Piloting the Naturalistic Methodology on Bicycles

M. Dozza*, J. Werneke#, A. Fernandez†

* Applied Mechanics – Vehicle Safety
Chalmers University of Technology
412 96, Göteborg, Sweden
e-mail: marco.dozza@chalmers.se

# Applied Mechanics – Vehicle Safety
Chalmers University of Technology
412 96, Göteborg, Sweden
e-mail: julia.werneke@chalmers.se

† Applied Mechanics – Vehicle Safety
Chalmers University of Technology
412 96, Göteborg, Sweden
e-mail: andre.fernandez@chalmers.se

ABSTRACT

Cycling is not a safe activity in Europe and claims over 2000 lives each year. In Gothenburg, Sweden, cycling safety has not significantly improved in the last five years, possibly due to the increasing number of cyclists. This paper describes the effort to adapt the naturalistic driving methodology to bicycles at SAFER (Vehicle and Traffic Safety Centre at Chalmers) in Gothenburg, and how this data can enable novel analyses and the development of accident countermeasures to increase bicycle safety.

The equipped bicycles that are collecting naturalistic data in Gothenburg as well as the tools and algorithms used to visualize and pre-process the data are described. Further, this paper reports on how naturalistic cycling data can enable novel safety analyses by mirroring previous analyses on naturalistic driving data. Finally, this paper offers some lessons learned from our current experience in collecting and analysing naturalistic cycling data and suggests future safety analyses and development for the naturalistic cycling methodology.

Keywords: naturalistic data, cycling safety, cyclist behaviour, bicycle dynamics.
1 INTRODUCTION

According to the European CARE (Community Road Accident Database) database, in 2009, bicycle fatalities accounted for 6.6% of all road accidents in Europe and claimed 2,109 lives. Cycling fatalities decreased by 33% in Europe in the last decade showing the same trend as driving fatalities (Fig. 1). In Sweden, the number of severely injured passengers in cars decreased by more than 30% in accordance with the European statistics. However, the number of severely injured cyclists decreased only by approximately 15%. In Gothenburg, the situation appears to be worse than the Swedish average. In fact, while the number of injured drivers decreased significantly in the last five years, the number of injured cyclists has been roughly the same (Fig. 2).

![Figure 1](image)

**Figure 1.** Number and proportion of cyclist fatalities in Europe between 2000 and 2009 [1].

A simple explanation for the lack of improvements in cycling safety compared to driving safety is the increasing number of cyclists which has already been documented in another Swedish city, Stockholm [2]. If this explanation is real, cycling in Gothenburg will become even less safe and very soon. In fact, starting 2013 new tolls will be in place to discourage the use of motorized vehicles in Gothenburg and will largely affect traffic split encouraging of alternative means of transportation. In the last years, Gothenburg has prepared by potentiating public transportation, e.g. building new dedicated lanes and stations, but very little was done so far for bicycles.

The severity of the current situation justifies and motivates new research on cycling safety in Gothenburg, possibly using the state-of-art tools developed for traffic safety. A few research projects to understand and improve cycling safety problems have been performed or are now ongoing at SAFER (Vehicle and Traffic Safety Centre at Chalmers). PreBikeSAFE was the first cycling project at SAFER and piloted the naturalistic methodology on bicycles. Here, naturalistic data is data collected in real traffic without interfering with the cyclists’ daily activities and is currently the most promising methodology to improve traffic safety and guide the development of traffic accident countermeasures. PreBikeSAFE took advantage of the experience, tools, and facilities developed at SAFER for naturalistic driving studies such as euroFOT [3] and SeMiFOT [4] and which made SAFER a European leader in management and analysis of naturalistic driving data. PreBikeSAFE adapted the naturalistic methodology from previous naturalistic driving studies such as the 100 Car Naturalistic Driving [5], IVBSS (Integrated Vehicle-Based Safety Systems) [6], and euroFOT (the first large-scale European Field Operational Test on Active Safety Systems) [3] to bicycles in order to demonstrate the potential of such data to tackle the cycling safety problems. The BikeSAFE and BikeSAFER projects are now employing the methodology and the tools developed in preBikeSAFE to carry on a larger data collection of
naturalistic cycling data in Gothenburg. Further, the BikeSing project is now looking into how to combine this new data with the existing accident data available from the Swedish accident database (STRADA) to cast light on the causes of single-bicycle accidents (which are responsible for about 70% of all cyclists’ injuries in Sweden). The BikeCom project is also analysing the naturalistic cycling data collected in BikeSAFE and BikeSAFER to guide the development of a prototype cooperative application to improve cycling safety. Another project, CiCity (http://www.cycity.se/), is also collecting real-traffic data in Gothenburg from GPS devices installed on bicycles. Even if the purpose of CiCity is to address mobility issues, the collected data may benefit safety analyses too.

