Doctoral Theses at NTNU, 2008:200

Thomas Engen

Use and validation of driving simulators

NTNU
Norwegian University of
Science and Technology

Doctoral Thesis

Thomas Engen

Theses at NTNU, 2008:200

NTNU
Norwegian University of
Science and Technology
Thesis for the degree of
philosophiae doctor
Faculty of Engineering Science and Technology
Department of Civil and Transport Engineering

NTNU

ISBN 978-82-471-1083-6 (printed ver.)
ISBN 978-82-471-1084-3 (electronic ver.)
ISSN 1503-8181

NTNU
Norwegian University of
Science and Technology

Thomas Engen

Use and validation of driving simulators

Thesis for the degree of philosophiae doctor

Trondheim, June 2008

Norwegian University of
Science and Technology
Faculty of Engineering Science and Technology
Department of Civil and Transport Engineering

NTNU
Norwegian University of
Science and Technology

NTNU
Norwegian University of Science and Technology

Thesis for the degree of philosophiae doctor

Faculty of Engineering Science and Technology
Department of Civil and Transport Engineering

©Thomas Engen

ISBN 978-82-471-1083-6 (printed ver.)
ISBN 978-82-471-1084-3 (electronic ver.)
ISSN 1503-8181

Theses at NTNU, 2008:200

Printed by Tapir Uttrykk

Use and validation of driving simulators

Thomas Engen

Preface

When I started work on this thesis, I thought it would involve the very narrow topic of validating the use of a driving simulator, which would have made it relatively easy to answer the questions I had asked. I now feel that I have learned a great deal, but instead of finding the answers to my initial questions, I have many more new ones. I also planned at the beginning to only deal with strictly technical issues regarding the driving simulator. I ended up by trying to learn about new subjects, such as behavioural science. It soon became evident that the use and validation of the driving simulator had to be related to other traffic and behavioural research studies. Instead of covering a narrow topic, I ended up addressing a huge topic. It has certainly been an interesting learning process and I hope in this thesis that I can communicate some of what I have learned.

I have had a great deal of help in my quest to finish this work. In retrospect it has been more difficult than I would like to admit to learn about new subjects, and even more difficult to try to communicate in writing what I have learned. I would like to thank my mentor, Professor Stein Johannessen at NTNU for his valuable support, even when I struggled to finish my work. I would also like to thank my colleagues at SINTEF, Terje Giæver, Ørjan Tveit, Kristian Sakshaug, and Torgeir Vaa for their contributions to and encouragement for finishing my work. I would also like to thank the Norwegian Public Roads Administration, and especially Even Myhre and Hans Skjelbred for their financial support and valuable input. Finally I would thank my family: my wife Nina, my daughter Aurora and son Odin, for being there when I needed them.

Thomas Engen, Trondheim 2008

Contents

Figures

Tables

Summary

NTNU and SINTEF have cooperated in the development of a behavioural research laboratory, which currently consists of three parts: a driving simulator, an instrumented vehicle and a traffic monitoring laboratory. Work is also under way in conjunction with the Norwegian Public Roads Administration to develop an instrumented road. These tools can be used to collect traffic data such as information about driver behaviour in traffic.

Driving simulators are becoming more and more common, particularly in driver's education courses. The ability to change modules and databases distinguishes a research driving simulator from a driving simulator intended for teaching. NTNU/SINTEF has acquired both the tools and the knowledge to create 3-D models based on several different sources and the ability to create complex scenarios in the driving simulator.

Observational studies have been most common in traffic research. Experimental studies have been difficult to conduct. Observational studies have been strongly criticized because of their shortcomings related to statistics. The use of both observational studies and experimental studies gives us new opportunities to both control our results, and to extend our understanding of the reasons for the results.

The alternatives to driving simulator data collection are to use either an instrumented vehicle or roadside data measurement equipment. Both have their advantages and disadvantages. Roadside data measurement equipment registers a great number of vehicles, but the measurements are usually related to one point. Instrumented vehicles measure vehicles over a stretch of road, but the number of vehicles measured is usually limited.

When we choose to use a driving simulator to conduct research, the results should hopefully be an accurate and correct representation of how we drive in real traffic. To obtain this knowledge we need to validate the research from the driving simulator, to see if the driver behaves the same way in a driving simulator as in real life.

The term validity can be described as "to refer to the approximate truth of an inference". Evidence of validity may come from other sources of knowledge, such as from previous findings and theories. Validity judgement cannot be absolute.

I have focused on different threats to validity, and discuss four types in this thesis:

- Statistical conclusion validity,
- Internal validity,
- Construct validity, and
- External validity.

I have tried to compare real world data with results from the driving simulator and to use the four threats to validity to ensure that the results really are accurate representations of the truth. Even if the results from real world studies and simulator studies are the same, it cannot necessarily be concluded that the driving simulator is

validated. There is both a need to compare real life data and driving simulator data, and to ensure that the threats to validity are reduced as much as possible.

Three case studies have been conducted to compare data from driving simulator experiments with real life data. These are: reaction time studies, speed and lateral position studies, and time gap studies.

The reaction time studies conducted in the driving simulator were compared to real life measurements, previous research, and measurements of reaction time in a video-based simulator.

The reaction time found in the driving simulator varied a great deal in differing situations, but this was reasonable and comparable to the results from all the other measurement methods. In this case study, the most important threat to the validity of the test was that subjects might learn the purpose of the study. In the case of reaction time it is very important that each situation will be a surprise to the test subjects, but we found that the subject's alertness level increased after only one incident.

In the case of speed and lateral position, it was found that the driving simulator gave results similar to those measured in the real world. It should be emphasized that even though the measurement of speed and lateral position is relatively easy, finding one real world speed that can be compared to data from the driving simulator is not easy. The difference in statistical mean could be just as small compared between simulator and roadside measurements as to between different roadside measurements.

The most important result from the study of speed and lateral position was that the driving simulator results have less standard deviation than real world measurements. This is to be expected, because real world measurements are more prone to influence from stochastic variability. The control of confounding variables possible in a driving simulator can create more exact results, but at the same time, there is a need for a good understanding of this confounding variable to be able to create sound scenarios.

In the case of time gap measurements, the importance of understanding confounding variables is even more evident. This case study was meant specifically for testing this method and not for finding the precise time gap. The measurements both in the driving simulator and the instrumented vehicle were done as add-ons to other research projects. The driving simulator was designed in an overly simplistic way as compared to a real world situation, which led to a very small time gap. Similarly, the lack of ability to control both the instrumentation and the traffic situation probably led to too large a time gap in recording as compared to a queued situation. As was found in the speed and lateral position case study, the standard deviation of the simulator study was much smaller than the standard deviation of the instrumented vehicle.

The driving simulator is an addition to existing methods of collecting traffic data and information about human behaviour in a traffic situation, but it cannot replace existing methods completely. Its primary strength is in controlling the confounding variables, but at the same time, this is its most challenging characteristic, because of the demand for a great deal of knowledge to create good scenarios.

