from the CAN bus in such time stamp, and the rest of the calculations provided in this section was not performed.

If  $T_s$  was not annotated, the bicyclist speed was calculated using the longitudinal distance provided by MobilEye: after fitting it with a linear polynomial, using the MATLAB<sup>®</sup> function *polyfit* (MathWorks, Polynomial Curve Fitting, 2019), the linear term of the polynomial was assumed to be the relative speed between the bicyclist and the ego vehicle (Nero, 2017) when the bicyclist appears. The bicyclist speed was calculated as in equation (3).

$$V_b = V_{ego}(T_x) - RV(T_x) \quad (3)$$

where  $V_b$  is the bicyclist speed,  $T_x$  is the time stamp in which the bicyclist appears,  $RV(T_x)$  is the relative speed of the bicyclist when the bicyclist appears, calculated from the longitudinal distance provided by MobilEye,  $V_{ego}(T_x)$  is the ego vehicle speed when the bicyclist appears. The bicyclist speed was assumed to be constant for the whole overtaking.

The bicyclist speed had to be in the range from 0 km/h to 40 km/h to be included in the statistical analysis described in Section 3.6: negative values were not possible since the ego vehicle and the bicyclist were travelling in the same direction, while values above 40 km/h are generally difficult to reach (What is a reasonable speed for long distances on a bike?, 2011) (Nero, 2017). The annotated road inclination and bicyclist action type (see Section 3.4.3) were used to support the calculated bicyclist speed.

### 3.5.4 Bicyclist longitudinal distance

After calculating the bicyclist speed as in Section 3.5.3, the time history of the relative speed throughout the whole segment was calculated using equation (4).

$$RV(T) = V_{ego}(T) - V_b \quad (4)$$

Where  $T$  is the time axis,  $RV(T)$  is the time history of the relative speed and  $V_b$  is the bicyclist speed, calculated as in Section 3.5.3.

The value of the longitudinal distance before the bicyclist appears is calculated as in equation (5).

$$\tilde{x}(T) = x_{ME}(T_x) + \int_{T_x}^T RV(\tau) \cdot |\tau - T_x| d\tau, \quad T_{begin} < T < T_x \quad (5)$$

where  $T$  is the time axis,  $\tilde{x}(T)$  is the time history of the longitudinal distance before the bicyclist appears,  $T_x$  is the time stamp when the bicyclist appears in MobilEye,  $T_{begin}$  and  $T_{end}$  are respectively the beginning and the end of the overtaking segment and  $\tau$  is an auxiliary variable to express the integral. Note that the absolute value

of the time is considered, to have positive values of longitudinal distance before the bicyclist appears, consistently with the longitudinal distance provided by MobilEye. The longitudinal distance before the bicyclist appears was joined to the longitudinal distance available from MobilEye to express the bicyclist longitudinal distance as in equation (6).

$$d(T) = \begin{cases} \tilde{x}(T), & T_{begin} < T < T_x \\ x_{ME}(T), & T_x \leq T < T_0 \end{cases} \quad (6)$$

where  $T$  is the time axis,  $T_x$  is the time stamp when the bicyclist appears,  $T_0$  is the time stamp when the bicyclist disappears,  $T_{begin}$  is the beginning of the overtaking segment,  $d(T)$  is the time history of the bicyclist longitudinal distance,  $x_{ME}(T)$  is the time history of the bicyclist longitudinal distance from MobilEye and  $\tilde{x}(T)$  is the time history of the longitudinal distance before the bicyclist appears in MobilEye, calculated as in Equation (5). The longitudinal distance was not calculated after the bicyclist disappeared from MobilEye, since the planning point is before bicyclist disappearance.

### 3.5.5 Bicyclist lateral distance at the planning point

The lateral distance between ego vehicle and bicyclist at the planning point,  $y_{pp}$ , was calculated using the distance of the ego vehicle to the lane markings, or the lane edges (lane position) and the bicyclist lateral distance, both detected by MobilEye.

Figure 7 illustrates the geometry used to perform the calculations in this section:  $y_{pp}$  is the lateral distance at the planning point,  $dL_{ego}$  is the lane position of the ego vehicle,  $dL_b$  is the lane position of the bicyclist,  $y_{ME}$  is the lateral distance provided by MobilEye. The information about the country (see Section 3.5.1) was used to choose the same sign convention in United Kingdom and in the other right-hand traffic countries. In the United Kingdom,  $y_{ME}$  was considered positive when the bicyclist was on the left side of the ego vehicle, and the lane position was set equal to the left edge (or marking) distance. In right-hand traffic countries,  $y_{ME}$  was considered positive when the bicyclist was on the right side of the ego vehicle, and the lane position was set equal to the right edge (or marking) distance.

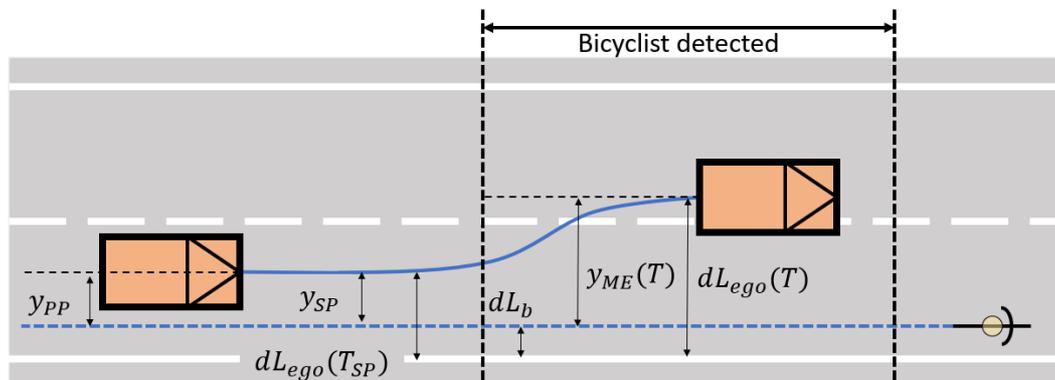


Figure 7 Geometry for the calculation of the lateral distance at the planning point  $y_{pp}$ . The dashed vertical lines highlight when the bicyclist is in the FoV of MobilEye.

The known values from the database are the lane position of the ego vehicle in every time stamp  $dL_{ego}(T)$  and the lateral distance in car-fixed coordinates of the bicyclist given by MobilEye  $y_{ME}$ .

The bicyclist lane position  $dL_b$  was calculated as the average of all the bicyclist lane positions calculated for each time stamp in which the bicyclist was detected, as in equation (7).

$$dL_b = \underset{T \in [T_x, T_0]}{\text{mean}} (dL_{ego}(T) - y_{ME}(T)) \quad (7)$$

where  $T$  is the time axis,  $T_x$  is the time stamp in which the bicyclist appears,  $T_0$  is the time stamp in which the bicyclist disappears,  $dL_b$  is the lane position of the bicyclist,  $dL_{ego}$  is the lane position of the ego vehicle,  $y_{ME}$  is the bicyclist lateral distance from MobilEye. The time average was performed to exploit all the data available during the bicyclist detection and in an attempt to filter eventual outliers in the lateral distance from MobilEye  $y_{ME}$ . Continuing, the lateral distance at the planning point was calculated as in equation (8).

$$y_{PP} = y_{SP} = dL_{ego}(T_{SP}) - dL_b \quad (8)$$

where  $y_{pp}$  is the bicyclist lateral distance at the planning point,  $y_{SP}$  is the lateral distance at the steer away point,  $T_{SP}$  is the steer away point,  $dL_{ego}(T_{SP})$  is the ego vehicle lane position at the steer away point and  $dL_b$  is the bicyclist lane position extracted as in equation (7). The assumption in this procedure is that the lateral distance at the planning point is the same as the lateral distance at the steer away point. This assumption was done because the lane position provided by MobilEye strongly depended on the infrastructure and the value of the signal at the planning point might have been considerably different from the steer away point: this happened not because of lateral motion of the ego vehicle, but because the lane markings' detection was different due to either a different visibility of the paint or a real change in where the lane markings were positioned. Choosing the steer away point, that is closer to when the bicyclist is detected, allowed to minimize the error due to this effect. The calculated value of lateral distance at the planning point was compared with the evidence from the video feed: overtaking manoeuvres in which the calculated lateral distance did not match with the evidence from the video feed were not included in the statistical analysis provided in Section 3.6.

### 3.5.6 Time to Collision

TTC was calculated applying the definition, given in equation (9).

$$TTC = \frac{d}{RV} \quad (9)$$

where  $d$  is the longitudinal distance and  $RV$  is the relative speed, calculated as in Section 3.5.3. The value of TTC was calculated at both the planning point and the steer away point, as in equation (10) and equation (11).

$$TTC_{PP} = \frac{d(T_{PP})}{RV(T_{PP})} \quad (10)$$

$$TTC_{SP} = \frac{d(T_{SP})}{RV(T_{SP})} \quad (11)$$

where  $T_{PP}$  and  $T_{SP}$  are the planning point and the steer away point respectively,  $TTC_{PP}$  is the TTC at the planning point,  $TTC_{SP}$  is the TTC at the steer away point,  $d(T_{PP})$  is the longitudinal distance at the planning point,  $d(T_{SP})$  is the longitudinal distance at the steer away point,  $RV(T_{PP})$  is the relative speed at the planning point and  $RV(T_{SP})$  is the relative speed at the steer away point. All the calculated values of TTC were positive, since no crash was observed during the analysis.

### 3.6 Generalised Linear Mixed Model (GLMM)

TTC and longitudinal distance at the planning point were analysed using generalised linear mixed models, or GLMMs.

GLMMs are used to describe the relation between a *response*, also called dependent variable, and one or more *factors*, also called independent variables. The factors can be either *observable* or *unobservable*. The *effects* represent the influence of the factors on the responses. The set of effects includes both *fixed effects* and *random effects* (from which the name *mixed* model).

The fixed effects express the influence of the observable factors on the response, while random effects express the influence of unobservable factors on the response: random effects need to be considered to correctly estimate the fixed effects. An unobservable factor could be, for instance, related to grouping data during collection. In this thesis, when analysing NDD, a grouping variable is the driver ID (see Section 3.5.1): it is reasonable to think that differences between two similar overtaking manoeuvres are due to the driver's driving style, that is intrinsically an unobservable factor. Using a mixed effect model allows to account for the influence of the driver's driving style and to correctly estimate the influence of the observable factors only. The model can be expressed as in equation (12) (MathWorks, 2019).

$$y_i|u \sim \text{Distribution}\left(\mu_i, \frac{\sigma^2}{w_i}\right)$$

$$g(\mu) = \eta \quad (12)$$

$$\eta = \mathbf{X}\beta + \mathbf{Z}u + \epsilon, \quad \mathbf{X} \in \mathbb{R}^{r \times n}, \mathbf{Z} \in \mathbb{R}^{r \times m}$$

$$u|\sigma^2, \theta \sim N(0, \sigma^2 D(\theta))$$

The  $i^{th}$  response out of  $r$  observations,  $y_i$ , is assumed to belong to a certain distribution,  $\text{Distribution}\left(\mu_i, \frac{\sigma^2}{w_i}\right)$ , with mean  $\mu_i$  and variance  $\frac{\sigma^2}{w_i}$ , where  $w_i$  is a weight associated to each observation. The response is modelled as a function of  $n$  observable factors and  $m$  unobservable factors. The mean vector  $\mu$  is calculated as a

function of a linear predictor  $\eta$ . Continuing,  $\eta$  is a linear combination of the fixed effects design matrix  $\mathbf{X}$ , weighted by the fixed effect vector  $\beta$ , and of the random effects design matrix  $\mathbf{Z}$ , weighted by the random effects vector  $u$ , plus a model offset  $\epsilon$ . The random effects vector is assumed to be normally distributed with null mean and standard deviation  $\sigma^2 D(\theta)$ , where  $D(\theta)$  is a semi-positive matrix, function of the parameter  $\theta$ . The relation between the linear predictor and the mean is given by the link function  $g(\mu)$ . The model is parametrized by  $\beta$ ,  $\sigma^2$  and  $\theta$ . The GLMM calculates the model parameters  $\beta$ ,  $\sigma^2$  and  $\theta$  that best fit the observed responses. After the calculation of the parameters, the mean of the fitted response is estimated as in equation (13).

$$\mu = g^{-1}(\hat{\eta}) = g^{-1}(\mathbf{X}\beta + \mathbf{Z}u) \quad (13)$$

In this thesis, the focus is on the fixed effects vector  $\beta$ , to estimate the influence of the observable factors on the responses to model (TTC and longitudinal distance at the planning point).