The scope of this paper is to present the current status of the ongoing effort to collect naturalistic cycling data at SAFER. The main results from our pilot, the preBikeSAFE project, are presented in relation to the current ongoing data collection in the BikeSAFE and BikeSAFER projects. The discussion of this paper encompasses 1) the technical performance of our equipped bicycles, 2) the potential of naturalistic cycling data to tackle cycling safety in a new and innovative way, and 3) the new challenges encountered while collecting naturalistic cycling data which were not obvious from our experience collecting naturalistic driving data.

Figure 2. Injured drivers and cyclists in Gothenburg (source Göteborgs Stad website, www.goteborg.se).

2 METHODS

A commercial bicycle (Scott Sub 40) was equipped with several sensors in order to collect naturalistic cycling data (Fig 3). The sensors included: two cameras (GoPro Hero; one directed forward and one to the cyclist’s face), a GPS, two inertial measurement units (Phidgets 1044; comprised of a 3-axis accelerometer, a 3-axis gyroscope, and a 3-axis magnetometer; one installed to the handlebar and one to the frame), two pressure brake sensors (Phidgets 3102; one on each of the brake pads), and a speed sensor (sensing the back wheel rotation). The bicycle was also equipped with a simple human machine interface consisting of a push button. This button was intended to enable the cyclist to report the time when safety relevant events occurred. By pressing this button the cyclist could save a time stamp in the data logger, so that the safety relevant event could be easily retrieved and discussed with the cyclist. The data logger was previously developed by the SAFER projects MASCOT and MASCOT2 [7]. Data collection started automatically once the cyclist sat on the saddle and was automatically interrupted at the end of the trip. All sensors (including cameras) were interfaced to and controlled by the MASCOT logger. Data from all sensors but cameras was saved with as 100-Hz sample rate (25 Mbyte/h). Cameras collected 30 frames per second full HD colour videos (1920x1080p; 6 Gbyte/h per camera). In addition, a wireless modem was installed on the bicycle so that data (all but videos) and diagnostic messages could be sent automatically to a server. Finally, a small LED was installed in front of the cameras for synchronization purpose. All data was saved on solid state supports.
Once the system had been verified, a volunteer rode the bicycle to and from work for one week. The data collected was pre-processed (to synchronize and interpolate data) and analysed for quality and consistency using a Graphical User Interface developed in Matlab and adapted from previous projects (Fig. 4; [8, 9]). Further, the video data was reviewed with the help of the volunteer to individuate safety relevant events. For each of the event a video comprising different data sources was generated to exemplify how of naturalistic cycling data can enable new analyses. Data analysis also assessed quality for each of the signals collected. Frequency, time, and correlation analyses were performed to verify the signals consistency with the physical constrains of the sensed measure.

Figure 3. Instrumented bicycle for analysis of naturalistic cycling data.

Figure 4. Screenshot from the preBikeSAFE data visualization tool.
Four more bicycles were then equipped in a similar fashion as the one in Figure 3. Since August 2012, 12 participants have completed the experiment, four more are collecting naturalistic data in Gothenburg, and a few more are scheduled. Each participant was asked to ride the bicycle for two weeks. In order to qualify to participate to this study, cyclists were required to be 25 to 60 years old and to use the bicycle for at least 40 minutes per day. Participants were asked not to transport children during the study. Subjective data, in form of questionnaires and interviews, is also being collected before and after the study. The purpose of these questionnaires includes accessing whether the two weeks of data collection looked in any way different from usual and how the bicyclist perceived potential safety-critical events. All collected data is securely stored at SAFER in dedicated analyses rooms with controlled access following the established procedures in previous studies (e.g. euroFOT and SeMiFOT) in order to guarantee data security and privacy according to the Swedish legal and ethical regulations.

