Understanding and modelling car drivers overtaking cyclists: Toward the inclusion of driver models in virtual safety assessment of advanced driving assistance systems

JORDANKA KOVACEVA

Department of Mechanics and Maritime Sciences
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2019
Understanding and modelling car drivers overtaking cyclists: Toward the inclusion of driver models in virtual safety assessment of advanced driving assistance systems

Jordanka Kovaceva
Understanding and modelling car drivers overtaking cyclists: Toward the inclusion of driver models in virtual safety assessment of advanced driving assistance systems

JORDANKA KOVACEVA

© JORDANKA KOVACEVA, 2019.

THESIS FOR LICENTIATE OF ENGINEERING
Technical report No. 2019:17

Department of Mechanics and Maritime Sciences
Chalmers University of Technology
SE–412 96 Göteborg, Sweden
Telephone + 46 (0) 31 – 772 1000

Typeset by the author using \LaTeX.

Printed by Chalmers Reproservice
Göteborg, Sweden 2019
To my beloved
Nikola
Abstract

The total number of road crashes in Europe is decreasing, but the number of crashes involving cyclists is not decreasing at the same rate. To help car drivers avoid or mitigate crashes while overtaking a cyclist, advanced driver assistance systems (ADAS) have been developed. To evaluate and further improve these ADAS to support drivers as they overtake cyclists, we need to understand and model driver behaviours.

This thesis has two objectives: 1) to extract and analyse cyclist-overtaking from naturalistic driving data and 2) compare driver behaviour models for overtaking manoeuvres that can be used in counterfactual simulations for evaluating ADAS safety benefits.

The drivers’ comfort zone boundaries (CZBs) when overtaking a cyclist were identified and analysed using naturalistic driving data. Three driver models that predict when a car driver starts steering away in order to overtake a cyclist were implemented: a threshold model, an evidence accumulation model, and a model inspired by a proportional-integral-derivative controller. These models were tested and verified using two different datasets, one from a test-track experiment and one from naturalistic driving data. Model parameters were obtained using a computationally efficient linear programming.

The results show that, when an oncoming vehicle was present, the drivers were significantly closer to the cyclist before steering away. This finding confirms that the presence of an oncoming vehicle is a crucial factor for the safety of the cyclist and needs to be taken into account for the development of ADAS that maintain safe distance to the cyclist. Furthermore, the quantification of the CZBs has implications for the development of ADAS which can estimate the time-to-collision to an oncoming vehicle or a cyclist to be overtaken, providing timely and acceptable warnings—or interventions—when drivers exceed their usual CZBs. A comparison of models shows that all three models are highly variable in detecting steering away time for different drivers. Furthermore, differences were discovered in detected steering away time between models fitted to test-track experiments and naturalistic driving data. Future work may focus on using larger, more diverse datasets and investigating more advanced models before including them in counterfactual simulations.

Keywords: Traffic safety, overtaking manoeuvres, cyclist, safety benefit, naturalistic data.
Acknowledgments

I would like to express my gratitude to my supervisors Marco Dozza and Jonas Bärgman for the guidance and the support throughout the studies.

My sincere gratitude to the colleagues and friends from Vehicle Safety for all the discussions, encouragement and friendship.

Thanks to my family for giving me the opportunity to pursue my dreams. And finally, thanks to Nikola for standing always by my side.

The research work was supported by the MICA project (Ref. No. 2017-05522) sponsored by the Swedish government agency for innovation (VINNOVA), within FFI program (strategic vehicle research and innovation), Sweden.

Jordanka Kovaceva
Göteborg, December 2019
# Contents

Abstract v

Acknowledgments vii

## I Introductory chapters 1

1 Introduction 3

1.1 Safety benefit assessment .................................. 3
1.2 Driver models ............................................. 4
1.3 Data collection methods for modelling driver behaviour .... 5
1.4 Driver behaviour during overtaking .......................... 6
1.5 The need for inclusion of driver models for safety benefit assessment 7
1.6 Aim and Objectives ......................................... 8

2 Driver models for overtaking a cyclist 9

2.1 Methodologies for data collection ............................ 9
2.2 Studying driver behaviour during overtaking .................. 10
2.3 Computational models of driver behaviour in the approaching phase 10
2.4 Parameter estimation ....................................... 12
2.4.1 Linear programming ...................................... 14
2.4.2 Quadratic programming ................................... 14
2.4.3 An example of optimal parameter estimation .............. 14

3 Summary of papers 17

4 Discussion 23

4.1 Driver behaviour during overtaking .......................... 23
4.2 Models of driver behaviour .................................. 24
4.3 Limitations .................................................. 25
4.4 Future work ................................................ 25

5 Conclusions 27

Bibliography 29
II  Appended papers  39
Part I

Introductory chapters
Chapter 1

Introduction

The total number of road crashes in Europe is decreasing, but the number of crashes involving cyclists is not decreasing at the same rate (European Commission Directorate General for Mobility and Transport, 2018). Meanwhile the number of cyclists in traffic is increasing, making car-cyclist interactions an important focus for future traffic-safety improvements (OECD/ITF, 2013). According to the European Commission’s Annual Accident Report (European Commission Directorate General for Mobility and Transport, 2018), 8% of road fatalities in the European Union (EU) were cyclists. Recent research on car-cyclist crashes in Europe has found that most fatalities occur when the car is travelling straight and the cyclist is moving in line with traffic (Wisch et al., 2017). To help the driver avoid or mitigate these car-to-cyclist crashes, advanced driver assistance systems (ADAS) have been designed to warn drivers or intervene in critical situations (Lindman, Ödblo, et al., 2010; Zhao et al., 2019). Safety benefit assessments are performed to investigate the effectiveness of ADAS in reducing crashes and personal injuries (by, for example, comparing the number of crashes of the vehicles with and without ADAS, after the systems are available in production vehicles on the road).