### 3.6.1 Factors and responses

The responses to model were the TTC at the planning point,  $TTC_{pp}$ , and the longitudinal distance at the planning point,  $d_{pp}$ .  $TTC_{pp}$  was modelled as an inverse gaussian distribution and  $d_{pp}$  was modelled as a gamma distribution, while the chosen link function was an identity function for both the models. The distributions were chosen by trial-and-error for each of the responses: the choice was between either a gamma or an inverse gaussian distribution, as  $d_{pp}$  and  $TTC_{pp}$  were known to be positive (see Section 3.5.4 and Section 3.5.6) (MathWorks, 2019).

The set of the possible observable factors, taken from the annotated attributes in Table 2 and the extracted data in Table 3, is provided in Table 4.

*Table 4 List of the single observable factors for the GLMMs of TTC and longitudinal distance at the planning point. The values in the table are observed at the planning point.*

Source	Variable
Continuous	Bicyclist speed $x_{v_b}$ [km/h]
	Ego vehicle speed $x_{v_{ego}}$ [km/h]
	Lateral distance $x_y$ [m]
	Driver's age $x_{age}$ [years]
Boolean	Oncoming traffic $x_{on}$ (present = 1, absent = 0)
	Overtaking strategy $x_{fl}$ (flying = 1, accelerative = 0)
	Driver's gender $x_M$ (M = 1, F = 0)

The model was built considering as observable factors all the single observable factors provided in Table 4 and all the possible two-way interactions between the single observable factors  $x_i: x_j$  (higher order interactions were not included in the model); the only unobservable factor was assumed to be a variation in the intercept of the response due to the driver and was symbolized as (1|ID). The overtaking strategy (see Table 2) and the driver's gender (see Table 3) were considered as Boolean variables in the statistical analysis since they originally were binary categorical variables.

### 3.6.2 Implementation in MATLAB®

The GLMM was implemented using the MATLAB® function *fitglme* (MathWorks, fitglme, 2019). The function requires an input formula, providing the response to model and the factors to include in the model, a distribution and a link function.

While the random effect was included by default, the fixed effects were analysed to select the significant ones only. This was done by using a stepwise algorithm, similar to the one proposed in a previous work (Boda, C.-N., et al., 2018). The stepwise algorithm: 1) used the MATLAB® function *fitglme* to fit the response including all the factors given in Section 3.6.1, 2) calculated the fixed effects, 3) discarded the observable factor associated to the fixed effect with the highest  $p$ -value, if this was greater than 0.05, 4) repeated again with the left observable factors, iteratively. The final result is a set of fixed effects with  $p$ -values lower than 0.05, meaning that the corresponding factors had a significant influence on the response, according to the chosen convention of  $p$ -value  $\leq 0.05$ . This algorithm was implemented to both  $TTC_{PP}$  and  $d_{PP}$ , using the distribution and the link function provided in Section 3.6.1. The formulas provided in the rest of this section are the result of this algorithm.

The formulas for the TTC at the planning point and the longitudinal distance at the planning point are respectively given in equation (14) and equation (15). The notation used is the one provided in MATLAB® documentation (MathWorks, fitglme, 2019): a response (left part of the equation) is distributed as ( “~” ) a list of factors (right part of the equation).

$$TTC_{PP} \sim 1 + x_y + x_{V_b} + x_{V_{ego}} + x_{on} + x_{male} + x_{V_{ego}:x_y} + x_{V_{ego}:x_{on}} + x_y:x_{male} + x_y:x_{age} + x_{V_b}:x_{age} + (1|ID) \quad (14)$$

$$d_{PP} \sim 1 + x_y + x_{V_b} + x_{on} + x_{V_{ego}:x_y} + x_{V_{ego}:x_{on}} + (1|ID) \quad (15)$$

The selected distribution for  $TTC_{PP}$  was an inverse gaussian distribution, the selected distribution for  $d_{PP}$  was a gamma distribution and the selected link function was an identity function for both, as mentioned in Section 3.6.1.

## 4 Results

In this chapter, the results of the methodology are provided. Section 4.1 provides an overview of the collected overtaking manoeuvres, Section 4.2 presents the collected TTC at the planning point, Section 4.3 presents the collected longitudinal distance at the planning point (results are provided considering the overtaking strategy, the presence of oncoming traffic and the ego vehicle speed to qualitatively identify any possible trend), Section 4.4 presents TTC at the planning point and at the steer away point, Section 4.5 provides the results from the GLMMs of TTC and longitudinal distance at the planning point.

### 4.1 Analysed overtaking manoeuvres

Starting from the full database, a data reduction phase gave a total number of 136 overtaking manoeuvres (see Section 3.3). Followed further annotation (see Section 3.4) and data extraction (see Section 3.5). The statistical analysis was performed on only 100 overtaking manoeuvres: 36 overtaking manoeuvres were excluded from the statistical analysis because the bicyclist speed was negative or above 40 km/h (see Section 3.5.3) or because the lateral distance was considerably different from the evidence in the video feed (see Section 3.5.5). Percentages are not reported, as the number of selected overtaking manoeuvres happened to be exactly 100.

40 drivers were involved in the analysis: 27 male drivers performed 61 overtaking manoeuvres (30 accelerative and 31 flying overtaking manoeuvres), 13 female drivers performed 39 overtaking manoeuvres (26 accelerative and 13 flying). The average age of all the drivers was 42 years (average age per gender was 45 years for male drivers and 37 years for female drivers). The number of overtaking manoeuvres per driver ranged from one overtaking to a maximum of eight overtaking manoeuvres, with a mean value of 2.5 overtaking manoeuvres per driver. The planning point indicator (see Section 3.4.2) was “eyes” in 46 overtaking manoeuvres and “pedals” in 54 overtaking manoeuvres. The number of overtaking manoeuvres per country, classified per overtaking strategy, is given in Table 5. The number of overtaking manoeuvres per country, classified per presence of oncoming traffic, is given in Table 6. 44 overtaking manoeuvres were flying while 56 were accelerative; 61 overtaking manoeuvres were performed with oncoming traffic while 39 were performed without oncoming traffic.

*Table 5 Number of overtaking manoeuvres per country, classified per overtaking strategy.*

Country	Flying	Accelerative	Total
Poland	20	29	49
United Kingdom	13	13	26
France	7	13	20
Netherlands	4	0	4
Germany	0	1	1
All countries	44	56	100

*Table 6 Number of overtaking manoeuvres per country, classified per presence of oncoming traffic.*

Country	Oncoming	No oncoming	Total
Poland	32	17	49
United Kingdom	17	9	26
France	10	10	20
Netherlands	1	3	4
Germany	1	0	1
All countries	61	39	100

The number of overtaking manoeuvres, classified per overtaking strategy and presence of oncoming traffic, is given in Table 7. The influence of the presence of oncoming traffic on the chosen strategy is noticeable, since 77 overtaking manoeuvres out of 100 were either accelerative with oncoming traffic or flying without oncoming traffic.

*Table 7 Number of overtaking manoeuvres, grouped per overtaking strategy and presence of oncoming traffic.*

	Flying	Accelerative	Total
Oncoming	14	47	61
No oncoming	30	9	39
Total	44	56	

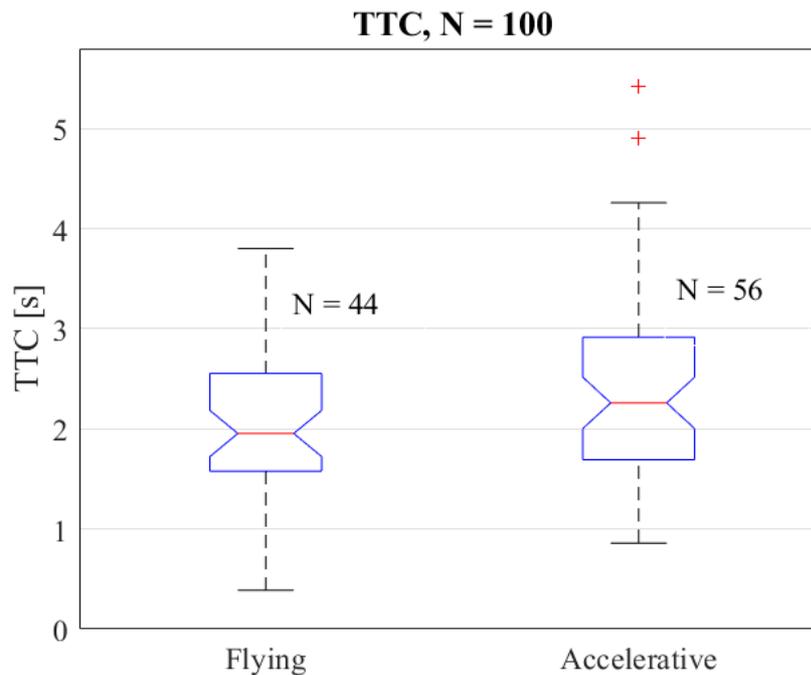
## 4.2 TTC at the planning point

An overview of the TTC at the planning point, in terms of mean and standard deviation, is given in Table 8. The mean does not seem to considerably vary due to presence of oncoming traffic in case of accelerative overtaking manoeuvres but has a more visible variation if flying overtaking manoeuvres are considered (it is higher if oncoming traffic is present). The standard deviation is higher for the groups with less observations (accelerative without oncoming traffic and flying with oncoming traffic, as provided in Table 7).

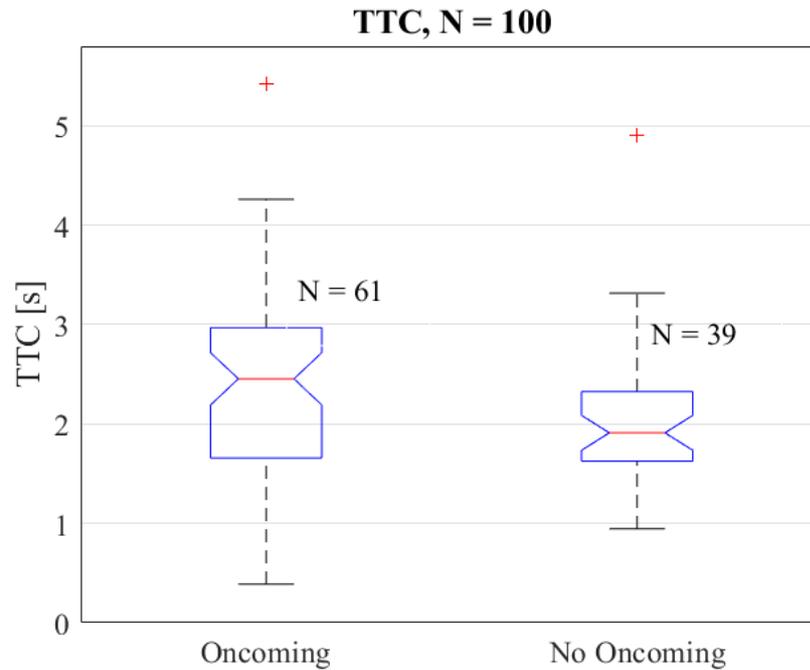
*Table 8 Mean and standard deviation of the TTC at the planning point, grouped per overtaking strategy and presence of oncoming traffic*

		Mean [s]	Standard deviation [s]
Accelerative	Oncoming	2.3683	$\pm 0.9062$
	No oncoming	2.3266	$\pm 1.1042$
Flying	Oncoming	2.4371	$\pm 0.9624$
	No oncoming	1.9505	$\pm 0.6306$

Values of TTC at the planning point, grouped per overtaking strategy and presence of oncoming traffic, are respectively provided in Figure 8 and Figure 9. Figure 8 shows that in case of flying overtaking manoeuvres, the median of TTC when the drivers started planning to overtake is lower compared to accelerative overtaking manoeuvres. Similarly, Figure 9 shows that the median of TTC when the drivers started planning to overtake is lower when oncoming traffic is absent compared to when oncoming traffic is present. Moreover, the range of values (region between the whiskers) of TTC when oncoming traffic is absent is fully included in the range of TTC when oncoming traffic is present.



*Figure 8* Boxplot of TTC at the planning point, grouped per overtaking strategy. The edges of each box represent the first and the third quartile, the whiskers give the range of the observations, the centre line represents the median, the notched region represents the 95% confidence interval for the median, the plus symbols represent the outliers.