1. Introduction

1.1 Background

NTNU[1]/SINTEF[2] have in cooperation developed a behavioural research laboratory. At the present time it consists of a driving simulator, an instrumented vehicle and a traffic monitoring laboratory. Work is under way in conjunction with the Norwegian Public Roads Administration to develop an instrumented road. These tools are used to collect traffic data such as information about driver behaviour in traffic. When I started work on this thesis I wanted to work with all of these research tools.

There was also a special need to look at the validity of the driving simulator as a research tool, both as a basis for our research at NTNU/SINTEF and as an asset when promoting the driving simulator for contract research. The validity and use of the driving simulator in research have to be compared to other ways of collecting traffic data. An important part of validation is to know what kinds of research experiments can be done with the driving simulator and the driving simulator's limitations. A general description of driving simulators, previous research and a comprehensive description of the NTNU/SINTEF driving simulator are therefore given in Chapter 2.

1.2 Objective of the thesis

The main focus of this thesis is to compare the research and the results from the driving simulator with whatever other sources are available, in order to provide more knowledge about the use and validity of driving simulators. This will hopefully be achieved through comprehensive literature studies of simulator validity research and comparison to alternative research tools, supplied by three case studies (Reaction time, Speed and lateral position, Time gap), carried out as part of my research work.

I hope this thesis will extend the use of driving simulators. We need a better understanding of the situations in which the driving simulator can best be used, and improved knowledge of the validity of the driving simulator as a research tool.

1.3 Traffic data

Road traffic is a complex system in constant flux. There is a need in the fields of traffic engineering and traffic safety research to describe the system itself and describe the effects of the changes in the system. Sometimes we assume a causal relationship between a measure and the results, such as in traffic safety outcomes, but often a cause – effect chain is needed to validate results. A great deal of traffic data has to be collected to do these things. The traffic system may be described by three components: the driver, the vehicle, and the road. The interactions among these components are equally important.

---

[1] NTNU = Norwegian University of Science and Technology

[2] SINTEF = The abbreviation SINTEF means The Foundation for Scientific and Industrial Research at the Norwegian Institute of Technology (NTH).

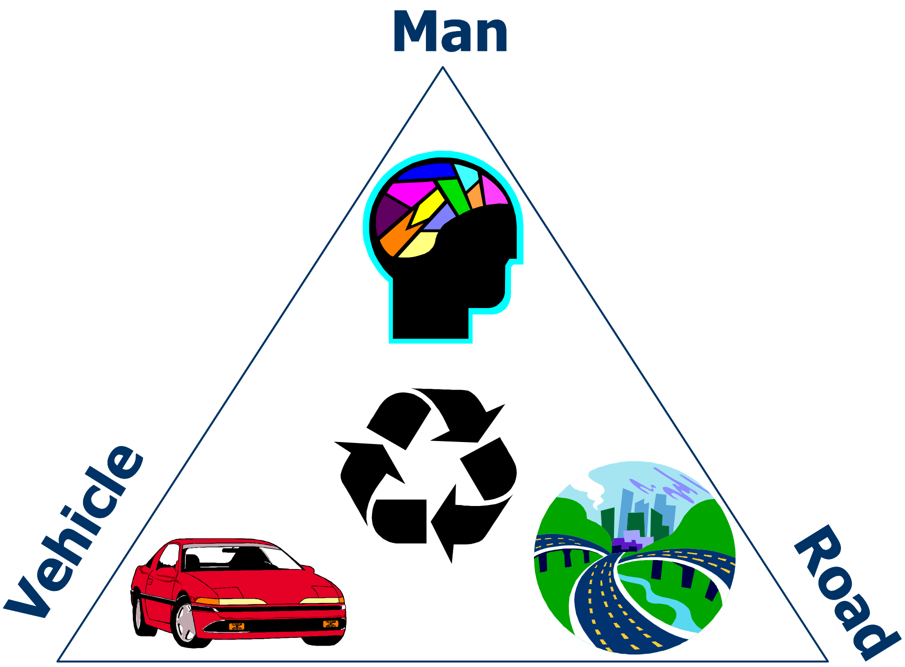


Fig. 1. The interactions among the driver, the vehicle and the road.

A driving simulator can be used to study all three factors and the interactions among them. In my research, I have focused on the use of a driving simulator to study the driver, the road and the interaction between driver-road and driver-vehicle. Car manufacturers are likely to have a different focus and be most interested in studying vehicle behaviour and the vehicle-driver and vehicle-road interactions.
Traffic and road user data are used to expand the understanding of road traffic and to evaluate projects designed to improve road traffic. We need traffic data from a number of areas in traffic research, including traffic safety, traffic regulation, traffic flow theory, road maintenance and environmental effects.

Within some of these areas, however, there is a lack of accurate data about how we behave in traffic in Norway. One example is in the use of micro-simulation models for traffic. Today, standard data for behaviour and variation of behaviour are often provided by the makers of the micro-simulation model. There is a need for research to make sure these data are valid for Norwegian conditions, or find better data for Norwegian conditions.

1.4 Behavioural research

The main focus of the NTNU/SINTEF simulator is behavioural research in realistic traffic conditions. The behaviour of the driver is related to his capabilities, the road and the vehicle. (Cozby 2003) describes four general goals of behavioural research: describe behaviour, predict behaviour, determine the causes of behaviour, and understand or explain behaviour.

The two main methods for behavioural research are observational studies and experimental work. In traffic research, observational studies are the most common. Experimental studies have been difficult to conduct.

In observational studies, behaviour is observed as it happens naturally. This is in contrast to experimental studies, which involve direct manipulation and the control of variables.

Observational studies have been strongly criticized because of their shortcomings related to statistics. A great deal of work has been undertaken to improve the quality or validity of the results from these kinds of studies.

(Hauer 1997) described two types of observational studies: before-after studies and cross-sectional studies. In before-after studies, the researcher looks at the change in behaviour when a change is introduced. In cross-sectional studies, a group with some common feature is compared with another group that does not have this feature.

(Elvik 1999) discussed the validity of evaluation research by means of meta-analysis. He described evaluation research as applied research designed to measure the effect of public measures taken to reduce social problems, such as road accidents. (Elvik 1999) provided validity criteria for evaluation research.

In experimental research, researchers can control the cause and measure the effect. It is possible to control variables in ways that are not possible in observational studies. On the other hand, it may be difficult to know if the experimental situation really represents the real world.

There are different tools in experimental research that range from simple medical and cognitive tests to more complex test scenarios. One such complex tool for a behavioural experiment for traffic research is the driving simulator. The simulator is becoming more and more common and researchers are getting more and more experience in using simulators in studying behaviour.

The use of both observational studies and experimental studies gives us new opportunities to both control our results, and to extend our understanding of the reasons for the results.