3 RESULTS

3.1 Data quality

After extensive tune up in the validation phase and a few improvements executed on the fly during the pilot, consistent data was successfully collected from all sensors. Video data was correctly and consistently collected in all trips. Glare, night time, and rain had a relatively little impact on the intelligibility of the videos (Fig. 5); however, when the street was not properly lit, video data was not as informative about the surroundings. Analysis of the time stamps reported the loss of 1 sample with a 0.012 probability and the loss of two consecutive samples with a 0.00004 probability. All signals provided values in the expected range, with expected accuracy, and dynamics with exception for the speed sensor. The GPS took between 1 and 11 minutes to sense the satellites (median 3 minutes). After the first connection, the GPS signal was missing at times for a few seconds (generally less than 10 seconds). Frequency analysis from the accelerometers showed a peak in the frequency spectrum at 1 Hz, most likely caused by pedalling. The brake force sensors timely signalled the onset and offset of the brakes. The sensor for automatic start-up failed in a few cases to recognize the cyclist. A hardware change solved this issue during the pilot. The speed sensor did not perform as expected providing values not consistent with the GPS speed, correlation between GPS speed and wheel speed was from $r=0.2$ to $r=0.93$ depending on the trip.

Figure 5. Video frames from known challenging situations. Panel A: glare from the sun with fast change of lighting (opening the garage door). Panel B: combined effect of rain and darkness. Panel C: night time on a lit street. Panel D: night time on a bicycle lane with no street lights.
3.2 Safety relevant events

Review of video data resulted in the selection of six events (18-30s long) exemplifying different safety relevant situations which were used to motivate the need for further collection and analysis of naturalistic cycling data. The six videos can be accessed on youtube by searching for preBikeSAFE (note that the videos on youtube have a much lower quality than the collected videos). A video frame from all six videos is presented in Table 1 together with a list of the data visualized, a short explanation of the video, and a list of possible analyses enabled by such data.

Table 1. Selected videos to exemplify the potential of naturalistic cycling data.

<table>
<thead>
<tr>
<th>ID</th>
<th>Video Frame</th>
<th>Description</th>
</tr>
</thead>
</table>
| 1  | ![Video Frame](image1) | **Data represented** - GPS in Google Earth environment (A), forward video (B), speed (C), longitudinal acceleration (D), rear wheel brake force (E).  
**Description** - A cyclist meets a bus at a construction site.  
**Suitable Analyses** - Bicyclist behaviour, infrastructure safety at construction works. |
| 2  | ![Video Frame](image2) | **Data represented** - Forward video (A), lateral acceleration (B), speed (C).  
**Description** - A cyclist brakes on ice and falls.  
**Suitable Analyses** - Bicycle accidents causation. |
| 3  | ![Video Frame](image3) | **Data represented** - GPS in Google Earth environment (A), forward video (B), lateral and longitudinal accelerations (C, D), handlebar rotation (E).  
**Description** - A cyclist passes an intersection with red light.  
**Suitable Analyses** - Bicyclist behaviour in relation to traffic rules, interaction among cyclists and drivers. |
| 4  | ![Video Frame](image4) | **Data represented** - Forward video (A), cyclist video (B), speed (C), lateral acceleration (D), and handlebar rotation (E).  
**Description** - A cyclist overtakes another cyclist.  
**Suitable Analyses** - Bicyclist risk taking behaviour, cyclist attention to traffic rules, and interaction with other cyclists. |
| 5  | ![Video Frame](image5) | **Data represented** - GPS in Google Earth environment (A), forward video (B), speed (C), longitudinal acceleration (D), and handlebar rotation (E).  
**Description** - A cyclist navigates through a complex multiple intersection.  
**Suitable Analyses** - Bicyclist behaviour in relation to traffic rules, interaction among cyclists and other road users. |
| 6  | ![Video Frame](image6) | **Data represented** - Cyclist video (A), forward video (B), longitudinal and lateral accelerations (C, D), and handlebar rotation (E).  
**Description** - A cyclist negotiates a bicycle lane with pedestrians.  
**Suitable Analyses** - Bicyclist and pedestrian behaviour in relation to traffic rules, interaction among cyclists and pedestrians, infrastructure safety. |
As Table 1 shows, the six selected videos illustrate how naturalistic cycling data can support bicycle safety in a similar way as naturalistic driving data supports vehicle safety. In fact, in video 1 the safety critical situation was visible from the video and the cyclist behaviour explained by the other collected signals. The video clearly showed that the cyclist was riding on a one way road against traffic and the presence of a construction work and a bus which clearly challenged the cyclist to pass thought a very narrow opening paved with gravel. The failed attempt from the cyclist to brake was shown by the onset of the brake signal in conjunction with an increase in lateral acceleration and a negligible change in speed, indicating skidding. GPS info complemented the picture by creating the link to map data where more details about the infrastructures are available.