*Figure 9* Boxplot of TTC at the planning point, grouped per presence of oncoming traffic. The edges of each box represent the first and the third quartile, the whiskers give the range of the observations, the centre line represents the median, the notched region represents the 95% confidence interval for the median, the plus symbols represent the outliers.

Figure 10 provides the TTC at the planning point, grouped considering the interaction between overtaking strategy and presence of oncoming traffic. It seems that the overtaking strategy does not influence the TTC at the planning point. Vice versa, presence of oncoming traffic has a more visible influence, causing a reduction of the median of TTC for both strategies. Hence, the main contributor in causing a difference in the TTC provided in Figure 8 is not the overtaking strategy, but the fact that the majority of flying manoeuvres were performed without oncoming traffic and the majority of accelerative overtaking manoeuvres were performed with oncoming traffic (see Table 7).

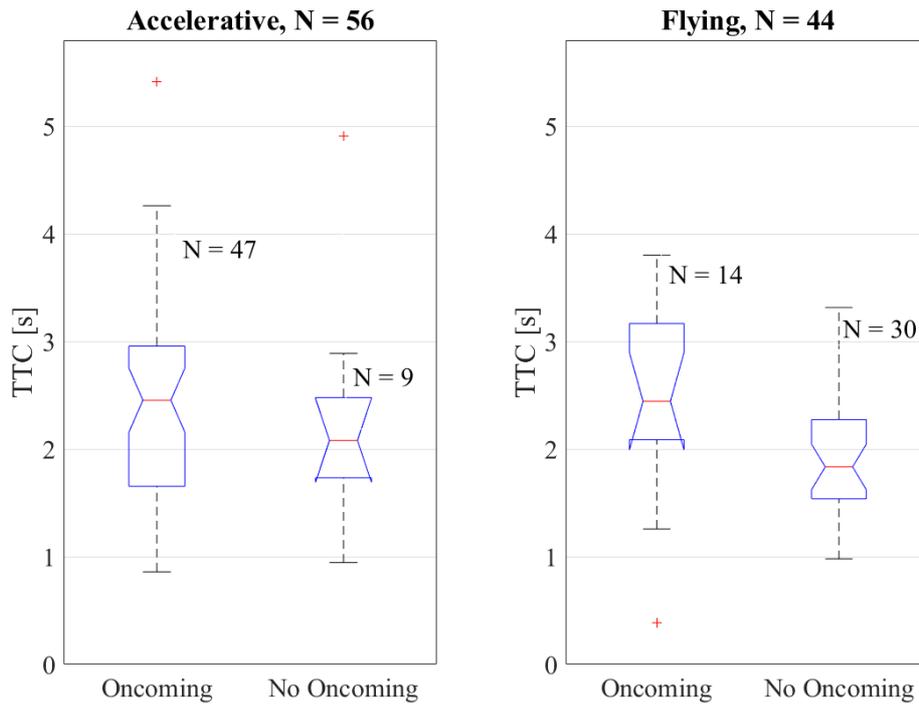


Figure 10 Boxplots of TTC at the planning point, grouped per overtaking strategy and presence of oncoming traffic. The edges of each box represent the first and the third quartile, the whiskers give the range of the observations, the centre line represents the median, the notched region represents the 95% confidence interval for the median, the plus symbols represent the outliers.

Scatter plots of the TTC at the planning point and the corresponding ego vehicle speed, grouped per interaction of overtaking strategy and presence of oncoming traffic, are provided in Figure 11 and Figure 12. The plotted lines are the result of a linear regression and are used to qualitatively identify an increasing or decreasing trend of the TTC with increasing ego vehicle speed. The fitting lines are plotted considering the overtaking strategy in Figure 11. The same scatter plot, with fitting lines plotted considering the presence of oncoming traffic, is provided in Figure 12. The TTC does not seem to be considerably influenced by the ego vehicle speed, since all the fitting lines have a low absolute value of the slope (maximum variation possible is roughly 0.5 s, when the vehicle speed increases from approximately 25 km/h to approximately 110 km/h).

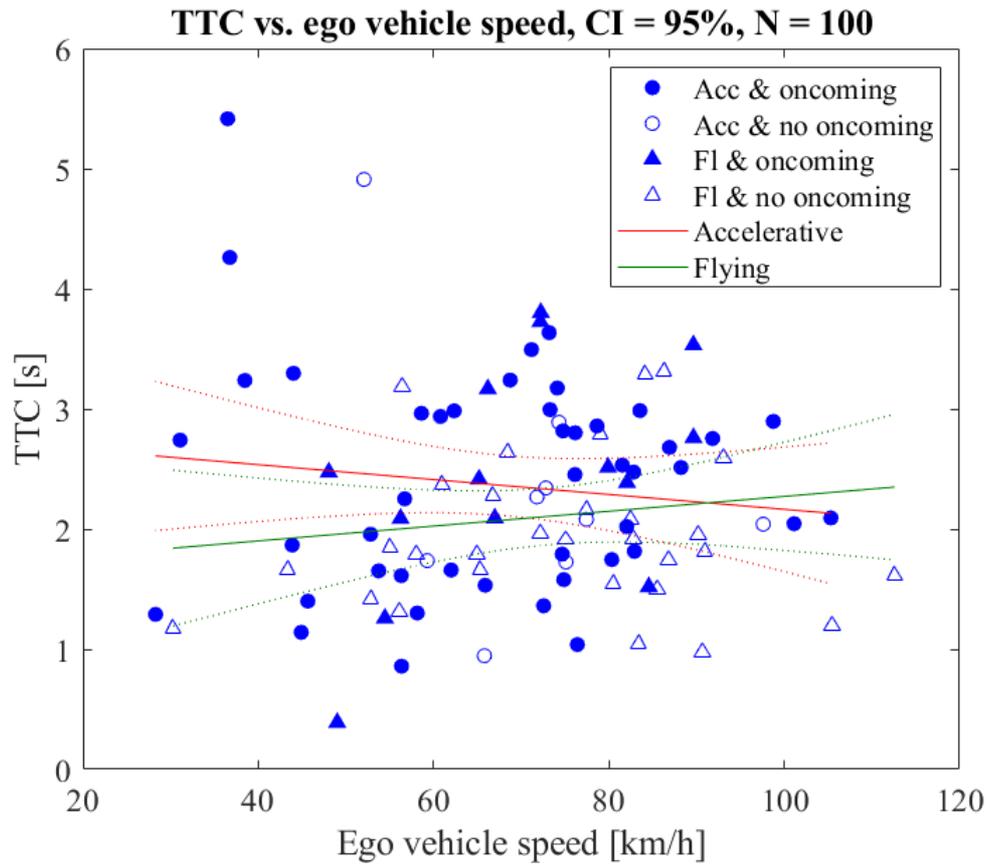


Figure 11 Scatter plot of TTC at the planning point and the corresponding ego vehicle speed. Fitting lines, with corresponding 95% confidence intervals (dotted lines), are plotted separately for flying and accelerative manoeuvres.

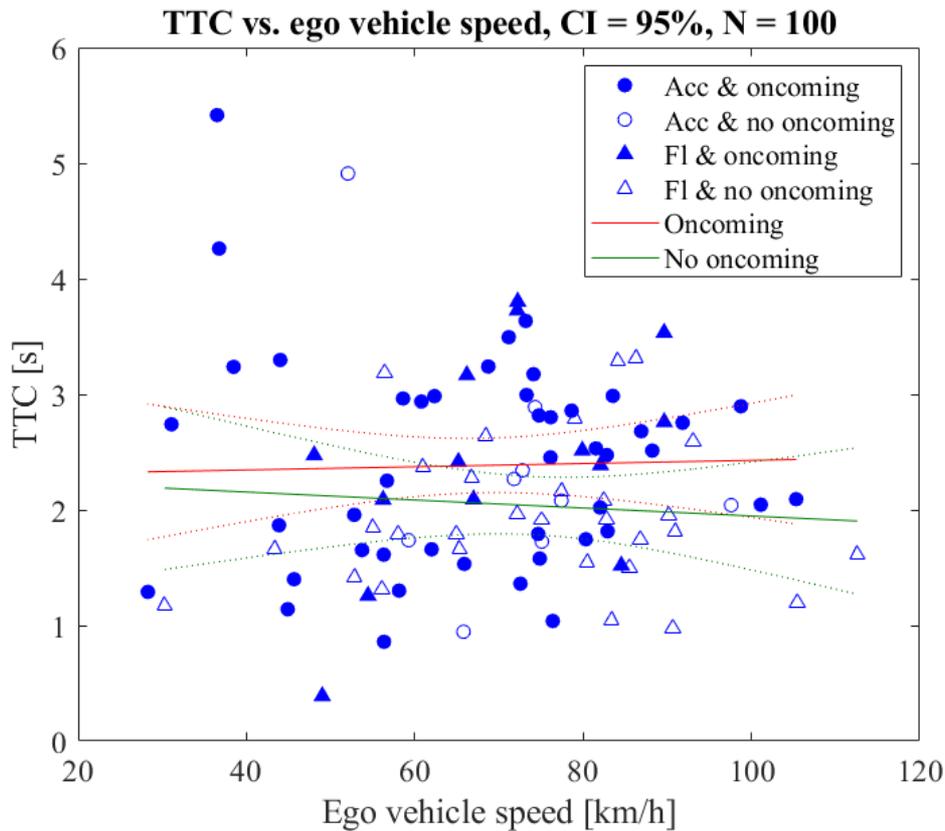


Figure 12 Scatter plot of TTC at the planning point and the corresponding ego vehicle speed. Fitting lines, with corresponding 95% confidence intervals (dotted lines), are plotted separately for present and absent oncoming traffic.

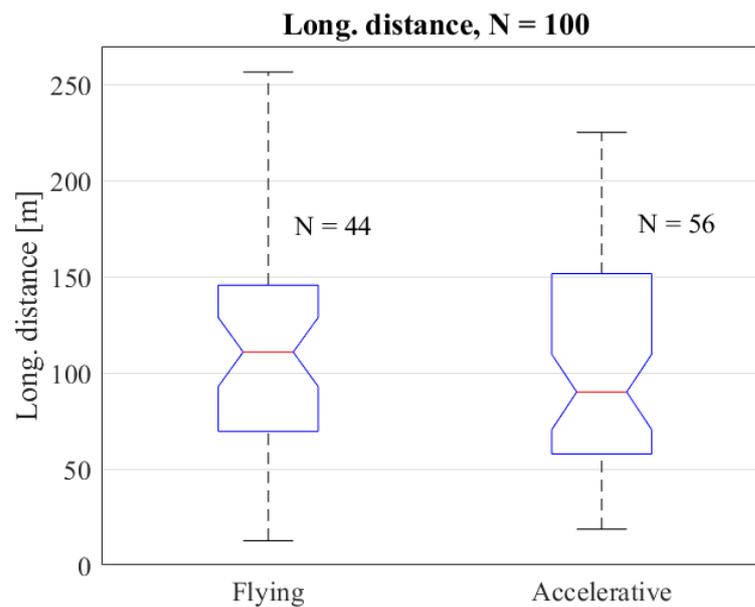
### 4.3 Longitudinal distance at the planning point

An overview of the longitudinal distance at the planning point, in terms of mean and standard deviation, is given in Table 9. When accelerative overtaking manoeuvres are considered, the mean of the longitudinal distance is higher when oncoming traffic is absent. On the contrary, when flying overtaking manoeuvres are considered, the mean of the longitudinal distance is higher when oncoming traffic is present. The standard deviation is higher when oncoming traffic is present compared to when oncoming traffic is absent for both strategies. The mean values are always above 100 m.