1.5 Traffic data collection at NTNU/SINTEF

NTNU/SINTEF is home to a laboratory for the analysis of road user behaviour, which is composed of three parts: Driving simulators, an instrumented vehicle, roadside logging equipment. A forth part, and an instrumented road, is under development.

One of the driving simulators is older and video-based, while the other is relatively new, with a graphical interface. Today the video-based driving simulator is primarily used to evaluate individuals to see if they have the physical abilities needed to hold a driving licence. This is done by measuring steering precision and reaction time to information presented on the video image. The graphical simulator is mainly used for research. The driving simulator can be equipped with both a lorry cabin and a full-size ordinary car body. The focus of this thesis is the graphical driving simulator and the ordinary car body; thus when the terms simulator or driving simulator are used in this thesis, this means the NTNU/SINTEF graphical driving simulator that has been equipped with the car body. The NTNU/SINTEF driving simulator is further described in Section 2.4

NTNU/SINTEF has acquired an instrumented vehicle that will make it possible to make a great number of measurements in a moving vehicle. The vehicle is a 2005 model year Volvo V70 2.4s with an automatic gear shift. The instrumentation makes it possible to measure the vehicle's handling along with the positions of nearby vehicles. The vehicle is described in detail in Section 3.4 .

The NTNU/SINTEF laboratory also contains equipment for roadside observations of traffic data. This laboratory includes different types of equipment for recording speed, lateral position and vehicle length. We have recently also acquired a trailer with a pole that can be fitted with video cameras that can be extended 15 metres into the air. See Section 3.2 for additional details about the roadside measurement equipment.

Work is under way to develop an instrumented road. The idea is to both use existing equipment on the road and to create a place for testing new equipment. The background for this new laboratory is presented in Section 3.3 .

1.6 Validation

When I started working on this thesis, it soon became clear to me that it would be impossible to make a general assumption about the validity of the driving simulator, as such, as a research tool. I realized instead that validity would have to be related to the specific research questions.

In all research projects it is important to know how accurately the research findings represent the real world. There is no easy way of finding the truth in traffic research. Both observational studies and experiments have shortcomings. There is also a need to validate the results of traffic research both when conducting observational research and experiments.

Physical validation and behavioural validation are two main approaches to validating the simulator. Physical validation is the validation of parameters such as how the car performs as compared to a real world car. Behavioural validation is an assessment of how the driver reacts and performs within the virtual world of a simulation. This thesis deals mostly with behavioural validation but will provide a limited discussion of physical validation.

We must determine how accurately the driver behaves in a simulator as compared to the real world. The simulator allows us to put a research subject in a virtual and controlled world and see how the driver reacts to virtual incidents. Even though the researcher wants to know how accurately the driving simulator represents the real world, the underlying, and fundamental question should be: are the results of the research valid?

Research projects involve a number of critical considerations. Some examples are:

- Is there enough data to be certain that the result is not created by chance?
- Would a new research project that uses the same method give the same results?
- Can the findings from the research projects be transferred to other places or situations?

All these questions relate to different considerations concerning the validity of the results from the project. Section 4.5.2 proposes different criteria for the validation of the driving simulator.

1.7 Scope of the thesis

The focus of this thesis will be on research conducted in a driving simulator, and specifically to what extent the driving simulator can be used to increase our understanding of the real world.

All data recording methods have their strengths and weaknesses. There is a need to compare the range of applications for driving simulators, instrumented vehicles and roadside logging equipment. This thesis contains an extensive introduction to these measurement methods as a means of comparison. My study will use instrumented vehicles, roadside equipment and observational studies primarily as alternative measurement methods to improve the validity of the use of the driving simulator.

My thesis contains the following chapters, which are new and based on my research:

> Chapter 2.5 “Data model creation” – This chapter describes a new method for developing data models for the driving simulator.
>
> Chapter 4.5 “Threats to driver simulator validation” – In this chapter, general models of determining validity of behavioural research have been applied to the research that has been conducted in driving simulators.
>
> Chapter 5 “Validation case 1 – Reaction time” – In this chapter, research conducted to determine reaction times is evaluated to improve the validity of using the driving simulator as a research tool.
>
> Chapter 6 “Validation case 2 – Speed and lateral position” – This chapter describes research conducted to find the effect of changing the road-, lane- and shoulder-widths to improve the validity of using the driving simulator as a research tool.
>
> Chapter 7 “Validation case 3 – Time gap” – This chapter presents the results of tests of the instrumented vehicle as a research tool and compares these results to findings from the driving simulator.

The rest of the chapters are primarily a summery of existing knowledge based on a literature review and conversation with colleagues and other researchers.

1.8 Outline of the thesis

This thesis contains three main parts. Chapters 2 – 4 present a theoretical perspective on the use of a driving simulator and alternative measurement methods to collect traffic data. Chapter 2 contains a description of driving simulators. Their use and one method of modelling the real world are presented. Alternative measurement methods are presented in Chapter 3. The equipment of an instrumented car and its possible use is the main focus. What are the advantages and weaknesses of a driving simulator compared to or used in conjunction with an instrumented vehicle or roadside logging equipment? How can we be sure the data from the driving simulator is a valid representation of the real world? Chapter 4 presents a basis for validating data from driving simulators.

Chapters 5 - 7 form the practical part of the thesis, where I present the results from the three research cases. The focus is both on the collection and use of traffic data and the validity of the data from the driving simulator. Chapter 5 presents a research case which has as its purpose the determination of reaction time in road traffic. The data are collected from several sources, which are not directly comparable. However, this case provides an opportunity to look at both the collection of data and the validation issue. Chapter 6 concerns the validation of results from the simulator using information from roadside data logging equipment. The data recorded consists of lateral placement and speed. Chapter 7 presents data from an instrumented car that is used to validate data collected from the driving simulator. The purpose is to establish the time gap between vehicles on the “same” stretch of road in both the simulator and in the real world.

Chapter 8 is the last main part of the thesis, where I review the results from my theoretical and practical research and present recommendations for improving the validation of the driving simulator and the use of roadside measurement equipment, instrumented vehicles and a driver simulator.

2. Driving simulators

2.1 Background

2.1.1 Use of driving simulators

Driving simulators are becoming more and more common in traffic research. Simple driving simulators have been available for decades. NTNU/SINTEF obtained its first video based driving simulator in 1988, which is still in operation today. The first graphical simulator was acquired in 1999.

(Allen et al. 2000) describe several trends in simulator applications. Advances in sensors, electronics, processing, storage capability, and computational algorithms have allowed significant advancements in simulation, vehicle instrumentation, and operator and vehicle modelling. The use of moderate to low-cost PC platforms is increasing. The development of simplified 3D visual database modelling procedures has reduced the effort required to produce visualizations. There is also a trend to create the ability to move through models in real time so that viewers can determine their own trajectory and point of view in reviewing proposed designs and developments.

Driving simulators are becoming better and cheaper, which has increased their availability. Thus, it is important to know the capabilities and the validity of research that is conducted using driving simulators.