1.1 Safety benefit assessment

Several approaches have been proposed for assessing the expected real-world safety benefit of ADAS (Carter and Burgett, 2009; Page et al., 2015; Sander, 2018). In general, a differentiation is made between retrospective and prospective safety benefit assessments. A retrospective assessment is based on observed real-world data after the systems are implemented in vehicles. Retrospective assessments have been performed, by analysing meta-data (Fildes et al., 2015), insurance claims data (Cicchino, 2017; Doyle et al., 2015; Isakså-Hellman and Lindman, 2016; Kuehn et al., 2009), national crash databases (Sternlund et al., 2017; Gärder and Davies, 2006), and naturalistic driving (ND) data (McLaughlin et al., 2008; Noort et al., 2012). This type of assessment aims to estimate the true effect of the systems, but it may require a long time until such systems are available in production vehicles (Eichberger, 2010).
In contrast, a prospective assessment is done before the systems are implemented in production vehicles (for example, by real-world testing of prototype systems using test tracks, driving simulator studies, or virtual assessment). Real-world testing on a test track, physical testing in a controlled environment, is often used to determine if the system works according to specifications (Edwards et al., 2015; Nilsson, 2014). This type of testing has the advantage of testing the actual system in a safe environment, which ensures high physical fidelity; however, the number of tests is usually limited due to the cost of performing the tests, and the interactions are typically with dummies and often driven remotely, without a driver in the vehicle. However, driving simulator studies, in which human drivers interact with the model of the system being evaluated, cost less (Aust et al., 2013). Further cost reduction can be obtained using virtual assessment; all components (driver, vehicle, environment) are modelled. In virtual assessment (Page et al., 2015), often taking form of counterfactual simulations, a re-analysis of real-world data (crashes or near-crashes) is performed (Bärgman et al., 2017a; Gorman et al., 2013; Rosen, 2013; Van Auken et al., 2011). The real-world data used as input in the virtual assessment provide the baseline scenarios. The baseline scenarios describe the scenario to be analysed without the ADAS under assessment. They are the basis for the simulation with the ADAS (Alvarez et al., 2017). The baseline scenarios can be derived from three types of data: original real-world scenarios (Kusano and Gabler, 2012; Lindman, Ödblom, et al., 2010; Sander and Lubbe, 2016), modified real-world scenarios (Bärgman et al., 2017a), and artificial scenarios generated by using, for example, distributions of crash-contributing factors from real-world scenarios (Dobberstain et al., 2017; Jeong and Oh, 2017; Yanagisawa et al., 2017). The virtual assessment approach allows a different number of simulations to be performed (depending on the availability of the baseline data). It can be applied in the early stages of ADAS development.

### 1.2 Driver models

In their early development stage, when ADAS are not yet available, models of the systems, vehicle, environment, and driver are needed for safety benefit assessments (Page et al., 2015). Driver models used in counterfactual simulations should be able to describe relevant aspects of driver decisions in the scenario that the ADAS is intended to address (Markkula, 2015). Driver models have been classified as conceptual, statistical, or process models (Markkula, 2015; Markkula et al., 2012). The conceptual models are not defined in rigorous mathematical formulations or implemented computationally. Instead they explain the driving process and how drivers interact with the world (some examples are zero-risk theory models (Näätänen and Summala, 1974), risk control models (Wilde, 1982; Janssen and Tenkink, 1988), and hierarchical models (Michon, 1985)). Statistical models explain the driver behaviour as distributions of, for example, reaction times (Green, 2000). Process models describe the driver reaction based on observed quantities (Macadam, 2003; McRuer, 1980; Nash et al., 2016). These models produce an output, such as an action (steering or braking), using recent and past measurements (Markkula, 2015; Boda, Dozza, et al., 2018). Statistical and process models (both also referred to as
computational models) are expressed in mathematical terms and are thus suitable for evaluating scenarios using counterfactual simulations. Depending on the scenario, existing computational models may not be able to capture relevant factors, and additional models may need to be developed (Benderius, 2012). It is also important to note that process models can include components of statistical models (Bärgman et al., 2017a; Markkula, 2015).

1.3 Data collection methods for modelling driver behaviour

To study and model driver behaviour and evaluate the models, we need to understand driver behaviour in the specific scenario which the ADAS is intended to address. Some studies can provide detailed data about how drivers behave, such as their choice of speed and the distance they maintain to surrounding objects and they are use of controls (steering wheel angle, gas and brake pedal).

Test-track (TT) experiment studies are suitable for studying driver behaviour in interactions with other road users and for designing driver models (Boda, Dozza, et al., 2018; Kiefer et al., 2003; Najm and D. L. Smith, 2004). These studies use real vehicles, in specific scenarios, in controlled settings. The participants drive on a TT while a researcher might sit in the passenger seat. The participants usually know that they are being tested, but the researchers may or may not reveal the real purpose of the study. The experiments are typically repeated, and different factors that may influence the driver behaviour in the specific scenario can be controlled. The data required to design driver models are collected by a Data Acquisition System (DAS) installed in the test vehicle(s), which records the driver’s controls, such as steering wheel angle and gas and brake pedal (Boda, Dozza, et al., 2018; Kiefer et al., 2003; Najm and D. L. Smith, 2004). Additional detailed data, such as the positions and speeds of the test vehicle as well as other road users in the scenario, are also typically recorded. However, in test-track experiments the other road users are typically inflatable cars or dummies and there are few drivers (e.g. typically 10-50) in a controlled environment. On the whole these experiments have limited ecological validity (Green, 2000; Hoffman et al., 2002), which refers to the degree to which the observed driver behaviour in the experiment reflects "real-world" behaviour patterns (what drivers typically do) (Shinar, 2017), (p. 50).

With the advent of big data, recent studies (Dingus et al., 2006; SHRP2 TRB, 2015) have emerged that are not performed in a controlled environment but rely, instead, on naturalistic driving (Shinar, 2017). In contrast to TT experiments, ND studies collect large amounts of continuous data on normal driving from many drivers, providing detailed information on how drivers behave in the real world—without the influence of instructions, predefined routes, and preselected environments (Shinar, 2017). ND data are suitable as input not only for designing driver models but also for developing and assessing ADAS (Bärgman et al., 2017a).

Other studies that investigate and record information about road crashes after they have happened (post-hoc) are less suitable for modelling driver behaviour.
These studies collect information with different levels of detail about crash causation mechanisms in macroscopic- and microscopic-level databases (Lindman, Isaksson-Hellman, et al., 2017); they do not collect time-series data of the response process. Actually, the macroscopic-level databases, which typically only include police-reported road crashes, only contain information about the crash time and location, and the person(s) involved (such as hospital care and specific injuries sustained). They do not contain information about the response process which could be used to model behaviour. On the other hand, microscopic-level, in-depth databases with crash reconstructions provide information in a time-series format on pre-crash trajectories. That is, in-depth databases often include data from a few seconds before the crash, including vehicle speed, distance between vehicles and driver deceleration and reaction time. Due to their level of detailed, in-depth databases have been used primarily to provide input to counterfactual simulations (Lindman, Ödblom, et al., 2010; Sander, 2018). However, as the time-series data in these databases are almost always created via reconstructions, they are already based on assumptions. This fact makes them much less suited for use in modelling than, for example, ND data—even though ND data usually only include near-crashes and normal driving behaviour, with very few actual crashes.