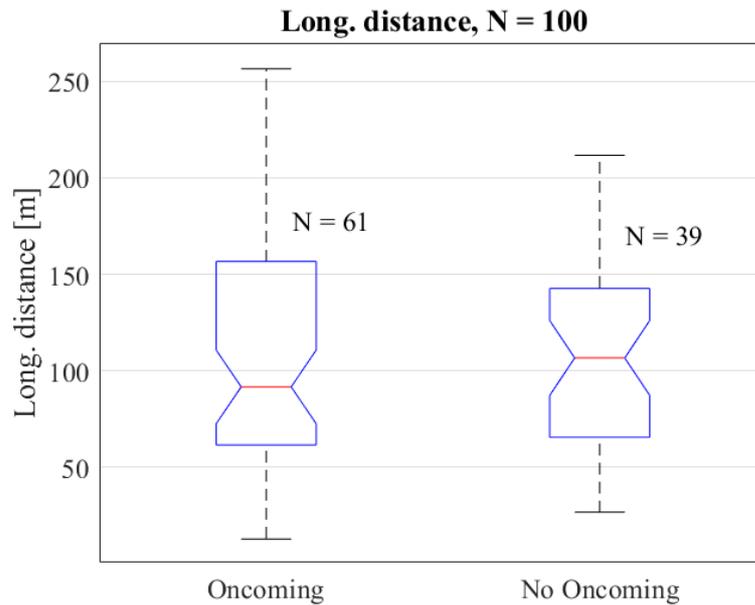
Table 9 Mean and standard deviation of the longitudinal distance at the planning point, grouped per overtaking strategy and presence of oncoming traffic

		Mean [m]	Standard deviation [m]
Accelerative	Oncoming	101.9882	$\pm 58.0464$
	No oncoming	107.4715	$\pm 45.9728$
Flying	Oncoming	122.4627	$\pm 63.3499$
	No oncoming	106.9782	$\pm 46.6257$

Figure 13 and Figure 14 provide the values of longitudinal distance at the planning point, respectively grouped per overtaking strategy and presence of oncoming traffic. It seems that drivers started planning to overtake when closer to the bicyclist (lower median of longitudinal distance) in case of accelerative overtaking manoeuvres compared to flying overtaking manoeuvres (see Figure 13) and if oncoming traffic was present compared to when it was absent (see Figure 14).

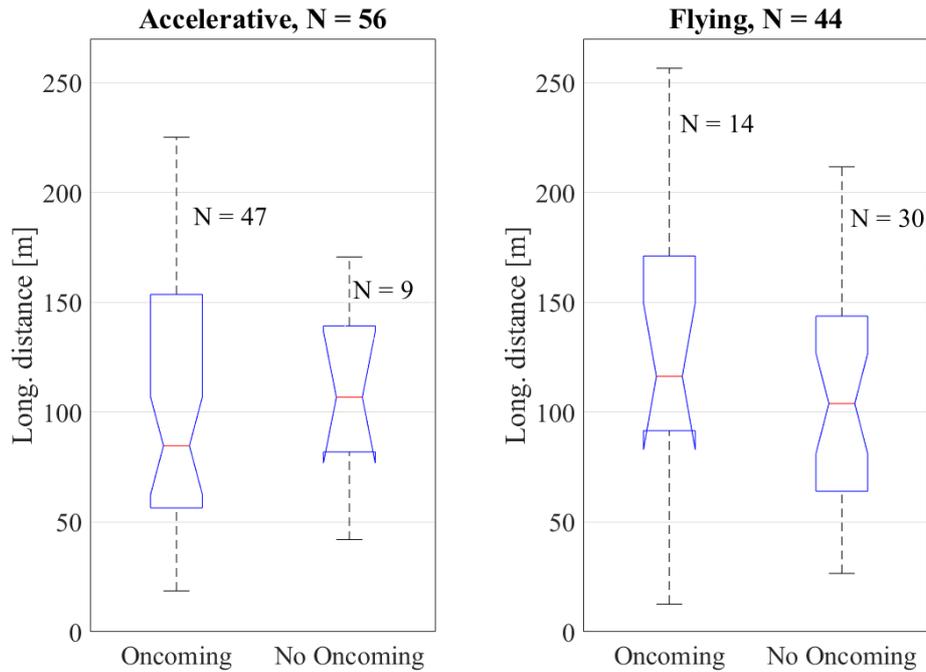


*Figure 13* Boxplot of longitudinal distance at the planning point, grouped per overtaking strategy. The edges of each box represent the first and the third quartile, the whiskers give the range of the observations, the centre line represents the median, the notched region represents the 95% confidence interval for the median.



*Figure 14* Boxplot of longitudinal distance at the planning point, grouped per presence of oncoming traffic. The edges of each box represent the first and the third quartile, the whiskers give the range of the observations, the centre line represents the median, the notched region represents the 95% confidence interval for the median.

Figure 15 provides the longitudinal distance at the planning point, grouped considering the interaction between overtaking strategy and presence of oncoming traffic. It can be noticed that the longitudinal distance at the planning point does not seem to depend on the overtaking strategy if there is no oncoming traffic. On the contrary, if there is oncoming traffic, the median of the longitudinal distance is higher for flying overtaking manoeuvres compared to accelerative overtaking manoeuvres.



*Figure 15* Boxplots of longitudinal distance at the planning point, grouped per overtaking strategy and presence of oncoming traffic. The edges of each box represent the first and the third quartile, the whiskers give the range of the observations, the centre line represents the median, the notched region represents the 95% confidence interval for the median.

Scatter plots of the longitudinal distance and the corresponding ego vehicle speed, grouped per interaction of overtaking strategy and presence of oncoming traffic, are provided in Figure 16 and Figure 17. The plotted lines are the result of a linear regression and are used to qualitatively identify an increasing or decreasing trend of the longitudinal distance with increasing ego vehicle speed. The fitting lines are plotted considering the overtaking strategy in Figure 16. The same scatter plot, with fitting lines plotted considering the presence of oncoming traffic, is provided in Figure 17. The longitudinal distance shows an increasing trend with the ego vehicle speed (roughly 150 m when the ego vehicle speed increases from approximately 25 km/h to approximately 110 km/h).

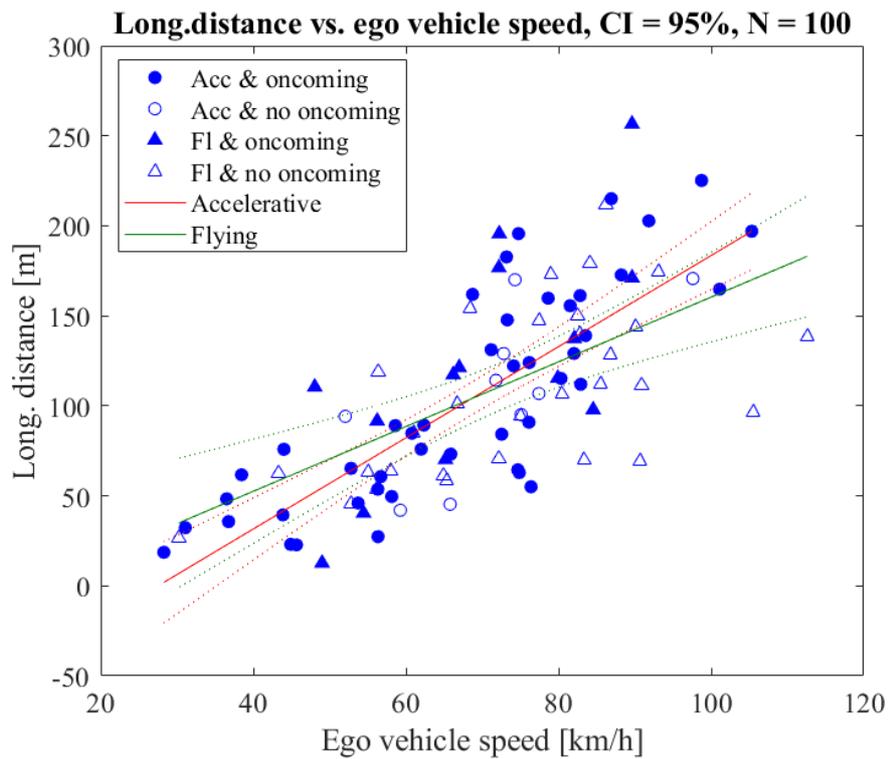


Figure 16 Scatter plot of the longitudinal distance at the planning point and the corresponding ego vehicle speed. Fitting lines, with 95% confidence interval (dotted), are plotted separately for flying and accelerative manoeuvres.

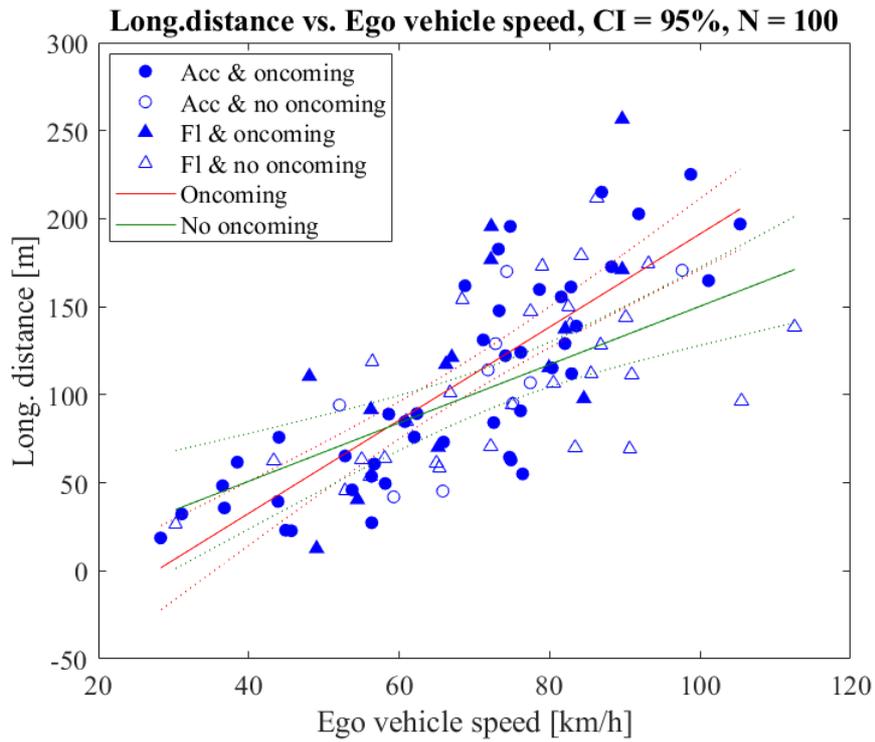


Figure 17 Scatter plot of the longitudinal distance at the planning point and the corresponding ego vehicle speed. Fitting lines are plotted separately for present and absent oncoming traffic.

## 4.4 TTC comparison between planning point and steer away point

TTC at the planning point is compared to TTC at the steer away point. A boxplot providing TTC at the planning point and at the steer away point is provided in Figure 18. Figure 19 (a) provides boxplots for the TTC at the planning point and at the steer away point grouped per oncoming traffic, Figure 19 (b) provides the same values grouped per overtaking strategy. From both figures it seems that the TTC at the steer away point is lower than TTC at the planning point (excluding the outliers), regardless of the strategy or the presence of oncoming traffic. The median of TTC at the steer away point is always below one second.

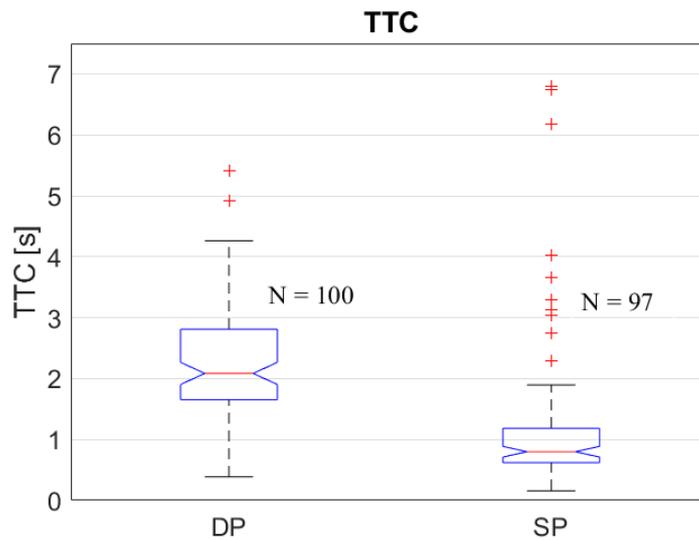


Figure 18 Boxplot of TTC at planning point (PP) and steer away point (SP). The edges of each box represent the first and the third quartile, the whiskers give the range of the observations, the centre line represents the median, the notched region represents the 95% confidence interval for the median, the plus symbols represent the outliers. Notice that three outliers were removed from the steer away point to ease comparison between the interquartile ranges.