2.1.2 Different components in driving simulators

Driving simulators are actually made up of several components or modules that simulate the real world. Different driving simulators might have different modules or different constellations of modules. I have distinguished between five modules I feel all driving simulators must have:

- Driver input module;
- Data output module;
- Sound, visual, and movement model;
- Dynamic module of the interactive car; and,
- Traffic module for autonomous vehicles.

Driver input module

The driver input module is basically how the driver interacts with the driving simulator. The simplest driving simulators have a steering wheel attached to a table and simple pedals. The most sophisticated driving simulators have an entire vehicle cabin fitted into the driving simulator.

Data output module

All driving simulators should be able to record data on speed, lateral position and the driver's interaction with the vehicle, such as brake pedal position. More sophisticated driving simulators also have data acquisition of driver reactions, such as heart monitors and eye trackers. It is also possible in some driving simulators to calculate data such as pollution caused by the "drive".

Sound, visual and movement model

This module gives the driver dynamic feedback; current models provide driver feedback through vision, sound and movement. Both the graphical output in the driving simulator and movement have improved greatly over the last ten years. Graphics are now much cheaper and better than just a few years ago.

Dynamic module of the interactive car

The dynamics of the interactive car is important, particularly in trucks, where one vehicle can behave in different ways, depending upon whether it is fully loaded or empty. However, the dynamic module is also important for a car, and can in some instances allow the researcher to the dynamics to enable the simulated car to behave like a sports car or family car. Some dynamic modules can be very sophisticated. However, the most sophisticated data models for vehicle behaviour that are used in car design are not actually able to run in real time and therefore cannot be used in driving simulators.

Traffic module of autonomous vehicles

Micro simulations are widely used today to simulate traffic behaviour. Nevertheless, these models might be simplistic, allowing for calculations only once a second or permitting a vehicle to stay in only one lane. A driving simulator needs some kind of traffic simulation of the autonomous vehicles. Calculations have to be done several times each second and positions have to be calculated with fewer restrictions of movements. Autonomous vehicles have to take into account both the vehicle characteristics and human behaviour of other drivers. This makes it very difficult to create good traffic modules for the driving simulator.

2.1.3 Types of driving simulators

Driving simulators are often divided into three groups. In (Kaptein et al. 1996) these are defined as:

- Low level driving simulator;
- Mid level driving simulator – Uses advanced imaging techniques, projection screens, a realistic cab, and possibly a simple motion base;
- High level driving simulator – Typically provides close to a 360° panorama and an extensive motion base.

Low level driving simulators

Driving simulators of this type usually use one regular PC for graphics, traffic and all the simulator controls. They may contain extra devices such as a steering wheel, brake, throttle and so forth, but have no or very little motion simulation. As a result, they are relatively cheap to develop and implement. These simulators make it possible even for regular driving schools to use them as a part of the training needed to obtain a driving licence.

The most advanced video games can be viewed as low level driving simulators if they provide at least a steering wheel, brake and a throttle.

Mid level driving simulators

These driving simulators vary a great deal with respect to their setup and how they are equipped. They are both more advanced and typically much more expensive than their low level cousins. As a result, these instruments are often found at research centres. In comparison to high-level simulators they often lack either an advanced motion-based system or visualization. The driving simulator that is presently owned by NTNU/SINTEF is an advanced mid-level driving simulator.

High level driving simulators

These driving simulators typically provide a close to 360° field of view and an advanced motion-based system. The most advanced motion-based systems allow 6 degrees of freedom, which provides the driver a good sense of motion. Two high level driving simulators are described below.

The National Advanced Driving Simulator (NADS) is located at Iowa State University. It has been funded by the U.S. Department of Transportation. The NADS is described in the U.S. National Highway and Transport Safety Administration's NHTSA pamphlet (NADS 2006). The NADS consists of a 24-foot-diameter dome in which entire cars and the cabs of trucks and buses can be mounted. Each vehicle cab is equipped electronically and mechanically using instrumentation specific to its make and model.

Fig. 2. The NADS driving simulator

The most advanced driving simulator in Scandinavia is located at VTI[3] in Sweden. The simulator is described on the website http://www.vti.se. Currently VTI has developed their third generation driving simulator. The VTI driving simulator III was introduced in April 2004. It is built on a real vehicle chassis and has an advanced motion system. The front environment is presented on three screens in front of the driver and the environment behind the driver is presented in three mirrors.

The simulator is modular based. The chassis can be fitted with either a car or truck chassis. Fig. 3 shows that the vehicle in the VTI simulator can be moved along one axis while the vehicle in the NADS simulator presented in Fig. 2 can be moved along two axes. The linear movement of the VTI driving simulator can have a maximum acceleration of ± 0.8 m/s^2. The chassis can be placed so it is possible to test either lateral motion or longitudinal motion. The VTI simulator can be tilted with both a pitch and rolling angle. In addition the chassis is mounted on a vibrating table, which gives the driver a feeling of driving on a road surface.

---

[3] VTI= Swedish National Road and Transport Research Institute

Fig. 3. The VTI driving simulator III

2.1.4 Applications

The driving simulator can be used for different tasks. Today the main applications are related to testing driver behaviour; these applications are:

- Introduction of new technology and in-car IT-based services
- New road design
- New roads and traffic regulations
- New traffic equipment
- New safety features in vehicles
- Safety effects of drowsiness
- Safety effects of drugs and alcohol

Driving simulator applications will be illustrated in more detail in later chapters, with examples both from the NTNU/SINTEF simulator and from a literature review of research that has been conducted in other driving simulators.

2.1.5 Measurement parameters

Most data about vehicle handling can be measured using a driving simulator. The exact parameters may vary, but some common data that can be tracked are speed, lateral position, steering wheel angle, and throttle and breaking pedal pressure. An overview of the data that can be collected at the NTNU/SINTEF driving simulator is presented in Section 2.4.5 .

(Östlund et al. 2006) list eight driving performance measures or indicators of driving performance:

- Speed
- Time to collision (TTC), time headway (HWT) and distance headway (HWD)
- Brake reaction time
- Lateral position
- Time to line crossing
- Reversal rate (Number of changes in steering wheel direction per minute)
- Steering wheel variations
- Self-reported driving performance

In Chapter 5-7 I use the first four bullet points to compare driving performance in the driving simulator and the real world.

Even more advanced methodologies can be used to collect behavioural data in driving simulators. For example, (Östlund et al. 2006) define two methods to measure the physiological mental workload:

- Heart rate
- Skin conductance

It is possible to use even more detailed models of human factors related to driving simulator use. (Gray et al. 2007) give an introductory web course on human factors. Cognition can be particularly important as related to driving simulator studies. Cognition is human mental activity, encompassing perception, mental imagery, thinking, remembering, problem solving, decision making, learning, language use, and conscious direction of motor activities. Studies in cognitive functioning are performed by both psychologists and neuroscientists. It is too complicated to give an in depth understanding of this subject in this thesis, but I have attempted to present different research projects in driving simulators that rely on cognition to some extent.