### 1.4 Driver behaviour during overtaking

One scenario in which cyclists are exposed to dangerous conflicts is overtaking scenario on rural roads; car drivers share the same lane as the cyclists and the difference between speeds of the car and cyclists is large. The research on car-cyclist interactions while overtaking started long ago (Kroll and Ramey, 1977) and continues to the present day. During these interactions, drivers try to minimize their risk by choosing to stay far enough away from potential hazards to feel safe and comfortable—that is, they strive to remain within their comfort zone (Summala, 2007). Drivers’ comfort zone boundaries (CZBs) while passing a cyclist have been summarized by lateral clearance, which is typically defined as the minimum lateral distance between the cyclist and the vehicle while the vehicle is passing the cyclist (Llorca et al., 2017). CZBs have implications for timely activations of ADAS, because too early activation of automated safety system may cause annoyance and too late activation may cause crashes (Lubbe and Davidsson, 2015).

Factors related to the infrastructure that influence lateral clearance include road grade, posted speed, and presence and width of a shoulder, as well as the presence of a cycling lane (Chapman and Noyce, 2012; Feng et al., 2018). Walker et al. (2014) and Chuang et al. (2013) have shown that bicyclists’ visible characteristics, such as gender, helmet-wearing, and clothing, also influence the lateral clearance. In addition, cyclist speed and speed variation have been shown to affect the lateral clearance (Chuang et al., 2013). Another factor that influences the lateral clearance is how the overtaking manoeuvre is performed: drivers may keep their vehicle speed relatively constant (flying strategy) or they may decelerate and follow the cyclist before passing (accelerative strategy) (Matson and Forbes, 1938).
Research has also shown that when oncoming traffic is present the lateral clearance is smaller (Goodridge, 2017; McHenry and Wallace, 1985). In fact, the presence of oncoming traffic has been identified as the principal factor affecting lateral clearance (Piccinini et al., 2018; Dozza et al., 2016). The authors (Piccinini et al., 2018) found a significant correlation between the overtaking strategy and the nominal time-to-collision (TTC) (between the overtaking and oncoming vehicle): as the TTC decreased, more drivers used the accelerative strategy, because they had slowed down and waited for the oncoming vehicle to pass before accelerating to overtake the cyclist. The study also found that the minimum lateral safety margins were larger in the accelerative than the flying strategy. Evans et al. (2018) show that the presence of a vehicle in the adjacent lane, travelling in the same direction as the vehicle overtaking the cyclist, has the largest effect on reducing the lateral clearance between a cyclist and an overtaking vehicle on urban and suburban roads.

By dividing the overtaking manoeuvre into four phases (approaching, steering away, passing, and returning), Dozza et al. (2016) were able to define new CZBs for the three new phases, analogous to the lateral clearance CZB for the passing phase. Since then, factors that influence the driver’s CZBs in all four phases have been studied in different experimental environments. For example, overtaking behaviour in simulator studies has been investigated by Piccinini et al. (2018). Other approaches include using an instrumented car (Schindler and Bast, 2015) or bicycle (Dozza et al., 2016). Semi-naturalistic studies of car-cyclist interactions have been performed by Parkin and Meyers (2010), Chuang et al. (2013), Dozza et al. (2016), Walker (2007), Walker et al. (2014), and Evans et al. (2018). These studies used bicycles equipped with data loggers and sensors (e.g., ultrasonic sensors, lasers, adn LIDAR or a GoPro video camera) to collect field data from the cyclist’s perspective. Although these studies were conducted in naturalistic settings, the bicyclists were instructed to ride on a specific road. Consequently these are not fully naturalistic studies.

1.5 The need for inclusion of driver models for safety benefit assessment

Cycling has been on the increase as a mobility choice over the past several decades in Europe (ECF, 2016). As long as cyclists and drivers share the road, or parts of the road, it is important that they adopt a safe strategy to interact with each other (Shinar, 2017). In these interactions, it is argued that the driver has the critical role, either changing the vehicle path or failing to adjust the vehicle position to accommodate the cyclist (OECD/ITF, 2013). To keep cyclists and drivers safe, ADAS that address car-to-cyclist scenarios are being developed; as a result guidelines for their prospective safety benefit assessment through counterfactual simulations are being proposed (Alvarez et al., 2017; Fahrenkorg et al., 2019; Wimmer et al., 2019). However, the knowledge about when and how ADAS should intervene in overtaking scenarios with cyclists is still limited. Recent research highlights the fact that the inclusion of driver models in ADAS may help produce timely and acceptable interventions, which in turn may increase their effectiveness (Dozza et al., 2016;
Moreover, recent studies have shown that the choice of driver model has a large effect on the estimated benefit of ADAS (Bärgman et al., 2017a). Factors that may be important in modelling driver behaviour have already been indicated, such as the presence of oncoming traffic (Dozza et al., 2016; Farah et al., 2019; Piccinini et al., 2018), but have not yet been confirmed in naturalistic driving studies. Because driver behaviour while overtaking cyclists has not yet been modelled, it is not yet included in the current ADAS or in the counterfactual simulations which assess them.

1.6 Aim and Objectives

The overall aim of this PhD work is to develop methodologies for the safety benefit assessment of ADAS that help avoid collisions with cyclists in overtaking scenarios. Specifically, these new methodologies will integrate behavioural models for interactions among road users. The objectives to achieve this aim are:

1. to extract and analyse overtaking manoeuvres (drivers overtaking cyclists) in naturalistic driving data;

2. to use the results from the first objective, to develop new and compare existing models of driver behaviour for overtaking manoeuvres that can be used in counterfactual simulations of safety benefit from naturalistic data;

3. to design new counterfactual simulation tools that include the driver behaviour model(s) from the second objective to assess ADAS in cyclist-overtaking scenarios; and

4. to apply these new tools to naturalistic driving data, in order to prospectively estimate the safety benefit of new ADAS that help avoid collisions during cyclist-overtaking scenarios.