In video 2 the causes of the accident were clear and mainly related to the environment. It was evident from the video that the ground was icy and a blind curve was coming. GPS data also added information about the road being downhill. Brake signals showed the effort from the cyclist to decrease speed. Finally lateral acceleration clearly showed the impact and the new geometrical configuration of the bicycle after the crash. Videos 3 to 6 also exemplify how combining different data sources increases accuracy when reconstructing an event from video. For video 4 and 6, it also is worth noting how the cyclist’s face video gave access to secondary tasks and gaze behaviour, while the lateral acceleration was a predictor of the traffic complexity. Finally, the event reconstruction from naturalistic objective data could be complemented and compared with an interview to the cyclist to achieve a more complete picture and/or to assess the extent to which subjective data from interviews would be reliable and redundant with the objective collected data.

### 3.3 Larger scale collection

Unfortunately, at the time of writing, no final result was yet available from the larger collection of naturalistic data ongoing in Gothenburg. So far, 12 (5 females and 7 males) participants finished data collection and four more are collecting data. Three participants collected less than 3 hours for only one week. The other participants collected about eight hours of data on average during the two weeks of collection.

### 4 DISCUSSION

#### 4.1 Potential of naturalistic studies for bicycle safety

Nowadays, naturalistic driving data is one of the most promising tools to improve vehicle safety and has already showed its ability to drive legislation [10] and enable innovative analyses [8]. Current studies on naturalistic cycling data show that this data has the same potential. A few studies already started to express the potential of naturalistic cycling data. For instance, Gustafsson and Archer [11] used GPS devices and camera to monitor 16 cyclists in Stockholm. Even if the cyclists were not totally free to decide their routes, the collection happened in a naturalistic fashion. This study also shows how safety and mobility analyses can be served by the same data collection and suggests that the data currently collected in Gothenburg may also be used to better understand cycling efficiency and cyclist mobility.

In Australia Johnson et al.[12], collected about 128 hours of video data from 13 cyclists in Melbourne. Data was limited to forward video from a helmet-mounted camera and was used to identify and analyze crashes and near crashes. The data presented in this study can complement analyses on safety critical events by Johnson et al. [12] by providing kinematics, location, and information on the cyclist distraction. This extra information helps understanding the dynamic and causes of a crash and may be used to derive measurable safety indicators for the development of intelligent countermeasures. In addition, previous research from naturalistic driving data shows how vehicle dynamics and controls are fundamental for analysing safety critical situations [5] and driver reactions [13] and motivates our efforts to extend collection beyond video and GPS data. Nevertheless, more data comes at the cost of increased complexity and expenses which span from the data collection to the data analysis phase [14].
The six videos selected in this study show the increased benefit of collecting continuous data from multiple sensors from a bicycle in a naturalistic set-up. Such collection strategy has the potential to serve many different types of analyses, provide a solid baseline, and enable simulation while inevitably resulting in higher costs and complexity. An event-driven strategy, in contrast, would significantly decrease costs and/or involve more participants. However, it may also serve a specific type of analysis, not capture exposure, and provide an arbitrary baseline. Continuous collection of naturalistic data limited to video data would also imply lower costs than the collection strategy suggested in this paper but may result in a more time-consuming and less accurate analysis. In fact, all videos would need to be watched (not having kinematics or the push button signals to guide the analyst) and information from videos may not be complemented by bicycle dynamics and cyclist controls, thus limiting the accuracy of the events reconstruction. Finally, signals from bicycle dynamics and cyclist controls help selecting, clustering and visualize safety-relevant point in the videos (e.g. where extreme values of acceleration are identified). Thus, when collection is limited to video, data is harder to browse, classify, and analyse, because manual annotation becomes necessary and needs to be systematically applied to the whole dataset.