One of the most important terms in cognitive measurement as related to driving simulator studies is workload measurement. (Gray et al. 2007) state that there are four generally accepted categories of workload assessment methods: subjective ratings, performance data, physiological measures, and analytical techniques. These assessment measurements are often used, but often with different technical terminology.

A review of the literature shows several examples of different methods that have been used to collect cognitive data measurements. The literature illustrates the diversity of both cognitive measurements and the use of indicators to assess these measurements. The technical terms used in the literature are adhered to as much as possible in this thesis to illustrate the diversity among researchers using driving simulators as a research tool.

Cognitive measurements used in driving simulator studies

Delayed event detection and degraded vehicle control were used in (Boer 2001) as a measure of general behavioural entropy. Behavioural entropy was used as a measure of driving performance. Reaction time to unpredicted peripheral events was used as a surrogate measure of event detection, while steering entropy was used as a measure of vehicle control.

Steering entropy was also used in a study by (Paul et al. 2005). The purpose of this study was to examine the decrement in driving performance caused by micro sleeping. The study tested the hypothesis that steering entropy is an indicator of increased erratic steering behaviour during micro sleep episodes in drivers with obstructive sleep apnea/hypopena syndrome (OSAHS). Steering entropy was calculated from a time-series history of steering angle data.

In a study by (Kircher et al. 2002), the aim was to identify the most relevant performance-based indicators to drowsiness. They tried to use alternative and easier psycho-physiological measures than brain activity (EEG). The most important performance-based indicators studied were steering wheel movement and steering wheel variability, time to line crossing (TLC), lateral position, and ocular dynamics (blink frequency and blink duration).

In a study by (Karlsson 2006), the most important dependent measures were found to be in-vehicle glance time and a steering wheel reaction time measure to evaluate driver distraction countermeasures. This study was conducted in a driving simulator with an in-vehicle information system as a distracter. Two countermeasures were used, a blue flash in the middle of the road and kinaesthetic brake pulse.

Eye movement measurements were used in a study by (Victor et al. 2005) to measure the demands of visual and auditory in-vehicle tasks as well as driving tasks. Two newer measures, percent road centre and standard deviation of gaze, were found to be more sensitive, more robust, more reliable, and easier to calculate than established glance-based measures.

(Slick et al. 2005) described workload as the amount of cognitive resources necessary to perform a task. They measured workload by manipulating and incorporating secondary tasks into a primary task such as driving. They studied the workload changes in teenager who were asked to drive while experiencing distractions. The study incorporated answering a phone and removing a plastic bottle top into a driving task. Objective performance indicators were velocity changes and lateral position in the lane, while subjective workload was measured with the NASA-TLX questionnaire.

In a study by (Campagne et al. 2004), driving errors were measured in terms of the number of running-off-the-road incidents and large speed deviations. The evolution of watchfulness level was measured by brain activity (EEG) recordings. The main question was if the occurrence of fatigue and drowsiness was accompanied by a modification in the driving performance and if this relationship partially depends on the driver's age.

(Wood and Hurwitz 2005) studied driver workload management during cell phone conversations. The study tested if intelligently suspending cell phone conversations during demanding driving situations would improve the driver’s performance and lessen subjective workload. The workload was measured by deceleration and delay in releasing the accelerator.

(Lenneman et al. 2005) studied the advantages that physiological measures can have over performance measures for detecting changes in the psychological processes required for driving related task performance. The heart rate was proposed as an indicator for evaluating the effects of conducting driving-related tasks. An environment with a single-task driving-only condition was compared to two dual-task, driving-with-a-secondary-working memory task condition.

Driver performance of drivers using cell phones with drivers who were legally intoxicated from ethanol was compared in a study by (Strayer et al. 2003). Speed, reaction time, distance to the vehicle immediately in front of the test vehicle, and the amount of force used while braking were used as indicators to making conclusions about the driving performance of the test subject.

2.2 Challenges in using driving simulators

2.2.1 The effects of different simulator properties

Research has been done on the effects of changing different variables in the driving simulator. These effects can also be viewed as opportunities for calibrating driving behaviour. Three important driving simulator variables are presented here: Visual information, sound and vibration, and motion base.

Visual information

There are several parameters related to vision that are important and can influence driver behaviour. The effect on driving speed as related to changes in optic flow and scene contrast were studied in (Pretto and Chatziastros 2006).The optic flow was manipulated by the motion of the road surface, while the effects of scene contrast were studied under fog conditions. The results showed that with an increased optic flow velocity, drivers slowed down, while with a slower optic flow they sped up. These behavioural effects emphasize the importance of optic flow for speed estimation in driving simulations. Furthermore, the simulated fog led to lower speed, and not to speeding. This result supports the interpretation that fog results in only peripheral portions of the scene being visible, where high angular velocities signal a higher driving speed.

(Kemeny and Panerai 2003) discussed how different parameters affected perceived speed. A special emphasis was placed on how the field of view affected the speed, but the effects of other parameters such as image resolution and motion were discussed. The researchers state that in driving simulators with a large field of view, longitudinal speed can be estimated correctly from visual information. On the other hand, they state that recent psychophysical studies have revealed an unexpectedly important contribution from vestibular cues in distance perception and steering, prompting a re-evaluation of the role of visual-vestibular interaction in driving simulation studies.

Sound and Vibration

The NTNU/SINTEF driving simulator has the ability to simulate vibration. Both the use of sound and vibration can influence driver behaviour. (Giacomin and Fustes 2005) have studied the subjective equivalence between steering wheel vibration and sound in two experiments. The first experiment used stimuli measured in an automobile when driving over a coarse asphalt road surface, while the second experiment used data obtained by driving over a 1.0 cm X 1.0 cm square metal bar. When all other conditions remained equal, the human response to the vibration was found to increase in relative importance with respect to sound in the case of short duration, transient, square metal bar stimuli.

Motion base

One of the main differences between high level and low/mid level driving simulators is a motion base. (Brünger-Koch et al. 2006) studied virtual driving with different motion characteristics. Braking manoeuvres were analysed and validated. Initial experiments were conducted in a new moving-base driving simulator to evaluate the participants' ability to control a car in a virtual environment with different motion platform characteristics. In addition, the subjects assessed the quality of the perceived motion. Driving behaviour in the simulator was compared to driving in the real world with an instrumented vehicle. The experimental data was analysed regarding the influence of motion-cueing parameter variations on braking behaviour. They conclude that difficulties with the correct estimation of speed and distance led to an overestimated time to collision, which caused a delayed braking initiation. Higher maximum deceleration and the occurrence of multi-modal braking profiles indicated a mismatch between intended and initiated decelerations. However, significant parameter effects showed the influence of the driving task and the presented motion characteristic on braking behaviour.