The first two objectives are addressed in this licentiate thesis, while the last two will be completed later on to achieve the PhD degree.
Chapter 2

Driver models for overtaking a cyclist

The objectives of this thesis work are to extract and analyse cyclist-overtaking manoeuvres from naturalistic driving data and compare driver behaviour models for overtaking manoeuvres that can be used in counterfactual simulations for evaluating ADAS safety benefits. Metrics that may be important for modelling driver behaviour while overtaking a cyclist have already been identified through different data collection methods. This section gives an overview of the data collection methods for modelling driver behaviour, used in Papers 1 and 2 (Section 2.1) and introduces the work covered in Paper 1, wherein the metrics are identified by analysing naturalistic driving data (Section 2.2). Driver models that can be used in counterfactual simulations are described in Section 2.3. In Section 2.4, an expansion beyond what is covered in Paper 2 is provided; the aim is to elaborate on how convex programming can be used to find the parameters for driver models.

2.1 Methodologies for data collection

Data used in Papers 1 and 2 is provided by two types of studies, ND and TT experiments. These two methods are often complementary—for example, when studying overtaking events where a driver overtakes a single bicyclist on a straight rural roads. The UDrive dataset, the largest naturalistic driving study in Europe (Bärgman et al., 2017b), was used in this study to extract overtaking events. The UDrive dataset includes data from 120 instrumented passenger cars, with almost 1.8 million kilometres travelled distance in one year of driving in six European countries. A Data Acquisition System (DAS) installed in the vehicles, registered: seven camera views (front left, front centre, front right, cabin view, cockpit view, driver face, and pedals), CAN bus data (vehicle speed, acceleration, steering wheel angle, and yaw rate), and GPS position. Unlike data from most previous naturalistic studies (Dingus et al., 2006; SHRP2 TRB, 2015), UDrive recorded continuous signals from a Mobileye smart camera system (Shashua et al., 2004). Thus the following information was also collected: the presence and type of other road users (cyclists, vehicles, and pedestrians) and their distances (lateral and longitudinal) from the
2.2 Studying driver behaviour during overtaking

Studying data about driver’s overtaking behaviour from two different datasets, described in Section 2.1, requires well-established definitions to make the comparative analysis easier. This work is based on the four phases of overtaking proposed in Dozza et al. (2016). Phase 1, approaching, starts when the cyclist is detected by the sensors; Phase 2, steering away, starts when the vehicle begins to divert from a collision course; Phase 3, passing, starts when the front of the vehicle is less than three metres behind the rear of the cyclist; and Phase 4, returning, starts when the rear of the vehicle is more than three metres in front of the cyclist (see Figure 2.1). In this work, we study the first three phases. The following driver CZBs were quantified: minimum approaching gap (MAG) for Phase 1; minimum distance (MDS) for Phase 2; lateral clearance (LC) for Phase 3; time-to-collision (TTC) between the vehicle and the nearest oncoming traffic (when present) at the boundary between Phases 1 and 2, 2 and 3, and 3 and 4; and TTC to the bicycle at MAG (TTC\(_b\)), just before the vehicle initiates Phase 2. The effects of the factors car speed, manoeuvre type, presence of oncoming traffic, and driver characteristics (age, gender, Arnett Inventory of Sensation Seeking score) on MAG, MDS, and LC were investigated using linear mixed-effects models, which incorporate both fixed and random effects (Bates et al., 2015).

2.3 Computational models of driver behaviour in the approaching phase

Poor timing or the driver’s failure to see the cyclist, at the end of the approaching phase (Phase 1, Figure 2.1) can both result in a rear-end collision with the cyclist. To help the driver avoid a rear-end collision with the cyclist, the ADAS should incorporate knowledge about typical driver behaviour. The approaching phase of the
Chapter 2. Driver models for overtaking a cyclist

Figure 2.1: Four overtaking phases while driver is overtaking the cyclists.

flying overtaking strategy is further studied in Paper 2. That is, Paper 2 compares driver models that compute an output value that indicates when a car driver starts steering away to overtake a cyclist on rural roads, using overtaking data from a TT study and UDrive. The comparison of driver models is warranted, since the choice of the driver model has been found to have a large effect on the estimated safety benefit of ADAS using virtual simulations (Bårgman, Boda, et al., 2017). In Paper 2 three computational models are compared: a threshold model, an accumulator model, and a model inspired by a proportional-integral-derivative controller.

A threshold model assumes a response threshold at which drivers start responding to the threat (Kiefer et al., 2003; D. N. Lee, 1976). The response threshold model has been investigated with different visual cues in a relatively large set of studies. For example, D. N. Lee (1976) suggested that drivers’ braking in rear-end scenarios is guided by the optical parameter $\tau$ and its derivative $\dot{\tau}$. Parameter $\tau$ is the ratio of $\theta$ to $\dot{\theta}$, where $\theta$ is the angular projection of an object on the subject’s retina and $\dot{\theta}$ is the angular expansion rate (the first derivative of $\theta$). In D. N. Lee (1976), drivers are assumed to start their braking when $\tau$ reaches a specific threshold value. M. Smith et al. (2001) proposed using $\theta$ and $\dot{\theta}$ as the criteria for driver braking in rear-end collision avoidance, while Kiefer et al. (2003) used an inverse $\tau$ threshold.

An alternative to the threshold model is the accumulator type model, which has been studied in domains such as psychology and neuroscience (Gold and Shadlen, 2001; Gold and Shadlen, 2007; Ratcliff, 1978; Ratcliff and P. L. Smith, 2004), as well as in driver modelling; see Markkula (2014) and Ratcliff and Strayer (2014). According to this model type, the driver’s action occurs after accumulation (or practically, integration), of sensory evidence.

Another type of model is the proportional-integral-derivative (PID) model,

$$y(t) = K_P \hat{z}(t) + K_I \int \hat{z}(\tau) d\tau + K_D \frac{d\hat{z}(t)}{dt}$$  (2.1)

where $y(t)$ is the control function, $\hat{z}(t)$ is the measured signal and $K_P$, $K_I$, and $K_D$ are the parameters of the proportional, integral and derivative terms, respectively. The PID controller model has been extensively used in many domains (Bennett, 1993; O’Dwyer, 2009; Rivera et al., 1986). This type of model has, for example, been applied also in control models for driver steering behaviour (Donges, 1978; Donges, 1999; Winner et al., 2016). The proportional term is proportional to the measured signal. The integral term takes the past values of the measured signal and
integrates them over time. The derivative term is an estimate of the future trend of the signal, based on its current rate of change. Using only the proportional term in (2.1) results in a threshold model, while using only the integral term in (2.1) results in an accumulator model equivalent.