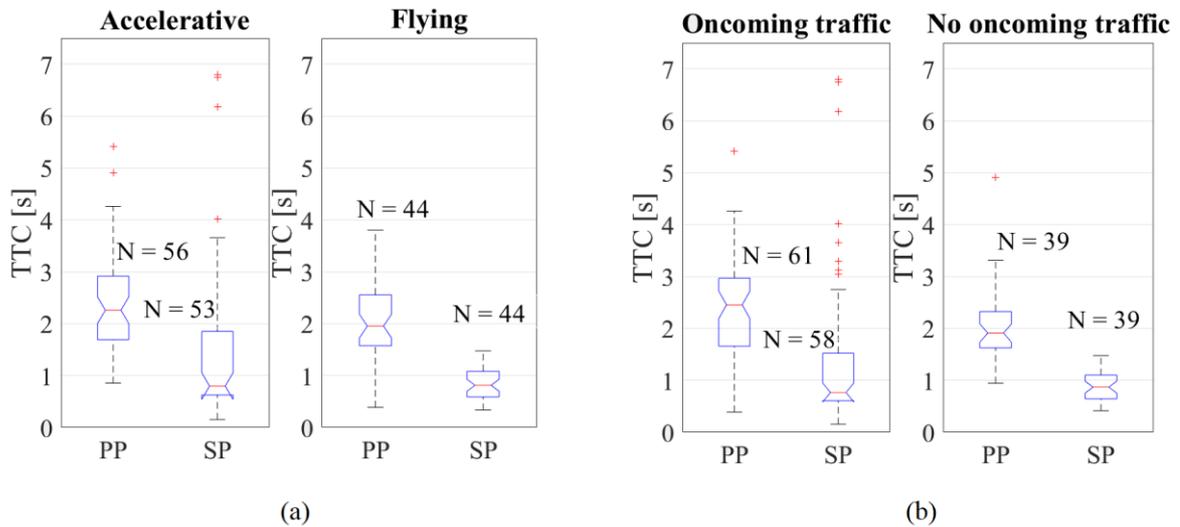


Figure 19 Time to collision at planning point (PP) and steer away point (SP). Data are grouped per overtaking strategy in Figure (a) and per presence of oncoming traffic in figure (b). The edges of each box represent the first and the third quartile, the whiskers give the range of the observations, the centre line represents the median, the notched region represents the 95% confidence interval for the median, the plus symbols represent the outliers. Three outliers were removed at the steer away point in case of accelerative manoeuvres and manoeuvres with oncoming traffic, to ease comparison between the interquartile ranges.

## 4.5 GLMM results

The fixed effects are provided for both the longitudinal distance and for the TTC at the planning point in Table 10 and in Table 11. Specifically, the tables provide the estimated value and the  $p$ -value of each of the fixed effects together with the ordinary and adjusted coefficient of correlation  $R^2$  and the Akaike Information Criterion (AIC). The effect “intercept” is the estimate of the response when all the influencing factors are null. The  $p$ -value is the result of a t-test with null hypothesis of null fixed effect. Smaller  $p$ -values indicate higher statistical significance of the given values.

The TTC at the planning point increases with an increase in the bicyclist speed, the interaction of ego vehicle speed with lateral distance, the interaction of ego vehicle speed with presence of oncoming traffic, and the interaction of lateral distance with driver’s age. TTC at the planning point decreases with an increase in the lateral distance, the ego vehicle speed, the interaction of the lateral distance with the driver’s gender and the interaction of the bicyclist speed with the driver’s age. TTC is higher for male drivers compared to female drivers and is lower when oncoming traffic is present compared to when it is absent. The three most influencing factors (where “most influencing” means higher absolute value of the fixed effect estimate compared to the others) are the bicyclist lateral distance, the presence of oncoming traffic and the driver’s gender.

The longitudinal distance at the planning point increases with an increase in the interaction of the ego vehicle speed with the lateral distance, the interaction between the ego vehicle speed and the presence of oncoming traffic and the interaction between the lateral distance and the driver’s age. The longitudinal distance at the planning point decreases with an increase in the lateral distance and the interaction between the bicyclist speed and the driver’s age. Moreover, the longitudinal distance is lower if oncoming traffic is present. The two most influencing

factors for the longitudinal distance are, similarly to TTC, the bicyclist lateral distance and the presence of oncoming traffic. The driver's gender does not show a significant influence on the longitudinal distance.

In both models, the overtaking strategy did not show any statistically significant influence on the response.

*Table 10 GLMM of the TTC at the planning point: fixed effects, adjusted and ordinary coefficient of correlation  $R^2$  and AIC. Level of significance was set to 0.05.*

Name	Estimate	p-value
Intercept	2.38	$9.7 \cdot 10^{-3}$
Lateral distance	-1.04	0.02
Bicyclist speed	0.09	$7.94 \cdot 10^{-4}$
Ego vehicle speed	-0.03	0.02
Oncoming traffic	-1.08	0.04
Male driver	1.42	$3.08 \cdot 10^{-3}$
Ego vehicle speed : lateral distance	0.01	0.03
Ego vehicle speed : oncoming traffic	0.02	$8.08 \cdot 10^{-3}$
Lateral distance : male driver	-0.75	$8.86 \cdot 10^{-4}$
Lateral distance : age	0.02	$2.09 \cdot 10^{-3}$
Bicyclist speed : age	-0.002	$7.68 \cdot 10^{-3}$
$R_{ord}^2 = 0.8370$ $R_{adj}^2 = 0.8186$ $AIC = 249.76$		

*Table 11 GLMM of the longitudinal distance at the planning point: fixed effects, adjusted and ordinary coefficient of correlation  $R^2$  and AIC. Level of significance was set to 0.05.*

Name	Estimate	p-value
Intercept	110.68	$5.86 \cdot 10^{-15}$
Lateral distance	-68.68	$6.03 \cdot 10^{-8}$
Oncoming traffic	-75.32	$2.11 \cdot 10^{-6}$
Ego vehicle speed : lateral distance	0.75	$5.30 \cdot 10^{-11}$
Ego vehicle speed : oncoming traffic	1.28	$2.50 \cdot 10^{-7}$
Lateral distance : age	0.54	$1.47 \cdot 10^{-3}$
Bicyclist speed : age	-0.03	$2.81 \cdot 10^{-5}$
$R_{ord}^2 = 0.8482$ $R_{adj}^2 = 0.8384$ $AIC = 984.03$		

The goodness of fit was briefly evaluated considering the coefficient of determination  $R^2$  for each of the models and by plotting the fitted values of the responses and the corresponding observed values, as in Figure 20 and Figure 21. What can be noticed from the figures is that the fitted values of the longitudinal distance and TTC, when plotted with the observed values, do not show a high dispersion around the red dotted line (a 45 degrees line representing where the fitted values are exactly equal to the observed values), supporting the high values of coefficient of correlation  $R^2$  (always higher than 0.8). Lower values of coefficient of correlation  $R^2$  for the TTC model compared to the longitudinal distance model could be explained considering that higher values of observed TTC were underestimated by

the model (in Figure 21, all the fitted values of TTC when the observed TTC is higher than 3 seconds are all placed below the red dotted line).

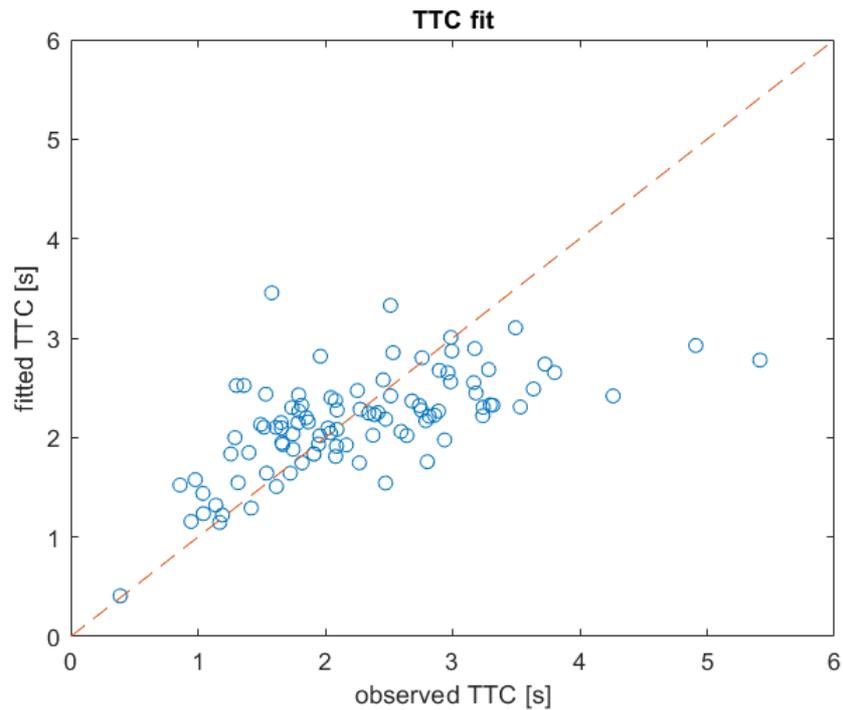


Figure 20 Comparison of the fitted and the observed values of TTC at the planning point. The red dotted line represents the region where fitted values are exactly equal to the observed values

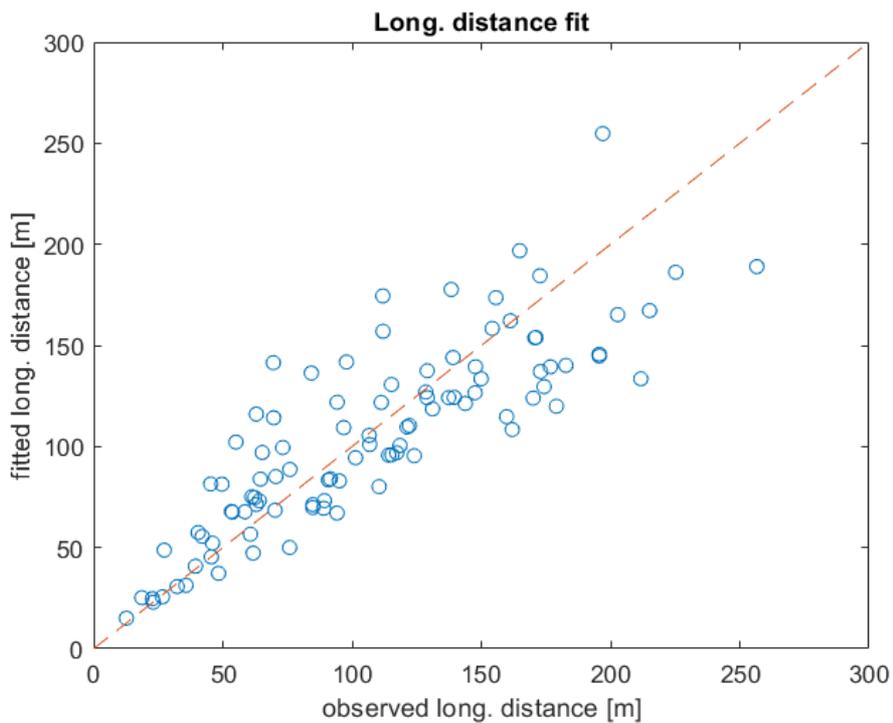


Figure 21 Comparison of the fitted and the observed values of longitudinal distance at the planning point. The red dotted line represents the region where fitted values are exactly equal to the observed value.

Appendix A and Appendix B provide additional Figures that allow to qualitatively see the influence of the driver's gender and the bicyclist lateral distance (factors highlighted by the GLMM in addition to the presence of oncoming traffic) on the responses. Figure 22 and Figure 25 respectively provide the TTC and the longitudinal distance at the planning point, grouped per driver's gender. As opposed to the result of the GLMMs, the driver's gender does not seem to influence the TTC and seems to influence the longitudinal distance at the planning point, which is higher for male drivers compared to female drivers. Scatter plots of the TTC at the planning point and the corresponding bicyclist lateral distance are provided in Figure 23 and Figure 24: the observations are grouped per presence of oncoming traffic and overtaking strategy, the fitting lines are plotted considering the overtaking strategy (Figure 23) or the presence of oncoming traffic (Figure 24). The TTC seems to increase with the bicyclist lateral distance, as opposed to the result of the GLMM. Scatter plots of the longitudinal distance at the planning point and the corresponding bicyclist lateral distance are provided in Figure 26 and Figure 27: the observations are grouped per presence of oncoming traffic and overtaking strategy, the fitting lines are plotted considering the overtaking strategy (Figure 26) or the presence of oncoming traffic (Figure 27). The longitudinal distance seems to show an increasing trend with the bicyclist lateral distance, as opposed to the result of the GLMM.

The explanation for the differences between the results of the GLMMs and the Figures in Appendix A and Appendix B is that boxplots and scatterplots do not consider all the factors together, justifying the need of a statistical model that considers all the factors.