2.2.2 Deducing effects on road safety

We often want to know the road safety effect of different measures. This can be difficult to measure directly in a driving simulator. Good scenarios that may result in a traffic accident in a driving simulator can be very difficult to design. (Lerner 2001) studied the lack of connection between driver behaviour/performance studies and crash experience related to young/inexperienced drivers. Lerner stated that quantitative measurement of driver behaviour has been central in much of the systematic research underlying highway safety issues during the past forty years. It has contributed to the way in which we design roads, vehicles, training programs, signs and markings, and intelligent transportation systems. Yet Lerner also states that methods we use to conduct driver behaviour experiments have little or no connection with the circumstances under which crash events occur. This is particularly evident in problems related to young, inexperienced drivers. The paper discusses some of the systematic biases that characterize the quantitative driver behaviour research base regarding young drivers.

One possible approach is to use alternative measurement parameters. Speed is probably the parameter for which we have the best knowledge regarding the effects on road safety. (Nilsson 2000) described the Power Model, which was further developed in (Nilsson 2004). The Power Model describes the relationship between the mean speed of traffic and the number of accidents or accident victims. The risk attributable to speeding can be estimated by applying the model once the change in mean speed has been estimated.

An evaluation of the Power Model was presented in (Elvik et al. 2004), who conducted an extensive review of relevant literature. The researchers used a meta-analysis to synthesize evidence from 98 studies containing 460 estimates of the relationship between changes in speed and changes in the number of accidents or accident victims. The results were broadly supportive of the power model. They concluded that speed has a major impact on the number of accidents and the severity of injuries and that the relationship between speed and road safety is causal, not just statistical.

In (Garay-Vega and Fisher 2005), eye movement data was used to study why novice drivers are more prone to accidents. In addition to traffic signs and other traffic control devices, there are many cues that help drivers further predict the presence of a potential risk in the driving environment. These cues are called foreshadowing elements. It was hypothesized that because younger adults have much less experience on the road, it is more difficult for them to predict where potential cues might be positioned when foreshadowing elements are not present. They found that in general, when novice drivers saw the foreshadowing element, they were almost two times as likely to recognize a risk as when the foreshadowing element was not present. However, the foreshadowing element by itself was not enough to equalize novice and experienced drivers' behaviour. The percentage of experienced drivers recognizing the risk, given that they saw the foreshadowing element was almost double the percentage of novice drivers recognizing the risk, given that they too saw the foreshadowing element.

2.2.3 Familiarization with driving in a simulator

When using a driving simulator, the driver is both introduced to a new vehicle and to a virtual reality. It takes time to adjust to both of these new situations. (Clarion et al. 2006) conducted a study of relevant indicators to assess driver habituation period in a dynamic simulator. The paper associated physiological measurements and behavioural indicators to assess habituation to a dynamic-based simulator. A 5-minutes test period was conducted, during which drivers were told to reach and maintain a 110 km/h speed on a virtual highway under weak traffic conditions. To assess drivers' arousal state, skin resistance was measured on drivers' fingertips. Longitudinal speed and lane position were also monitored. Arousal level was found to be stabilized 2 minutes after starting test. Similar findings were observed for speed. Nevertheless, no stabilization was found for lane keeping during the 5-minute test. These findings suggest that at the end of the 5 minutes of familiarization, drivers were sufficiently used to driving in the simulator and were able to perform an experimental task without interference between task requirements and driving simulator habituation.

(McGehee et al. 2001) examined older drivers' steering adaptation on a high performance driving simulator. The objective of the study was to examine how long it takes for older drivers to adapt their steering control to a fixed-base driving simulator. The hypothesis was that older drivers achieve maximum training benefit in the first few minutes of a driving simulation. The results showed that older drivers needed about three minutes to adapt and get the "feel" of the simulator. Before this time driving behaviour in the simulator may not be representative of actual driving performance. These results provide preliminary support for assuming that an adaptation period as short as five minutes may enable drivers to adapt to the driving simulator and drive normally.

2.2.4 Simulator sickness

Simulator sickness and discomfort must be considered when conducting simulator experiments. (Allen et al. 2006) analysed simulator sickness as a function of age and gender. The primary objective of the project was to develop a PC-based program in conjunction with a low-cost driving simulator that could be used for screening and potentially retraining the psychomotor, attention, and cognitive skills of older drivers. A large range of sensory, perceptual, psychomotor, and cognitive tests were administered to validate the driving simulator and/or provide a comprehensive older driver assessment test battery. Sickness ratings after each of the 5 driving simulator sessions were obtained using the Simulator Sickness Questionnaire (SSQ). They found that consistent with the literature, the results suggested a higher incidence and prevalence of simulator sickness for older drivers and for females. The results also showed no effect of simulator sickness on simulator performance.

2.3 Video-based driving simulator at NTNU/SINTEF

The driving environment for the NTNU/SINTEF video-based driving simulator is based on an ordinary car that is cut in half right behind the driver. The road is presented on large screen in front of the vehicle. The image presented is a combination of a video image and computer graphics. The video image is stored on a laser disc that is controlled by a computer.

Fig. 4. Video-based simulator at NTNU/SINTEF

SINTEF introduced the video-based simulator in 1988. At the time, it would have been considered a mid-level driving simulator, but today it lacks the graphical capabilities to be placed in this class.

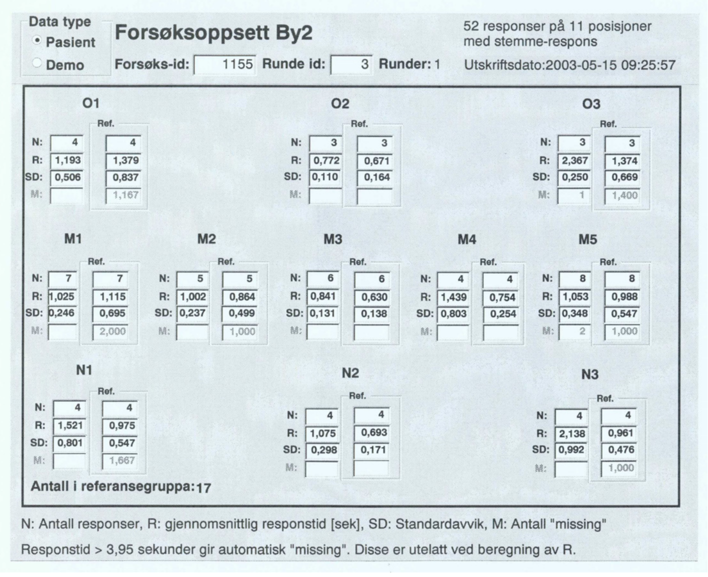
This video-based simulator is today primarily used to evaluate patients to see if they have the physical skills needed to hold a driving licence. These are patients who have undergone medical examination by a general practitioner and are sent to SINTEF for further tests.