Developing a system for continuous collection of cycling data from video and extra sensors required several technical choices and trade-offs. The following section describes the reasons that drove the technical development of our logger.

**4.2 Technical discussion: equipped bicycle and collected data**

The equipped bicycle presented in this paper was inspired by previous studies on naturalistic driving data such as the 100 Car Naturalistic Driving Study [5], IVBSS [6], and euroFOT [3]. The adaptation of a data logger to a bicycle is not obvious. For instance, new technical requirements rise from installing a logger on a bicycle compared to a truck [15] or a car [16]. The bicycle logger presented in this paper is small, light, waterproofed, tolerant to shocks and vibrations, low cost, and use as little energy as possible. For cars and trucks, such requirements are of minor or none importance. A crucial point was also to determine which signals to collect keeping into account that the collected signals define the analysis framework. Following an analysis driven-approach, we compiled a list of variables of interest able to serve data analysis by best capturing the cyclist behaviour, the bicycle dynamics, and the surroundings. Such list was inspired by previous analysis work (e.g. [5, 6]) and our experience from the SeMiFOT and euroFOT projects (e.g. [17, 18]).

The variables collected in this study result from our best effort to collect as many variables as possible from our list while complying with technical and budget requirements. For example, video cameras covering 360-degree around the bicycle would be optimal to understand the surroundings and the interaction with other road users —especially when the bicycle is overtaken— however cameras are expensive and installations time-consuming. In this respect, this study had budget for two cameras and facing these two cameras one forward and one to the cyclist was a relatively obvious choice considering previous study on distraction (e.g. [10, 17]) and rear-end collisions (e.g. [5, 19]). We also considered the alternative to use a higher number of more economical cameras. However, less expensive cameras offered much lower image quality than the one chosen in this study and/or were not water-resistant.

Handlebar movement and brakes activity were highly prioritized because they represent an access to the cyclist response time —as steering wheel and pedal controls are an access for the driver response time [8]. The movement of the bicycle frame was also considered important to identify safety critical situations [20] and enable analyses of crashes and near crashes [21]. Our analysis shows that accelerations from the inertial measurement units installed on the handlebar and the frame correctly represent the bicycle dynamics; however the significant frequency component found at 1 Hz may limit the employment of these inertial measurement units to 1) estimate pose, 2) increase GPS accuracy [22], or 3) derive other kinematic variables.

The GPS sensor was also an obvious choice favoured by the low-cost and easy-installation of such device. We did not expect for the GPS signal to take so long (up to 11 minutes) to be reliable. A possible justification is that bicycles are often riding close to buildings which may shield
the signals from the satellites to the GPS sensor. Since GPS speed is normally not sufficiently reliable for analyses of safety-critical events, we equipped the bicycle with one speed sensor. However, such sensor did not perform satisfactorily as shown by the large variance found in the correlation with the speed recorded from the GPS. New sensors, able to estimate accurately speed from individual tires, are now being tested.