Given a driver model (for example the PID model discussed above), the next step is to estimate its parameters so that the model best fits the collected data.

2.4 Parameter estimation

In the literature, different approaches have been used to estimate parameters. This section will introduce the background of convex programming, related to the contents of Paper 2, and give an example of its application to tuning the parameters of a well-known driver model.

Parameters have been tuned by hand by, for example, Gordon and Magnuski (2006) and Salvucci (2006), while heuristic optimization methods have been used by Benderius (2012), for example. Heuristic optimization is general and can be applied without simplifying the studied systems, but it does not guarantee a globally optimal solution. In contrast, methods that guarantee global optimality often require intractable computation time that increases exponentially with problem size. Convex programming methods are an exception, able to solve a convex problem in polynomial computation time, while simultaneously providing a proof or certificate that the solution is indeed a global optimum (Boyd and Vandenberghe, 2004) (p.242). Furthermore, publicly available solvers are available. However, the downside of this approach is that many optimization problems cannot be cast as convex programs (Boyd and Vandenberghe, 2004). Convex programming is useful for solving a “relaxed problem” (see Figure 2.3): a problem that provides a lower bound to the original non-convex problem, or by solving a subproblem that is locally convex, see Figure 2.2 (Boyd and Vandenberghe, 2004). In this thesis we pursue the latter approach: a generally non-convex problem of parameter estimation is solved by a combination of convex programming and a grid search. Parameters that appear in a non-convex form are gridded within a given range, and for each grid value a convex subproblem is solved to obtain the remaining parameters. Although the parameters optimised by convex programming are locally optimal, the parameters obtained by grid search are generally suboptimal, since the fixed grid resolution typically applied is confined to specific discrete values of the parameters. Thus solution quality depends on the grid resolution, which is a trade-off between optimality and computational efficiency.

The convex subproblems encountered in this thesis are of particular forms, a quadratic program (QP) and a linear program (LP), which may be seen as the simplest forms of convex programming. The two froms will be described briefly in the following section.
Figure 2.2: Illustration of a nonlinear and non-convex function to be minimized. Within the shaded regions the function is locally convex. The minimum of the function in the red region is a local optimum, while the minimum in the blue region is the global solution.

Figure 2.3: Illustration of an optimisation problem with a non-convex feasible set, illustrated by the blue region. In order to make the problem convex, the set is relaxed by enlarging it with the red region. The global optimum of the relaxed problem provides a lower bound (it is always below or equal) to the global optimum of the original non-convex problem.
2.4.1 Linear programming

Any optimisation problem that can be stated in the form

$$\min_x \quad c^T x$$ \hspace{1cm} (2.2a)
subject to: $Ax \leq b, \quad x \in \mathbb{R}^n$ \hspace{1cm} (2.2b)

is called a linear program (LP). Here $x$ is a vector of $n$ decision variables (i.e., quantities controlled by the decision maker). The matrix $A \in \mathbb{R}^{m \times n}$ and the vectors $c \in \mathbb{R}^n$ and $b \in \mathbb{R}^m$ are given coefficients, where $\mathbb{R}$ denotes the set of real values. The scalar function $c^T x$ is called an objective function or a performance index. It provides a value system for ranking the possible solutions, in order to identify the optimal solution $x^*$ that minimises the objective function (2.2a). Equation $Ax \leq b$ in (2.2b) enforces $m$ constraints in the problem. The constraints represent physical or other restrictions on the numerical values that can be assigned to the vector of decision variables $x$.

2.4.2 Quadratic programming

Any optimisation problem that can be stated in the form

$$\min_x \quad \frac{1}{2} x^T H x + c^T x$$ \hspace{1cm} (2.3a)
subject to: $Ax \leq b, \quad x \in \mathbb{R}^n$ \hspace{1cm} (2.3b)

is called a quadratic program (QP). The decision variables and the coefficients $A$, $b$ and $c$ are defined as in the LP formulation (2.2). However, the QP formulation (2.3) allows an additional quadratic term in the objective, where $H \in \mathbb{R}^{n \times n}_{\geq 0}$ is a positive semidefinite matrix. It is clear that a QP is a more general form than an LP, since by setting $H = 0$ the QP transforms directly to an LP.

2.4.3 An example of optimal parameter estimation

As an optimisation example of parameter estimation, we revisit the well-known driver model for steering, proposed by Salvucci and Gray (2004). The model adjust the steering angle $y_j(x)$ as a function of measurements

$$\tilde{z}_j(\tau) = \begin{bmatrix} \tilde{\theta}_{nj}(\tau) & \tilde{\theta}_{ij}(\tau) \end{bmatrix}^T, \quad \tau \in [t_0, t], \quad j = 1, \ldots, N_d$$ \hspace{1cm} (2.4)

that include the visual direction angles to near and far points ahead, denoted by $\tilde{\theta}_{nj}$ and $\tilde{\theta}_{ij}$, respectively. The symbol $\tilde{\cdot}$ is here used to denote measured data from $N_d$ drivers, over the time interval from $t_0$ to $t$. Salvucci and Gray (2004) propose a proportional-integral driver model for steering that consists of a proportional gain to both the near and far points ahead, and an integral gain to the near point,

$$y_j(x) = K_{Pn} \tilde{\theta}_{nj}(t) + K_{Pf} \tilde{\theta}_{ij}(t) + K_{In} \int_{t_0}^t \tilde{\theta}_{nj}(\tau) \, d\tau$$ \hspace{1cm} (2.5)
where
\[ x = \begin{bmatrix} K_{Pn} & K_{Pf} & K_{In} \end{bmatrix}^T \] (2.6)
are unknown parameters. The goal is to estimate the best values of the parameters such that the error
\[ \| y_j(x) - \tilde{y}_j(t) \| \] (2.7)
between the steering angles \( y_j \) obtained by the model (2.5) and the measured angles \( \tilde{y}_j \) for all the drivers \( j = 1, \ldots, N_d \) is minimised. The function \( \| \cdot \| \) may, in principle, denote any norm, although in practice norms 1 and 2 are most commonly used.