This simulator allows the measurement of steering precision and reaction time. The driver can not influence the presentation of the video/vehicle, only give response to stimuli. The computer graphics are used to measure reaction time. Different symbols are presented to the driver, who must response to these stimuli. The symbols are presented to the driver in pseudo-random time intervals. This means that the driver perceives that the stimuli are being presented in a random order, although all the test drivers will have the same sequence of symbols presented at the same time intervals. This means that the test conditions are the same for all drivers.

There are two main response options, either by voice or by using buttons on the steering wheel. Both response options can be used in research where one wants a response to a symbol that is being displayed. It is also possible to make the task more complex by giving the test driver an option of pushing different buttons, depending upon the symbol being displayed. One example of this may be that the test driver pushes the left button when there is a round symbol, while the right button should be pushed when there is a triangle symbol.

Fig. 5. Response keys on the steering wheel (red on the left and green on the right)

When there is no response, the symbol will be removed after 4 seconds, and it will be recorded that there was no response. The accuracy of the measurements is 5 ms.



Data type
● Pasient
○ Demo

Forsøksoppsett By2

Forsøks-id: 1155 Runde id: 3 Runder: 1

52 responser på 11 posisjoner
med stemme-respons
Utskriftsdato:2003-05-15 09:25:57

	O1	Ref.	O2	Ref.	O3	Ref.
N:	4	4	3	3	3	3
R:	1,193	1,379	0,772	0,671	2,367	1,374
SD:	0,506	0,837	0,110	0,164	0,250	0,669
M:		1,167			1	1,400

	M1	Ref.	M2	Ref.	M3	Ref.	M4	Ref.	M5	Ref.
N:	7	7	5	5	6	6	4	4	8	8
R:	1,025	1,115	1,002	0,864	0,841	0,630	1,439	0,754	1,053	0,988
SD:	0,246	0,695	0,237	0,499	0,131	0,138	0,803	0,254	0,348	0,547
M:		2,000		1,000					2	1,000

	N1	Ref.	N2	Ref.	N3	Ref.
N:	4	4	4	4	4	4
R:	1,521	0,975	1,075	0,693	2,138	0,961
SD:	0,801	0,547	0,298	0,171	0,992	0,476
M:		1,667				1,000

Antall i referansegruppa: 17

N: Antall responser, R: gjennomsnittlig responstid [sek], SD: Standardavvik, M: Antall "missing"

Responstid > 3,95 sekunder gir automatisk "missing". Disse er utelatt ved beregning av R.

Fig. 6. Output sheet from the video driving simulator

Fig. 6 shows the output sheet for the comparison of norm data (mean, SD, and missing responses), matched by position in visual field, age, gender and km driven pr year. A set of border values for the parameters has been defined to assess if the patient should be deemed fit to hold a driving licence.

There are three different scenarios used to measure the reaction time in the video-based driving simulator. In all three scenarios, the driver has two tasks: centre a marker on the road using the steering wheel and respond to different stimulus. The scenarios are:

Viggja
The response to stimuli is given by one of two keys, making this a "choice reaction". In all there are 20 stimuli presented at 6 positions on the screen. The stimuli are traffic signs with a size equal to a reading distance of 10 metres. The driving scenario lasts 4 minutes.

Town1
The driver responds to stimulus by voice, making this a "simple reaction". In all there are 20 stimuli presented at 6 positions on the screen. The stimuli are traffic signs with a size equal to a reading distance of 10 metres. The scenario lasts 9.5 minutes.

Town2
As for the town1 scenario, the driver responds to stimulus by voice, making this a "simple reaction". In all there are 52 stimuli presented at 11 positions on the screen. The stimulus is a round symbol equal in size to child at a reading distance of 30 metres. The scenario lasts 11 minutes.

2.4 NTNU/SINTEF's graphical driving simulator

2.4.1 History

The graphical driving simulator was ordered from Autosim AS in 1998. It was financed mainly by the Research Council of Norway, but also by the Norwegian Public Roads Administration and the Nord-Trøndelag University College, Faculty of Education of Driver Instructors. The simulator was officially opened on November 24 1999. The graphical driving simulator is in frequent use and has been under constant improvement since then.

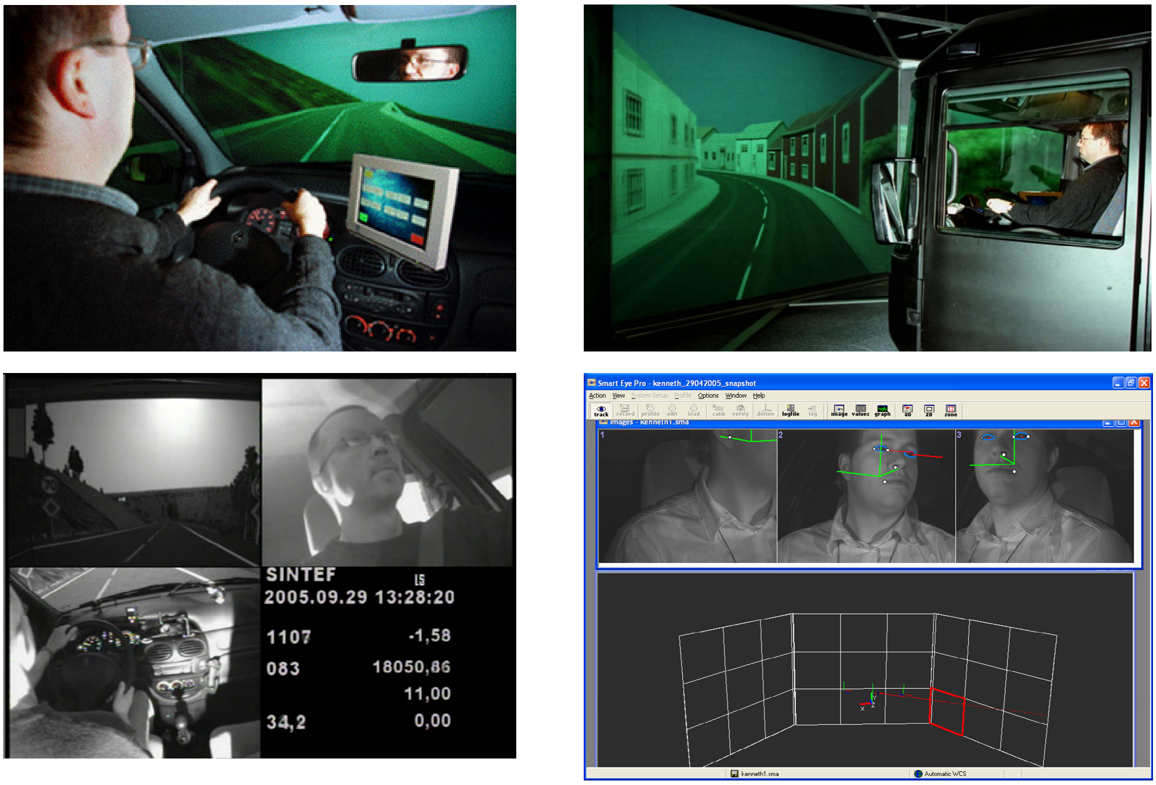


Fig. 7. NTNU/SINTEF driving simulator

2.4.2 Hardware setup of the driving simulator

Both a lorry cabin and an ordinary car cabin are available for experiments. The physical cabins and the dynamic module for the simulator software can be exchanged within hours. The car cabin has been the most frequently used of the two cabins and is the focus of this thesis.