## 5 Discussion

This chapter provides an interpretation of the results and comments about the methodology. Section 5.1 is a discussion about the identification of the planning point, Section 5.2 is a qualitative analysis of the results, Section 5.3 provides a discussion about the GLMMs and an interpretation of the output of the GLMMs, Section 5.4 is a discussion about the main limitations of this thesis.

### 5.1 When is the planning point?

The identification of the planning point was performed during manual annotation of the overtaking manoeuvres. The two indicators that allowed to identify the planning point were the eyes and the pedals only, without including turn indicator and steering wheel angle (see Section 3.4.2). In particular, eyes were used to check the surroundings and to assess if it was allowed and if there was the opportunity to overtake (see Section 2.2). Regarding the pedals, the analysed drivers used to release the gas pedal before checking the surroundings, for instance to have more time to decide whether to overtake or not and to be ready in case the decision is to brake and follow the bicyclist. In case of flying overtaking manoeuvres, after releasing the gas pedal for a small amount of time (roughly no more than one second), the driver pressed the gas pedal again to perform the overtaking without considerably changing the speed. For accelerative overtaking manoeuvres the gas pedal was not pressed, and the brake was possibly used, depending on the ego vehicle speed at the planning point.

Even if the turn indicator was initially considered as a possible planning point indicator (see Section 3.4.2), the annotation performed on the validated segments showed that the turn indicator was activated only after planning to overtake, when the decision is made (see Section 2.2). Similar principle applies for the steering wheel angle: it is a clear indicator of the steer away point, that could be coincident with the decision point if the driver decides to immediately change its lane position without getting closer to the bicyclist, but not of the planning point.

### 5.2 Qualitative analysis

A qualitative analysis can be performed by looking at the figures provided in Section 4.2, Section 4.3 and Section 4.4. What can be noticed from Figure 9 is that, although the higher number of observed overtaking manoeuvres with oncoming traffic (61 out of 100), the TTC at the planning point has higher dispersion when there is oncoming traffic compared to when there is no oncoming traffic. An explanation for this could be that the planning point depends on when the oncoming traffic is perceived as well as on its speed and distance (not analysed in this thesis), while if there is no oncoming traffic the driver is free to steer away as soon as possible, so that the planning does not have any dependence on other factors that could make it more dispersed. The same interpretation applies when comparing the longitudinal distance with and without oncoming traffic (see Figure 14).

When the longitudinal distance at the planning point is analysed considering the interaction between overtaking strategy and presence of oncoming traffic (see Figure 15), it can be noticed that the presence of oncoming traffic seems to have an opposite effect depending on the strategy: for accelerative overtaking manoeuvres, the presence of oncoming traffic causes a decrease of the median compared to when there

is no oncoming traffic; vice versa, for flying overtaking manoeuvres, the presence of oncoming traffic causes an increase of the median compared to when there is no oncoming traffic. The median values of both the strategies with no oncoming traffic are close (both medians are slightly above 100 m). These values are in accordance to the assumption of 100 m for the longitudinal distance at which the driver is deciding the overtaking strategy (Farah, H., et al., 2019). Regarding the overtaking manoeuvres with oncoming traffic, the inter-quartile range of longitudinal distance for flying overtaking manoeuvres is almost fully included in the inter-quartile range of longitudinal distance for accelerative overtaking manoeuvres.

Unlike the longitudinal distance, the presence of oncoming traffic seems to have the same effect on the TTC at the planning point regardless of the strategy (see Figure 10): for both strategies, oncoming traffic seems to increase the median value of TTC as well as its dispersion.

While the longitudinal distance at the planning point has a clear increasing trend with increasing ego vehicle speed (see Figure 16 and Figure 17), also confirmed by the results of the GLMM (see Table 11, where the ego vehicle speed appears in the interaction with bicyclist lateral distance or with presence of oncoming traffic), it is difficult to identify a trend for the TTC at the planning point by just looking at the scatter plots (see Figure 11 and Figure 12), since the slope of the fitting lines changes its sign depending on how the observed values of TTC are grouped.

The TTC at the steer away point has a lower median and a lower dispersion compared to the TTC at the planning point, as shown in Figure 18. The outliers were far above the value of seven seconds at the steer away point. The main reason for this is that, while the longitudinal distance has a finite positive value at the steer away point, the ego vehicle speed and the bicyclist speed might be close, causing the denominator of TTC to be close to zero, causing an overall high value of TTC. This happens when the driver has to follow the driver for a while before steering away, hence outliers with considerably high value of TTC at the steer away point only appear in accelerative overtaking manoeuvres, or in overtaking manoeuvres with oncoming traffic (see Figure 19). The TTC at the steer away point has a higher dispersion for accelerative overtaking manoeuvres or when oncoming traffic is present.

### 5.3 GLMM

GLMM was chosen to model TTC and longitudinal distance at the planning point for three reasons.

The first reason is the assumption of linearity (*linear* model), so that the responses are based on a linear combination of the factors (see Section 3.6): this is a starting point, and was assumed to be true by default. A non-linear model would have been chosen if the results with the assumption of linearity were not acceptable or if more time was allotted to this thesis.

The second reason why GLMM was chosen is that it is a *mixed* model: giving as an input all the factors together, including the random effect due to the driver, was crucial to better estimate the fixed effects. This can be seen considering how the results from the GLMM were not visible from a qualitative analysis of the results.

The third reason why GLMM was used is that it allows to choose a distribution of the responses when creating the model (*generalised* model): it was

important to choose a distribution defined only for positive values, since the observed values of the responses (TTC and longitudinal distance at the planning point) were always positive. The tested distributions for each response were a gamma distribution and an inverse gaussian distribution: the distribution that gave the highest coefficient of correlation  $R^2$  was chosen to model the response. The best distribution was a gamma distribution for the longitudinal distance at the planning point and an inverse gaussian distribution for the TTC at the planning point. The normal distribution was also tested: it resulted in considerably lower values of coefficient of correlation  $R^2$ , justifying the use of a *generalised* model.

The chosen link function was an identity function: the mean of the response was set equal to the linear predictor. This choice was made to ease the interpretation of the fixed effects, but it also improved the model (higher coefficient of correlation  $R^2$ ) compared to the canonical link function, used by default by the MATLAB<sup>®</sup> function *fitglme* according to the chosen distribution of the response (MathWorks, fitglme, 2019).

Follows a critical interpretation of the results from the GLMM. The fixed effects for TTC and longitudinal distance at the planning point are provided in Table 10 and Table 11.

### 5.3.1 Lateral distance

Both the longitudinal distance and the TTC at the planning point considerably decrease as the bicyclist lateral distance increases (TTC is 1.04 s lower and longitudinal distance is 68.68 m lower when the bicyclist lateral distance increases of 1 m). A possible explanation for this could be that the bicyclist was perceived later if its lateral distance was higher, thus the driver started planning to overtake when closer to the bicyclist both in terms of longitudinal distance and TTC.

### 5.3.2 Bicyclist speed

An increase in bicyclist speed causes a decrease of the longitudinal distance at the planning point if the interaction with the age is considered: this means that the driver plans to overtake when closer to the bicyclist, and this effect is more pronounced for increasing driver's age. Given a fixed driver's age, an explanation for how an increase in bicyclist speed causes a decrease in longitudinal distance could be that, if high speed of the bicyclist was perceived, the driver could afford getting closer to the bicyclist before planning to overtake. Increasing bicyclist speed has an opposite effect on TTC, which increases as the bicyclist travels faster. This can be explained by considering that, even if the longitudinal distance at the planning point decreases with the bicyclist speed, the relative velocity at the denominator of TTC decreases as well. Hence, the decrease in the relative speed prevails over the decrease in longitudinal distance, causing an overall increase of TTC.

### 5.3.3 Ego vehicle speed

The ego vehicle speed has no significant influence on the longitudinal distance at the planning point but influences the TTC at the planning point: if the ego vehicle speed increases of 1 km/h, TTC decreases of 0.03 s. This could be explained considering that, while the longitudinal distance is not influenced by this factor, the relative speed

at the denominator of the TTC increases with the ego vehicle speed, causing an overall decrease of TTC (see equation (9)).

### 5.3.4 Presence of oncoming traffic

The presence of oncoming traffic causes a considerable decrease of the TTC and of the longitudinal distance at the planning point. TTC decreases of 1.08 s and longitudinal distance decreases of 75.32 m when oncoming traffic is present. An explanation for this could be that, if oncoming traffic is perceived, the driver decides to wait before planning to overtake, having a smaller TTC and longitudinal distance. It is worth mentioning that the influence of oncoming traffic on TTC is considerably less significant ( $p$ -value = 0.04) compared to the influence on the longitudinal distance ( $p$ -value =  $2.11 \cdot 10^{-6}$ ).

The influence of oncoming traffic on the overtaking manoeuvres was also noticed when studying the Comfort Zone Boundaries (CZBs). Oncoming traffic caused a reduction in minimum lateral clearance (Dozza, M., et al., 2015) (Kovaceva, J., et al., 2018): the results of this thesis provide that oncoming traffic has an analogous reduction effect on the longitudinal distance and TTC at the planning point.

### 5.3.5 Overtaking strategy

The overtaking strategy did not show any significant influence on both TTC and longitudinal distance at the planning point. An explanation for this could be that most of the accelerative overtaking manoeuvres were characterized by presence of oncoming traffic and, vice versa, most of the flying overtaking manoeuvres were characterized by absence of oncoming traffic (see Section 4.1, Table 7). While the overtaking strategy did not show any influence at the planning point, it made a difference when analysing the CZBs (Dozza, M., et al., 2015). This might be a proof that the overtaking strategy is chosen while planning, thus it does not affect the planning point but affects the CZBs.

### 5.3.6 Gender

The gender influences the TTC at the planning point. The observed values of TTC at the planning point were higher for male drivers compared to female drivers, meaning that male drivers start planning earlier. This is in contrast with previous studies (Farah, 2011) showing that male drivers are more aggressive than female drivers. A reason for this contrast might be the small number of female drivers involved in the analysis (13 only). An analysis involving more drivers, with a similar amount of male and female drivers, could make the results more accurate and possibly similar to the findings in the literature (Farah, 2011). Nevertheless, 26 out of 39 overtaking manoeuvres performed by female drivers were accelerative, while 30 out of 61 overtaking manoeuvres performed by male drivers were flying, showing that female drivers preferred to choose an accelerative overtaking strategy, in accordance to the literature (Farah, 2011).

### 5.3.7 Age

The age appears in two-way interactions in the model for TTC. TTC increases with the increase of two-way interaction between lateral distance and age and decreases

with the two-way interaction between bicyclist speed and age. It can be noticed that the coefficients for the interaction of bicyclist speed with the age and lateral distance with the age are low compared to the single fixed effects (bicyclist speed and lateral distance).

## **5.4 Limitations**

The main limitations of this thesis consist in the small dataset (100 overtaking manoeuvres used in the statistical analysis), in subjective error during manual annotation (see Section 5.4.1) and in the assumptions needed to extract data at the planning point (see Section 5.4.2).

### **5.4.1 Manual annotation**

Manual annotation of the overtaking manoeuvres might cause a strong influence of the annotator's subjective perception on the results. In an attempt to make the results as objective as possible, new definitions were given for the purpose of annotating (see Section 3.4). This could have led to some bias in the results. For instance, overtaking manoeuvres in which the oncoming traffic appeared only after the passing phase: in such a case, it is difficult to assess whether the oncoming traffic had an influence on the planning point or not. Oncoming traffic could have an influence on the planning point if it was perceived before the planning point. Vice versa, if oncoming traffic was too far or if the infrastructure did not make it visible before the planning point, oncoming traffic could not have an influence on it. Another example regards the annotation of the overtaking strategy. In some overtaking manoeuvres the driver reduced the speed but steered away without following the bicyclist. This kind of overtaking is a borderline case, close to a flying overtaking because the driver does not follow the bicyclist, but also close to an accelerative because the driver is reducing its speed. Some overtaking manoeuvres that were flying in principle could have been classified as accelerative, or vice versa, to be consistent with the definition given in Section 3.4.1.

### **5.4.2 Data extraction**

As discussed in Section 3.5, data from the ego vehicle was always defined, but data about the bicyclist was not available at the planning point. Hence, some assumptions were done to extract data about the bicyclist at the planning point.

The bicyclist speed was assumed to be constant. This assumption is reasonable for a bicyclist on rural roads, as the bicyclist is just travelling straight without encountering any intersection or pedestrian crossing typical of an urban environment. One drawback of this assumption is that the bicyclist might have changed its speed as the ego vehicle approached.