**Parameter estimation with linear programming**

Consider norm 1 (or, identically, the mean absolute error)
\[ \min_x \frac{1}{N_d} \sum_{j=1}^{N_d} |y_j(x) - \tilde{y}_j(t)|. \] (2.8)
At first glance, problem (2.8) may appear nonlinear due to the absolute value function. However, the problem can be formulated as a linear program,
\[ \min_{x, e_j} \frac{1}{N_d} \sum_{j=1}^{N_d} e_j \] (2.9a)
subject to:
\[ e_j \geq y_j(x) - \tilde{y}_j(t), \quad j = 1, \ldots, N_d \] (2.9b)
\[ e_j \geq -(y_j(x) - \tilde{y}_j(t)), \quad j = 1, \ldots, N_d \] (2.9c)
\[ [x^T, e_1, \ldots, e_{N_d}]^T \in \mathbb{R}^{3+N_d} \] (2.9d)
with the help of new variables \( e_j \) and two inequality constraints per driver that represent the absolute error in a linear form. Let
\[ \tilde{\omega}_{nj}(t) = \int_t^T \tilde{\theta}_{nj}(\tau) d\tau \] (2.10)
represent the integral, for simplicity, and denote the augmented vector of decision variables as
\[ \tilde{x} = \begin{bmatrix} K_{Pn} & K_{Pf} & K_{In} & e_1 & \cdots & e_{N_d} \end{bmatrix}^T \] (2.11)
By defining coefficients
\[ A = \begin{bmatrix} \tilde{\theta}_{n1}(t) & \tilde{\theta}_{n1}(t) & \tilde{\omega}_{n1}(t) & -1 & 0 & \cdots \\ -\tilde{\theta}_{n1}(t) & -\tilde{\theta}_{n1}(t) & -\tilde{\omega}_{n1}(t) & -1 & 0 & \cdots \\ \tilde{\theta}_{n2}(t) & \tilde{\theta}_{n2}(t) & \tilde{\omega}_{n2}(t) & 0 & -1 & 0 & \cdots \\ -\tilde{\theta}_{n2}(t) & -\tilde{\theta}_{n2}(t) & -\tilde{\omega}_{n2}(t) & 0 & -1 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \tilde{\theta}_{nN_d}(t) & \tilde{\theta}_{nN_d}(t) & \tilde{\omega}_{nN_d}(t) & 0 & 0 & \cdots & 0 & -1 \\ -\tilde{\theta}_{nN_d}(t) & -\tilde{\theta}_{nN_d}(t) & -\tilde{\omega}_{nN_d}(t) & 0 & 0 & \cdots & 0 & -1 \end{bmatrix} \] (2.12)
\[ b = \begin{bmatrix} \tilde{y}_1(t) & -\tilde{y}_1(t) & \tilde{y}_2(t) & -\tilde{y}_2(t) & \cdots & \tilde{y}_{N_d}(t) & -\tilde{y}_{N_d}(t) \end{bmatrix}^T \] (2.13)
\[ c = \begin{bmatrix} 0 & 0 & 1/N_d & \cdots & 1/N_d \end{bmatrix}^T \] (2.14)
2.4. Parameter estimation

Problem (2.9) can be written in the standard LP form

\[
\begin{align*}
\min_{\tilde{x}} & \quad c^T \tilde{x} \\
\text{subject to:} & \quad A\tilde{x} \leq b, \quad \tilde{x} \in \mathbb{R}^{3+N_d}.
\end{align*}
\]  

(2.15a) (2.15b)

Then, the optimal values for the parameters are the first three values in \(\tilde{x}^*\), where \(\tilde{x}^*\) is the optimal solution of problem (2.15).

This example showed how convex programming, and in particular linear programming can be used to estimate parameters of driver model.
Chapter 3

Summary of papers

Paper 1

Paper 2
Kovaceva, J., Bärgman, J., Dozza, M., Enabling counterfactual analyses to estimate the safety benefit of advanced driving assistance systems: A comparison of driver models using naturalistic and test-track data from cyclist-overtaking manoeuvres, to be submitted to an international scientific journal.
Paper 1: Drivers overtaking cyclists in the real-world: Evidence from a naturalistic driving study

Introduction

The total number of road crashes in Europe is decreasing, but the number of crashes involving cyclists is not decreasing at the same rate. When cars and bicycles share the same lane, cars typically need to overtake them, creating dangerous conflicts—especially on rural roads, where cars travel much faster than bicycles. During these manoeuvres, drivers try to minimize risk in the complex traffic environment by staying in their comfort zone while overtaking the cyclist.

Aim

The paper quantified drivers' comfort zone boundaries (CZBs) and investigated the combination of factors that affect the CZBs while drivers overtake cyclists in a naturalistic setting.

Method

This study developed a four-step procedure to identify and extract overtaking manoeuvres from naturalistic driving data from UDrive. The effects of the factors car speed, manoeuvre type, presence of oncoming traffic, and driver characteristics (age, gender, Arnett Inventory of Sensation Seeking score) on CZBs were investigated using linear mixed-effects models.

Results

The results show that the higher the car speed the larger the CZBs while approaching and passing, but the presence of an oncoming vehicle significantly decreased the CZB during passing. The drivers' age, gender, and Arnett Inventory of Sensation Seeking score were not found to have a statistically significant impact on the CZBs.

Discussion

The presence of an oncoming vehicle is a crucial factor for the safety and comfort of the cyclist and needs to be taken into account for the development of advanced driver-assistance systems that maintain safe clearance to the cyclist. The results help identify which of the CZBs during an overtaking manoeuvre may be related to the risk of an accident in different scenarios. For example, the TTC to the oncoming vehicle at the end of the passing phase might help identify the risk of a head-on collision with an oncoming vehicle, while TTC to the bicycle in the approaching phase might estimate the risk for a rear-end collision with the bicycle. The findings of this study provide implications both for the design of road safety intervention programs that increase safety for all road users and for the development of advanced driver-assistance systems that could interact with cyclists.
Paper 2: Enabling counterfactual analyses to estimate the safety benefit of advanced driving assistance systems: A comparison of driver models using naturalistic and test-track data from cyclist-overtaking manoeuvres

Introduction
Advanced driver assistance systems (ADAS) for car-cyclist interactions, as well as their safety assessments, can be improved by understanding and modelling scenario-dependent driving behaviours. Previous studies have focused on describing driver behaviour in these interactions, but different models have not been compared.