While the logger software guarantees synchronization of inertial measurement units, GPS, pressure brake sensors, and speed sensor, it does not guarantee the synchronization of these sensors with the cameras. In fact, even if the logger directly controls the cameras, cameras boot-up may vary up to 2 s from time to time. Currently the synchronization procedure is semi-automatic and is enabled by and LED light captured by the camera and time stamped by the logger. For larger collections this solution for video synchronization may not be sustainable and would require to be fully automated. Newer firmware promises better control on the camera and may solve this issue, however another solution might be to use lateral and vertical acceleration from the inertial measurement unit and see at which time lag they best match the relative motion processed from the camera.

Automatic start and stop of the system was a very hard requirement to meet. Nevertheless, the only alternative was to rely on the cyclists to start and stop the system and this alternative had two substantial drawbacks. The first was that having to start and stop the unit would have been an obvious reminder to cyclist that she/he was going to be monitored; the second was that, if the cyclist would have forgotten to start or stop the system, either loss of data or battery drainage would have occurred.

Many other sensors were considered for this study, most of them technically possible to interface to the logger. Out of these, 1) sonar sensors able to sense passing bicycles and 2) an encoder able to record the pedal movement are now prioritized. Our preliminary test show that the major drawbacks with sonar sensors are the cost and the complexity of the recorded signals, whereas sensing pedal movement is mainly time-consuming due to the extra wiring and not easy installation for the encoder.

4.3 Lesson learned from the larger ongoing collection

Although, no final result is currently available from our larger collection of naturalistic cycling data, a few lesson learned from the data collection and pre-processing phases are worth a mention. The amount of data collected from our subjects has a large variability and three subjects collected data only for one week instead of the planned two weeks. Reasons for the shorter collections include: 1) sickness and following convalescence, 2) bad weather, and 3) more convenient alternatives to cycling tempting the cyclist not to stick to her/his commitment. All these reasons influence naturalistic driving collection to a lesser extent and suggest that, when planning a cycling collection, a higher dropout than for driving collection can be expected. Privacy may be also a larger concern for our cycling studies than for our previous driving studies. The full HD colour videos used in this study have a significantly higher resolution compared to traditional videos from cars and trucks. As a consequence it is possible to appreciate new details, for instance read vehicles plates and recognize oncoming pedestrians and cyclists. Further, cyclists often take the bicycle all the way inside their households and it is hard to guarantee for the logger to be shut down by then. Finally, loggers on bicycle are more exposed to theft, tempering, and shocks, than loggers on cars and trucks, especially when bicycles are parked outside. Our recommendation is to fasten the equipment tidily to the bicycle and not hesitate to use locks to prevent access to cameras and circuitries. Also, our experience suggest that participants do not always store the bicycles in safe places and may leave them outside and unattended thus increasing the probability of theft and tempering. Finally, in our preliminary video analysis from our larger ongoing collection, we could already appreciate the advantages of having an event push button on the bicycle, which was also strongly suggested by the 2BESAFE project[23]. In fact, cyclists seem to use the push button responsibly, and, besides pointing the analyst to take a close look to a specific data segment, this information gives insights into what is actually perceived as dangerous by the cyclist himself/herself.
5 CONCLUSIONS

In conclusion, our studies show that collection of naturalistic cycling data can be executed in a similar fashion to naturalistic driving data, collecting similar signals and supporting similar analyses. In this respect, naturalistic cycling studies have much to learn from the more established naturalistic driving studies; nevertheless new challenges have also to be faced. Considering the severity of the cycling safety issue and the increasing number of cyclists, the potential of naturalistic cycling data to help improve traffic safety seems to be at least as large as the one for driving data. Unfortunately, the budget for bicycle studies is, at least in some countries, lower than the budget available for studies addressing other vehicles. Nevertheless, the lesson learned from naturalistic driving studies can improve efficiency and economy of naturalistic cycling studies compensating, at least in part, for the lower budget. In any case the current situation calls for joining efforts on cycling studies and for looking into new innovative ways to collect and analyse data, possibly leveraging on the higher level of openness that cycling data may offer compared to driving data.

6 ACKNOWLEDGEMENTS

The authors would like to acknowledge SAFER for the facilities. This study was supported by VINNOVA and Trafikverket via the projects preBikeSAFE, BikeSAFE, and BikeSAFER.

REFERENCES