The lateral distance was calculated assuming constant lane position of the bicyclist, using the lane position calculated during bicyclist detection from MobilEye, when the ego vehicle was very close to the bicyclist. If allowed, the bicyclist could have changed its lane position to give more space to the driver to perform the overtaking. Hence, the lateral distance at the planning point might have been overestimated because of this effect. Moreover, the pressure zone created by the ego vehicle might have caused lateral oscillations of the bicyclist when the ego vehicle was close to it (Nero, 2017) (Schindler, R., Bast, V., 2015) (Kato, Y., et al., 1981).

The error due to the oscillations might have been reduced by the time average, as described in Section 3.5.5. The assumption of lateral distance at the planning point equal to lateral distance at the steer away point is reasonable, since the steer away point was defined as the moment when the vehicle starts diverging from the original lane position: a possible difference in lane position between planning point and steer away point might only be due to small trajectory adjustments performed by the driver, which can be considered negligible.

The heading angle was used to rotate the bicyclist longitudinal and lateral distance given by MobilEye in a car-fixed coordinate system before extracting the values of bicyclist speed, longitudinal distance and lateral distance at the planning point. The underlying assumption in the calculation of the heading angle as in Section 3.5.2 is that the road is straight from the steer away point to the bicyclist disappearance. This is not an issue in case of straight roads but might cause errors if the overtaking was performed while travelling on a curve. However, all the analysed overtaking manoeuvres that were not performed on a straight road were performed on curves with a visibly high turning radius: hence, the effect of the curvature on the heading angle was negligible with respect to the effect of the overtaking manoeuvre.

The overtaking manoeuvres without yaw rate were analysed assuming negligible heading angle. This is a strong assumption, made to include as many overtaking manoeuvres as possible in the statistical analysis. An analysis was made on a few overtaking manoeuvres where the heading angle was available to check if this assumption strongly affected the results (the bicyclist speed and the bicyclist lateral distance). The difference between the results calculated with and without using the heading angle was relatively low, meaning that it has not a strong influence on the results.

## 6 Conclusion and future works

This thesis provided a methodology for the extraction of bicyclist overtaking manoeuvres on rural roads and for the extraction and estimation of TTC and longitudinal distance at the planning point, from naturalistic driving data.

Data reduction allowed to identify bicyclist overtaking manoeuvres on rural roads in the database. Continuing, manual annotation and data extraction were performed to collect the data needed to estimate TTC and longitudinal distance at the planning point. After data extraction, and after removing observations with non-valid data (for instance, non-realistic bicyclist speed), a statistical analysis was performed on the collected data.

GLMMs of TTC and longitudinal distance at the planning point were developed to understand which factors influenced them and to quantify their influence. The conclusion from the GLMMs of TTC and longitudinal distance at the planning point is that they both considerably decrease with increasing lateral distance of the bicyclist and they both considerably decrease if oncoming traffic is present. Moreover, TTC at the planning point is higher for male drivers than female drivers.

Knowledge of TTC at the planning point could contribute in the improvement of active safety systems by making them less conservative. Knowing when the driver is planning to overtake could help to understand how to prevent driver annoyance when AEB is activated (activation of AEB while the driver is still planning would be perceived as too early). Moreover, knowledge of TTC at the planning point might be useful to assess if the driver did not start planning to overtake yet, even if planning was expected to occur, and could be used by the safety systems algorithms as a metric to decide when to intervene. The long-term scope of TTC estimation is to increase the efficiency of AEB algorithms to reduce bicyclist fatalities in rear-end collisions, typical of overtaking scenarios on rural roads, where the high speed difference increases the risk of collisions resulting in severe injuries or fatalities.

### 6.1 Future works

This thesis consisted in extracting and estimating the TTC and the longitudinal distance at the planning point during an overtaking of a bicyclist using NDD from the UDRIVE project.

During manual annotation, the planning point was annotated as in Section 3.4.2: this thesis focused only on the first evidence that the driver is planning to overtake. It would be interesting to annotate all the points in which one different action is performed, to have a wider knowledge of what the drivers do during and after the decision-making process.

Future studies might also collect more overtaking manoeuvres to include in the statistical analysis: specifically, more flying overtaking manoeuvres with oncoming traffic and more accelerative overtaking manoeuvres without oncoming traffic could be collected to make the population more homogeneous.

The set of possible influencing factors in the GLMM could be extended, including more accurate description of oncoming traffic (which, in this thesis, was just a Boolean): for instance, TTC or longitudinal distance of the oncoming traffic could be used. The goodness of fit of the proposed model can be better assessed using more advanced methods than just looking at the coefficient of correlation  $R^2$ . Moreover, different statistical models, such as nonlinear regressive models, could be possibly used to either confirm or increase the accuracy of the results obtained using GLMM.

## 7 Bibliography

- Boda, C.-N., et al. (2018). Modelling how drivers respond to a bicyclist crossing their path at an intersection: How do test track and driving simulator compare? *Accident Analysis and Prevention*, 111, 238 - 250.
- Boufous, S., et al. (2012). Risk factors for severe injury in cyclists involved in traffic crashes in Victoria, Australia. *Accident Analysis and Prevention*, 49, 404 - 409.
- Bärgman, J., et al. (2017). *The UDrive dataset and key analysis results. UDRIVE Deliverable D41.1.*
- Castermans, J. (2017). *Overview of the Data Collection. UDRIVE Deliverable D30.1.*
- Collins Dictionary. (2019). *Overtaking*. Retrieved from <https://www.collinsdictionary.com/dictionary/english/overtaking>
- Dozza, M., et al. (2015). How do drivers overtake cyclists? *Accident Analysis and Prevention*, 88, 29 - 36.
- Duan, J., et al. (2017). Driver braking behavior analysis to improve autonomous emergency braking systems in typical Chinese vehicle-bicycle conflicts. *Accident Analysis and Prevention*, 74 - 82.
- EuroNCAP. (2018). *AEB Cyclist*. Retrieved from <https://www.euroncap.com/en/vehicle-safety/the-ratings-explained/vulnerable-road-user-vru-protection/aeb-cyclist/>
- European Commission. (June 2018). *Traffic Safety Basic Facts on Cyclists*. European Commission, Directorate General for Transport.
- Farah, H. (2011). Age and gender differences in overtaking maneuvers on two-lane rural highways. *Transportation Research Record: Journal of the Transportation Research Board* No. 2248, *Transportation Research Board of the National Academies, Washington*, 30 - 36.
- Farah, H., et al. (2019). Modelling overtaking strategy and lateral distance in car-to-cyclist overtaking on rural roads: A driving simulator experiment. *Transportation Research Part F*, 226 - 239.
- Feng, F., Bao, S. (2018). Drivers overtaking bicyclists - An examination using naturalistic driving data. *Accident Analysis and Prevention*, 115, 98 - 109.
- Hegeman, G., et al. (2005). Opportunities of advanced driver assistance systems towards overtaking. *European journal of transport and infrastructure research EJTI*, 5(4), 281 - 296.
- Kato, Y., et al. (1981). *Aerodynamic Effects to a Bicycle Caused by a Passing Vehicle*. Society of automotive engineers, inc.
- Kovaceva, J., et al. (2018, September 8). *Safety science*. Retrieved from <https://doi.org/10.1016/j.ssci.2018.08.022>
- MathWorks. (2019). *fitglm*. Retrieved from <https://it.mathworks.com/help/stats/fitglm.html>
- MathWorks. (2019). *Generalized Linear Mixed-Effects Models*. Retrieved from <https://it.mathworks.com/help/stats/generalized-linear-mixed-effects-models.html>
- MathWorks. (2019). *Polynomial Curve Fitting*. Retrieved from <https://it.mathworks.com/help/matlab/math/polynomial-curve-fitting.html>
- Matson, T. M. , et al. (1939). Overtaking and passing requirements as determined from a moving vehicle. *Highway Research Board Proceedings*, 18.

- Nero, G. (2017). *Quantifying Drivers' Behaviours when Overtaking Bicyclists on Rural Roads, Master's Thesis*. Gothenburg: Chalmers University of Technology.
- Observed vs Unobserved variables [duplicate]*. (n.d.). Retrieved from <https://stats.stackexchange.com/questions/159886/observed-vs-unobserved-variables>
- Panero, G. (2018). *Drivers' comfort zone boundaries when overtaking pedestrians, Master's Thesis*. Gothenburg: Chalmers University of Technology.
- Pucher, J., Dijkstra, L. . (2003). Promoting Safe Walking and Cycling to Improve Public Health: Lessons From The Netherlands and Germany. *American Journal of Public Health*, 93, 1509 - 1513.
- Pucher, J., et al. . (2011). Bicycling renaissance in North America? An update and re-appraisal of cycling trends and policies. *Transportation Research*, 93, 451 - 475.
- Rasch, A. (2018). *Modelling driver behaviour in longitudinal vehicle-pedestrian scenarios, Master's thesis*. Gothenburg: Chalmers University of Technology.
- Räsänen, M., Summala, H. (1998). Attention and expectation problems in bicycle-car collisions: an in-depth study. *Accident analysis and prevention*, 30, 657–666.
- Schindler, R., Bast, V. (2015). *Drivers' comfort boundaries when overtaking a cyclist, Master's thesis*. Gothenburg: Chalmers University of Technology.
- SWOV. (2012, July). *Vulnerable road users*. Retrieved from [https://www.swov.nl/sites/default/files/publicaties/gearchiveerde-factsheet/uk/fs\\_vulnerable\\_road\\_users\\_archived.pdf](https://www.swov.nl/sites/default/files/publicaties/gearchiveerde-factsheet/uk/fs_vulnerable_road_users_archived.pdf)
- The Royal Society for the Prevention of Accidents. (2017, November). *Road Safety Factsheet*. Retrieved from <https://www.rospa.com/rospaweb/docs/advice-services/road-safety/cyclists/cycling-accidents-factsheet.pdf>
- What is a reasonable speed for long distances on a bike?* (2011). Retrieved 04 2019, from <https://bicycles.stackexchange.com/questions/1103/what-is-a-reasonable-speed-for-long-distances-on-a-bike>
- Wilson, T., Best, W. (1982). Driving strategies in overtaking. *Accident Analysis and Prevention*, 14(3), 179 - 185.
- Zago, M., et al. (2008). Visuo-motor coordination and internal models for object interception. *Exp Brain Res*, 571 - 604.

# Appendices

## A Time to Collision at the planning point – additional figures

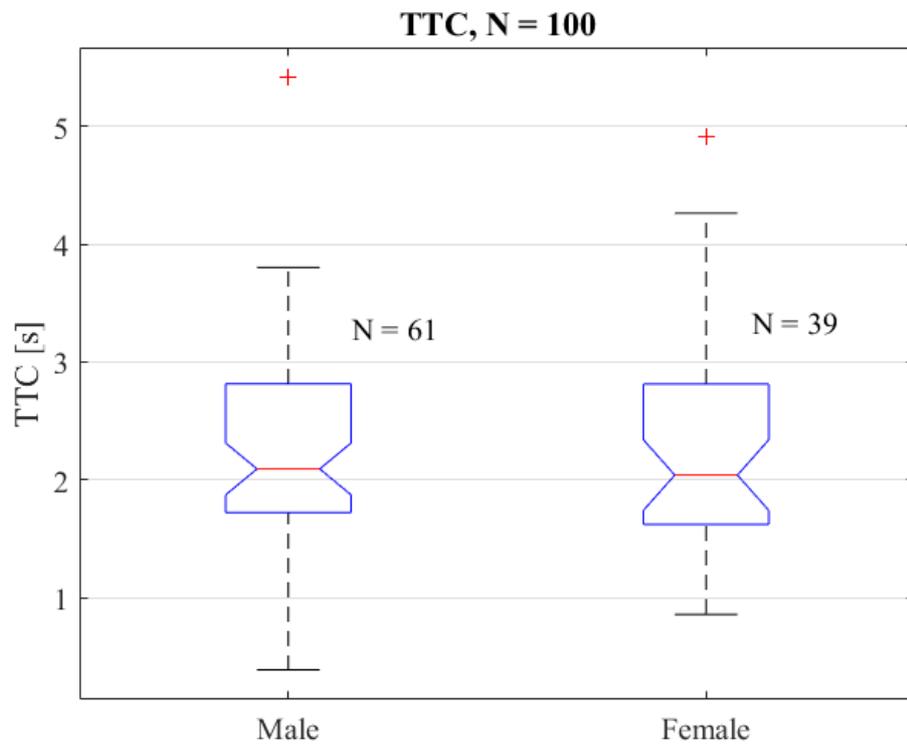


Figure 22 *TTC at the planning point, grouped per driver's gender. The edges of each box represent the first and the third quartile, the whiskers give the range of the observations, the centre line represents the median, the notched region represents the 95% confidence interval for the median, the plus symbols represent the outliers.*