Aim
This study compares driver models that compute an output value that indicates when a car driver starts steering away to overtake a cyclist on rural roads.

Method
Three models are compared: threshold model, accumulator model and a model inspired by a proportional-integral-derivative controller. These models are tested and verified using two different datasets, one from a naturalistic driving (ND) study and one from a test-track (TT) experiment. Two perceptual variables, expansion rate and inverse tau, are tested as input to the models. A linear cost function is proposed to obtain the optimal parameters of the models by computationally efficient linear programming.

Results
The results show that the models based on inverse tau fitted the data better than the models that include expansion rate. All three models give high variability in detecting steering away time for different drivers. Furthermore, differences were discovered in detected steering away time between models fitted to test-track experiment and naturalistic driving data, and further analysis is required, perhaps using better models.

Discussion
These tested models have implications for the development of counterfactual simulations, which can estimate the safety benefits of ADAS, such as forward collision warning and autonomous emergency braking, by simulating realistic driver behaviour. The linear cost function and the linear programming approach used in this paper have the potential to be used for parameter optimization of models such as those used in ADAS.
Chapter 4

Discussion

4.1 Driver behaviour during overtaking

Paper 1 used ND data to investigate factors that affect drivers’ CZBs while they overtake cyclists. The CZBs discussed below include the minimum distance between the car and the cyclist in the approaching and steering away phases, the lateral clearance during the passing phase (LC), the TTC to the cyclist at steering away ($TTC_b$) and TTC between the car and the oncoming vehicle (when present) at each of the phase boundaries (see Figure 2.1).

The results of Paper 1 support the findings from the previous studies by Dozza et al. (2016) and Piccinini et al. (2018), and extend the knowledge about driver behaviour in car-to-cyclist overtaking scenario in naturalistic driving beyond what was reported in literature. The findings of Paper 1 show that as car speed increased, the LC also increased, although not much ($0.01 \text{ m per } 1 \text{ km/h}$). In contrast, neither Dozza et al. (2016) nor Mehta (2015) found a significant influence of car speed on LC. However, our findings are in line with cyclists’ expectations that higher speeds require a larger LC (Llorca et al., 2017). The presence of oncoming vehicles significantly decreased the LC during the passing phase.

No significant differences were found between the accelerative and flying manoeuvres with respect to the TTCs to the oncoming vehicles at each of the phase boundaries. For all manoeuvres, the average TTC at the boundary between Phases 3 and 4 was only 1.8s. This short TTC may not be sufficient to trigger a driver reaction to a warning in critical situations when the driver doesn’t see the cyclist. In fact, the literature suggests that perception reaction time is at least one second (Van Der Horst and Hogema, 1993). This finding may motivate the development of ADAS that prevent a head-on collision (Clarke et al., 1999) if the driver decides to perform an overtaking manoeuvre that is unsafe because the TTC to the oncoming vehicle is too short at the time the driver decides to overtake the cyclist.

Furthermore, $TTC_b$ was not found to be significantly different between accelerative and flying manoeuvres. We found $TTC_b$ to be 2.1s, so a warning at 3s TTC, which is suggested for warning systems intended to avoid rear-end crashes (S. Lee et al., 2004), may not be acceptable to drivers. However, delaying this warning may reduce its effectiveness, thus requiring alternative intervention strategies. Therefore, a
forward collision warning alone may not be viable for avoiding rear-end collisions with cyclists, and autonomous emergency braking may be the only alternative solution (Boda, Dozza, et al., 2018).

Paper 2 used the analysis from Paper 1 to further study the flying overtaking strategy and to incorporate the findings into a computational driver model. Section 1.4 reported that few models for cyclist-overtaking behaviour exist in the literature. Therefore, Paper 2 compared existing models to understand which models may be suitable to describe the approaching phase. The results show that the models based on inverse $\tau$ fitted the data better than the models that included optical expansion rate, $\dot{\theta}$. However, the models only describe the driver’s action in the approaching phase of the flying overtaking strategy; future work should extend the models to cover all overtaking strategies and phases.

Paper 1 showed that the drivers’ age, gender, and score from the Arnett Inventory of Sensation Seeking score, which all drivers completed, did not statistically significantly influence the CZBs. However, previous research (Farah, 2011; Farah, 2013) indicates that there are behavioural differences in the overtaking manoeuvres of older compared to younger drivers. It may be that the data sample in Paper 1 was relatively homogeneous in terms of such characteristics, and/or the sample size was simply too small to show statistical significance. As a result, in Paper 2, the models are tuned on the data collected from all drivers and provide a single set of model parameters for an average driver. A possible further improvement would be to tune the model on each individual driver or on different groups of drivers (e.g., young and old, aggressive and calm) when more data are available.

### 4.2 Models of driver behaviour

Paper 2 demonstrated how the driver behaviour from a model tuned on TT data can be verified with ND data, and vice versa, to perform model validation. Previous research (Benderius et al., 2011; Markkula et al., 2012) noted that efforts to compare and tune driver models have been rather limited. Paper 2 also demonstrates how linear programming can be used for tuning model parameters in a computationally efficient way, which ensures that a global minimum solution for the chosen error and cost function is obtained. However, the process of model validation is not trivial. Further analysis is needed before including the model in counterfactual simulations.

The parameter fitting approach in Paper 2 is particularly suitable for large optimization problems, constructed by a large dataset of drivers or large set for parameter search, and is promising for optimizing model parameter, e.g., those connected to ADAS, intended for counterfactual simulations. However, not all problems can be defined by a convex or a linear cost function, and other methods may need to be used in such cases. Additionally, it may not be possible to transform the problem as a linear program; in particular, it might not be possible to define a meaningful error which is the convex function of the parameters.
4.3 Limitations

Data from UDrive are unique; they are unlike those from previous large naturalistic studies (Dingus et al., 2006; Hankey et al., 2016), because Mobileye detects interactions with vulnerable road users, as long as they occur in the daytime. As a result, many automatically identified car-to-cyclist overtaking manoeuvres can be analysed with high ecological validity. However, there are some limitations to the UDrive data. As noted, the smart-camera could only detect cyclists in daylight, which prevented us from investigating CZBs at night, and we can only analyse those cyclists detected by the camera; we do not know whether there were any false positives (cyclists that were not detected) in daylight, either. Another limitation is that both the UDrive and TT datasets had limited generalizability. The drivers were driving only one car brand and there was limited representation of young and old drivers (Bärgman et al., 2017a) in the ND data (Paper 1). In the TT data, all drivers were driving the same car, and the experiment with robots was precisely repeated for each of the drivers (Paper 2).

The parameter optimization of the models in Paper 2 was carried out using only data from trials in which drivers steered away. In the future, more situations in which drivers have not steered should be added to the optimization. Furthermore, the models were tested with only one cost function. In the future, other cost functions, parameter constraints, or parameter optimization methods could be compared by defining different functions (e.g., polynomial, exponential, trigonometric) and different numbers of parameters (e.g., 2nd or 3rd-order polynomial), and providing more measurements (e.g., different signals).

In this work, only three types of models were tested, as they have been widely used in rear-end scenarios. There are other quantitative models that could be compared which capture more aspects of the steering process (Hildreth et al., 2000; Salvucci, 2006). Furthermore, only two input measurements to the models were used, $\dot{\theta}$ or inverse $\tau$, supported by the theory of D. N. Lee (1976). In future work, a combination of these and other types of inputs may be considered e.g. those proposed by Boda, Lehtonen, et al. (2019). Additional input to the models, which takes into account the road profile, could be the visual direction angle to a near and far point ahead (Salvucci and Gray, 2004).

4.4 Future work

In this work, we utilized data from ND and TT datasets, which could, in the future, be harmonized with crash databases in order to contribute to a further understanding of injury mechanisms and conditions for cyclist’ crashes, improve and validate driver models, and assess ADAS using counterfactual simulations. In particular, the data from crash databases will provide situations to estimate benefits in terms of crash avoidance and speed reduction and to evaluate true positives (system intervenes when intervention is needed) and false negatives (missed interventions). The normal driving events from ND data, non-crashes, can be used for evaluating false positives (unnecessary interventions) and true negatives (system does not intervene when
intervention is not needed). Care should, however, be taken when using reconstructed crashes as input to the driver models.

Furthermore, the in-depth crash data can provide information in cases where the driver missed to steer avoiding the cyclist. Driver models that address the steering-away, passing, and returning phases of an overtaking manoeuvre, in addition to the approaching phase, could be developed to increase the operational design domain of the ADAS and could be implemented in the counterfactual simulations.

The model comparison from Paper 2 could be extended with other optimization methods to gain deeper knowledge about the performance of existing models, as well as to understand the differences between parameter fitting methods. A larger set of models could be assessed (with multiple parameter fitting techniques as well), to better determine which models would be best suited for implementation in counterfactual simulations.

In the future, safety benefit estimation will be developed into software to assure rapid assessment of the effect of different ADAS algorithms and their comparison. The assessment of the safety benefit will be on ADAS that can avoid crashes during all phases of the overtaking manoeuvre, head-on collisions with oncoming traffic, and side swipes of the cyclist in the passing and returning phases. Finally, in the future, a method to combine the effects of active and passive safety systems could be investigated to assess the benefit of an integrated system.
Chapter 5

Conclusions

UDrive naturalistic driving data have given us valuable information about how car drivers overtake cyclists on rural roads. In Paper 1, we quantified drivers’ Comfort Zone Boundaries (CZBs) and investigated the combination of factors that affect the CZBs while drivers overtake cyclists in a naturalistic setting. The results of Paper 1 helped identify which of the CZBs during an overtaking manoeuvre may be related to the risk for an accident in different scenarios. For example, the TTC to the oncoming vehicle at the end of the passing phase might help identifying the risk of a head-on collision with an oncoming vehicle, while the TTC to the bicycle in the approaching phase might help estimate the risk for a rear-end collision with the bicycle. Thus, the findings of Paper 1 have implications for the development of ADAS that can measure the TTC to an oncoming vehicle or a bicyclist to be overtaken, providing timely and acceptable warnings—or interventions—when drivers exceed their usual CZB. In addition, drivers’ CZBs are influenced by car speed: the higher the speed the larger the CZBs (in the approaching and passing phases) maintained by overtaking cars. This result shows that drivers may have perceptions similar to those of cyclists, who expect a larger clearance when being overtaken faster. However, the extent to which drivers increase CZB sufficiently to preserve cyclist safety and comfort as they drive faster is still to be investigated. It was also found that drivers were significantly closer to the cyclist when an oncoming vehicle was present, confirming that the presence of an oncoming vehicle is a crucial factor for the safety of the cyclist. This result needs to be taken into account when developing ADAS to ensure that drivers maintain a safe clearance to the cyclist.

Paper 2 used the analysis from Paper 1 to further study the approaching phase of the flying overtaking strategy and incorporate the findings into a computational driver model. In Paper 2, the models based on the inverse $\tau$ fitted the data better than the models that included $\theta$ according to the values of the cost function. This may indicate that the drivers are responding to a higher-order variable (higher-order optical primitive e.g., inverse $\tau$) and are indifferent to changes in the underlying variables (lower-order primitives, e.g., $\theta$ and $\dot{\theta}$) that leave the higher-order variable the same. This finding may be relevant when deciding which measurements to include as input to more comprehensive driver models that address all phases of the overtaking manoeuvres.
The comparison of the threshold, accumulator and PID models shows that all three were highly variable at detecting steering away time for different drivers. Furthermore, differences were discovered in the detected steering away times between models fitted to test-track experiment and naturalistic driving data. The models should be further developed as part of the counterfactual simulations.


pp. 29–36. ISSN: 00014575. DOI: 10.1016/j.aap.2015.12.008 (cit. on pp. 7, 8, 10, 23).


Lubbe, Nils and Johan Davidsson (June 2015). “Drivers’ comfort boundaries in pedestrian crossings: A study in driver braking characteristics as a function of pedestrian walking speed”. In: *Safety Science* 75, pp. 100–106. ISSN: 09257535. DOI: 10.1016/j.ssci.2015.01.019 (cit. on pp. 6, 8).


Noort, Martijn van, Frederick Faber, and Taoufik Bakri (2012). “EuroFOT Safety Impact Assessment Method and Results”. In: Transportation Research Board (cit. on p. 3).


Part II

Appended papers