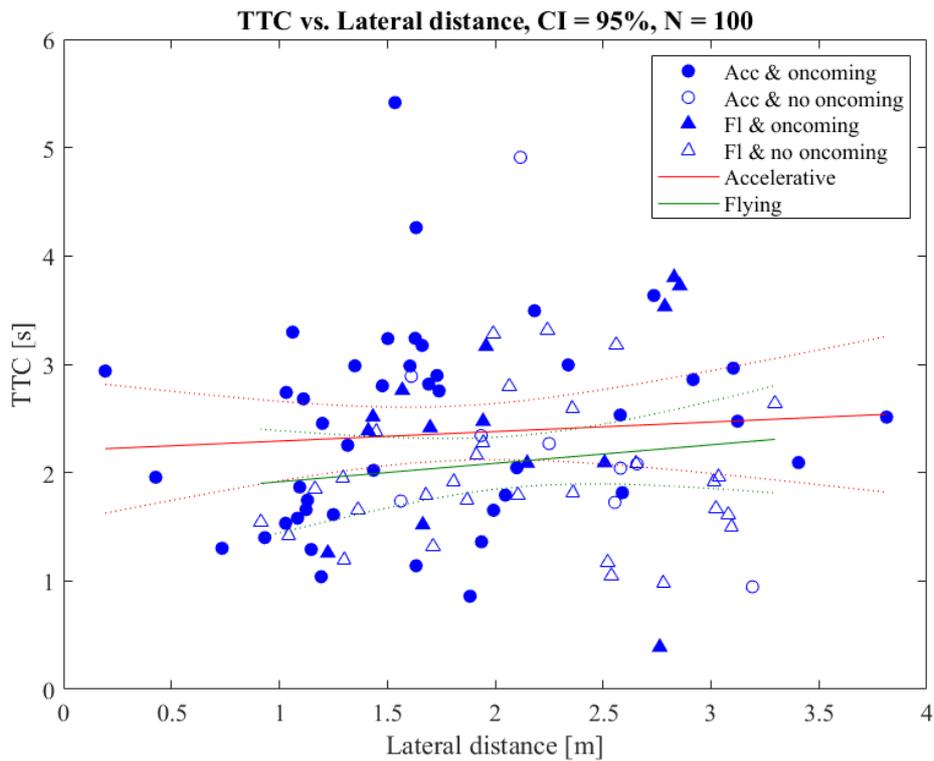


Figure 23 Scatter plot of TTC at the planning point and the corresponding bicyclist lateral distance. Fitting lines, with corresponding 95% confidence intervals (dotted lines), are plotted separately for flying and accelerative manoeuvres

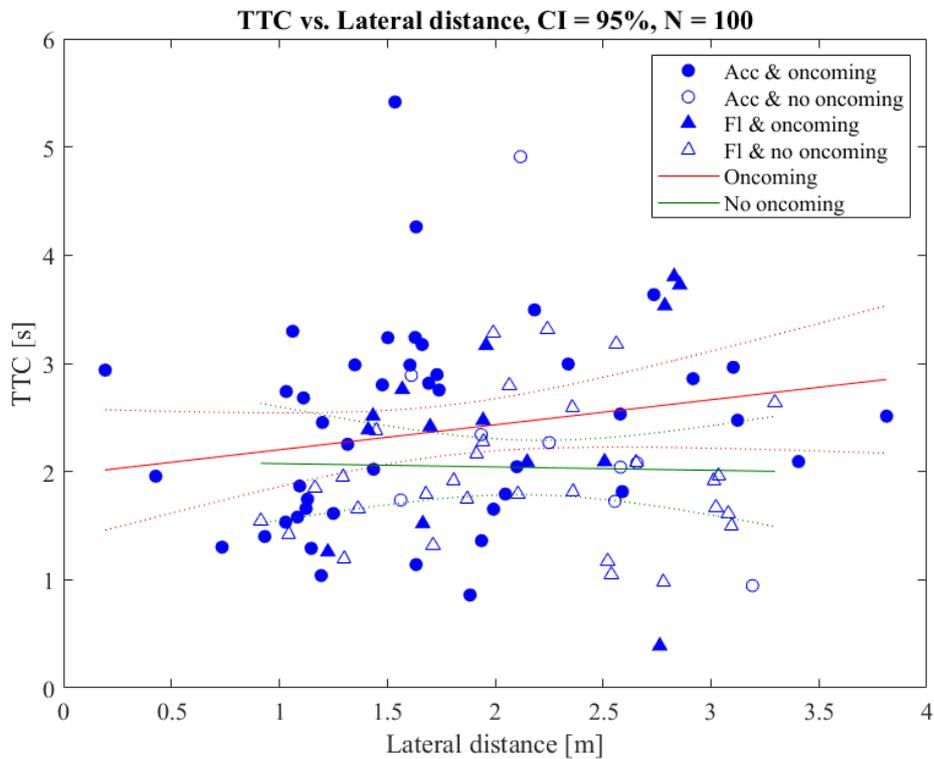
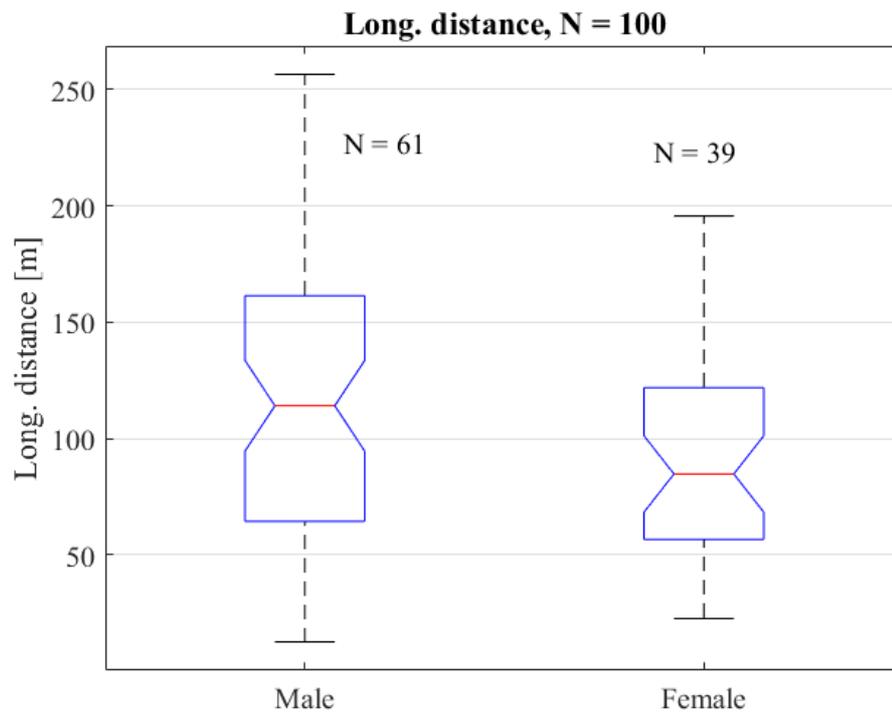


Figure 24 Scatter plot of TTC at the planning point and the corresponding bicyclist lateral distance. Fitting lines, with corresponding 95% confidence intervals (dotted lines), are plotted separately for present and absent oncoming traffic.

## B Longitudinal distance at the planning point – additional figures



*Figure 25 Longitudinal distance at the planning point, grouped per driver's gender. The edges of each box represent the first and the third quartile, the whiskers give the range of the observations, the centre line represents the median, the notched region represents the 95% confidence interval for the median.*

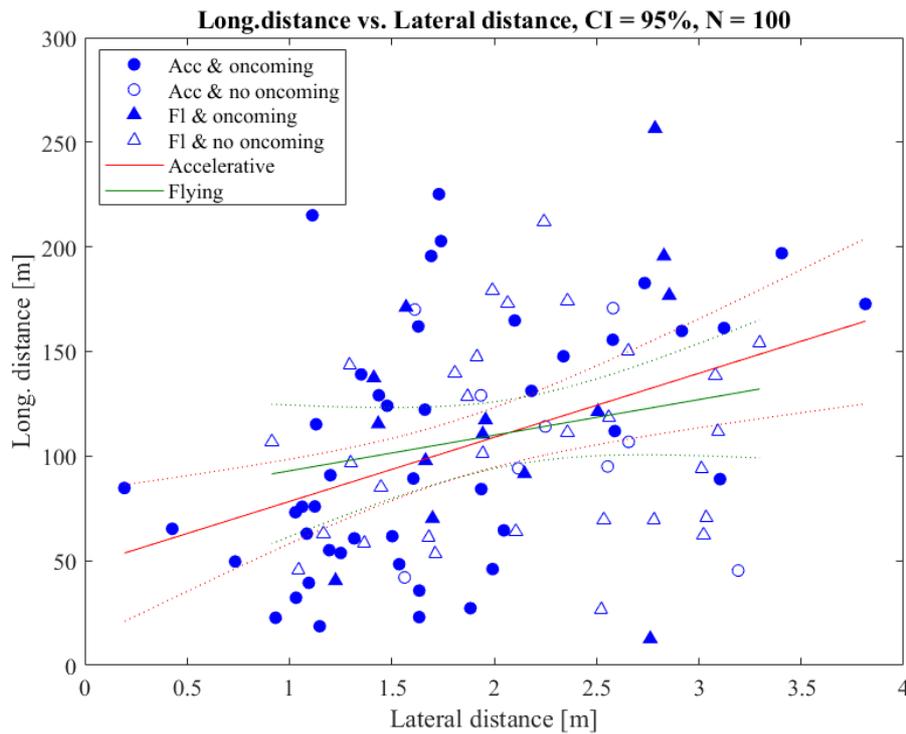


Figure 26 Scatter plot of longitudinal distance at the planning point and the corresponding bicyclist lateral distance. Fitting lines, with corresponding 95% confidence intervals (dotted lines), are plotted separately for flying and accelerative manoeuvres.

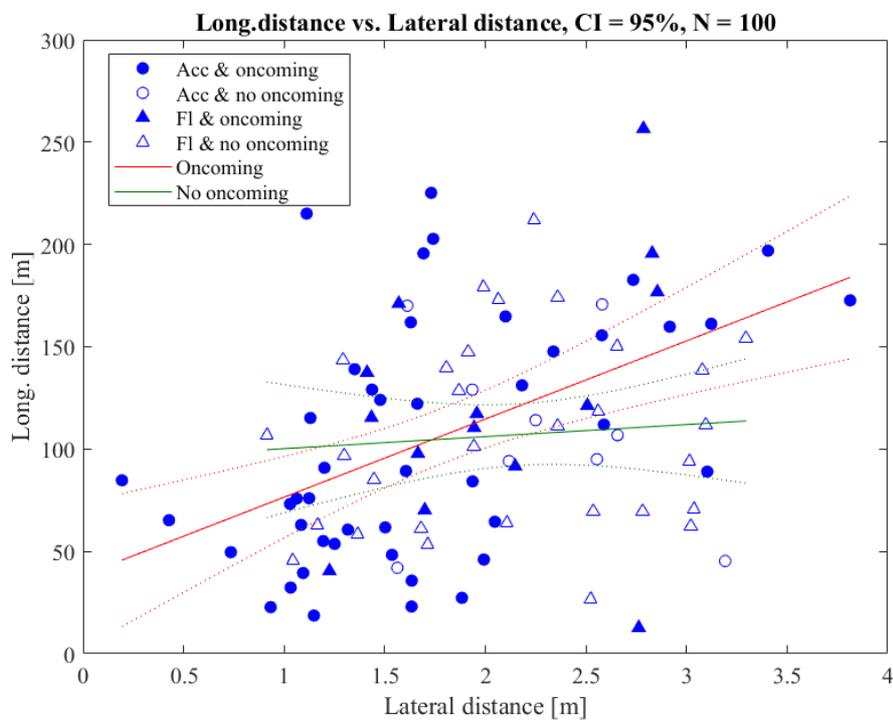


Figure 27 Scatter plot of longitudinal distance at the planning point and the corresponding bicyclist lateral distance. Fitting lines, with corresponding 95% confidence intervals (dotted lines), are plotted separately for present and absent oncoming traffic.