The driving simulator started out as a mid level driving simulator. A Silicon Graphics Onyx Reality Engine 2 was used for the graphics of the front channels. This computer was later exchanged for three ordinary PCs running LINUX. Today all PCs employ the Windows operating system. The simulator initially had no moving platform, only a vibration system. A simple motion system with 3 degrees of freedom was later added.

Despite the development of the driving simulator, it is still considered a mid level simulator, but is now on the border of being a high level driving simulator.

The driving environment is today a Renault Scenic 1997 year model with a three-axis moving platform, a vibration system in the chassis and a four-channel sound system. The steering wheel is equipped with a motor to give force feedback.

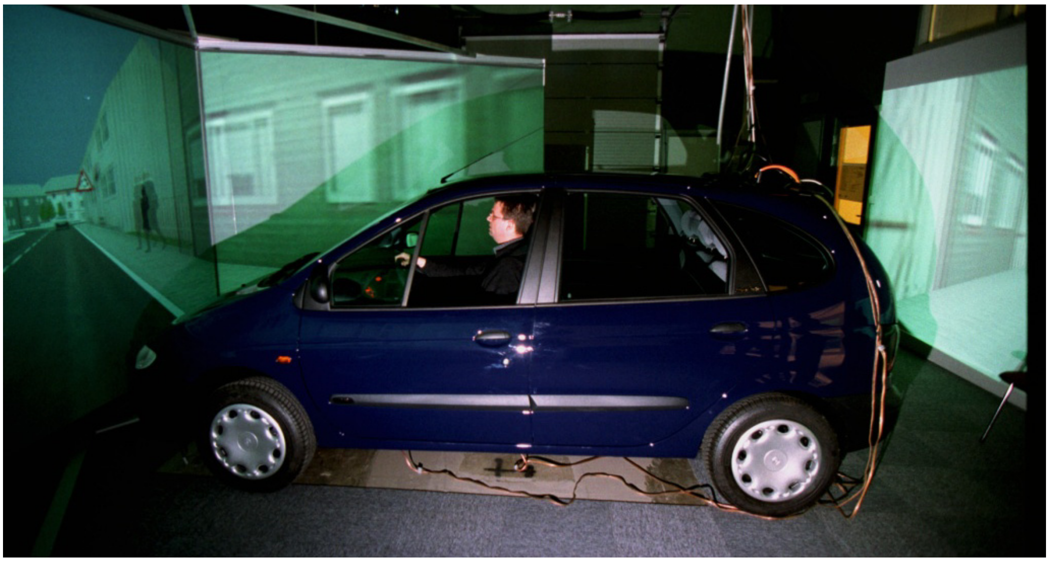


Fig. 8. The driving environment of the NTNU/SINTEF driving simulator

The visual representation of the road is presented on three screens in front of the driver and two screens behind the driver, for a total of five projectors. Each screen is 2.4 metres high and 3.1 metres wide. The resolution of all the projectors was initially 1024 x 768 pixels, but the centre front projector has now been updated to 1400 x 1050 pixels.

The visual system is based on PCs that run a Windows operating system. There are three PCs that run the front projection system and two PCs that run the back projection system. The three front screens are rear projected and provide in sum a 180° horizontal field of view and 47° vertical field of view. The two screens behind the vehicle provide in sum a 90° horizontal field of view and 47° vertical field of view.

The whole driving simulator is placed inside a separate housing. The driving simulator has facilities such as a break room for test subjects, a restroom, a special operator area and the ability to make presentations for larger groups of visitors. Fig. 9 shows the layout of the facilities. The figure shows a lorry cabin in use, while the car cabin is stored behind the screens.

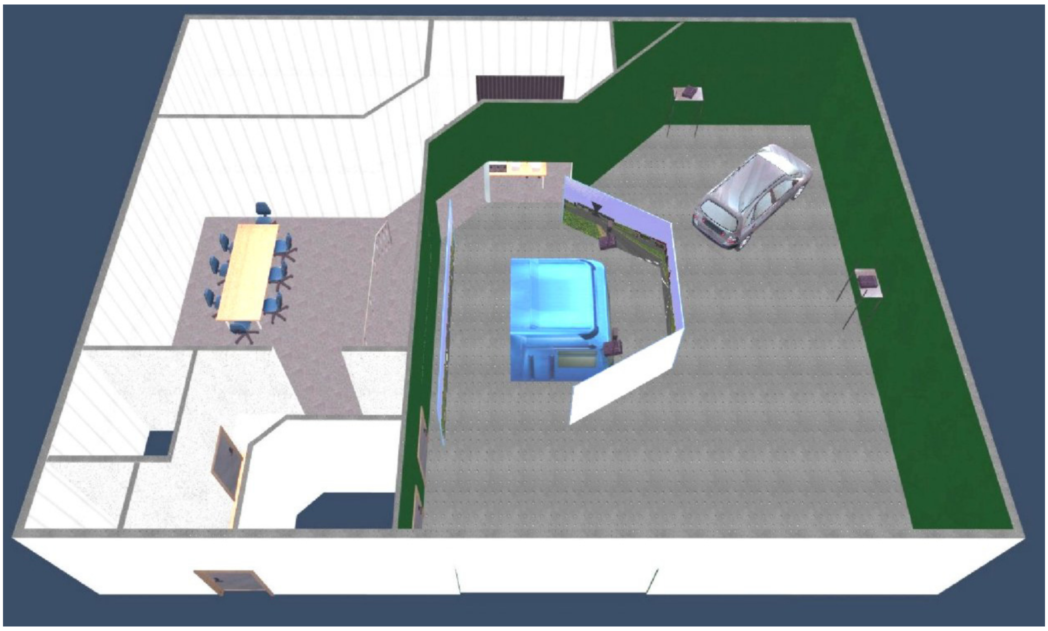


Fig. 9. The graphical driving simulator at NTNU/SINTEF

The car cabin has been equipped with a motion system. The motion system uses four actuators, one for each wheel, to obtain three axis of motion (pitch, roll and heave). The frequency of motion is below 10 Hz. Two vibration units are connected to the car body. These are able to reproduce frequencies above 10 Hz.

Sound is provided by a four-channel high fidelity sound system with loudspeakers inside the cabin and a subwoofer in the trunk. In addition, the system provides sound from the driver's vehicle as well as from other vehicles, and lets the driver experience both directional and Doppler effects.

Fig. 10 shows the hardware setup for the simulator. It shows the computers, network and the physical locations of the modules running the simulator.

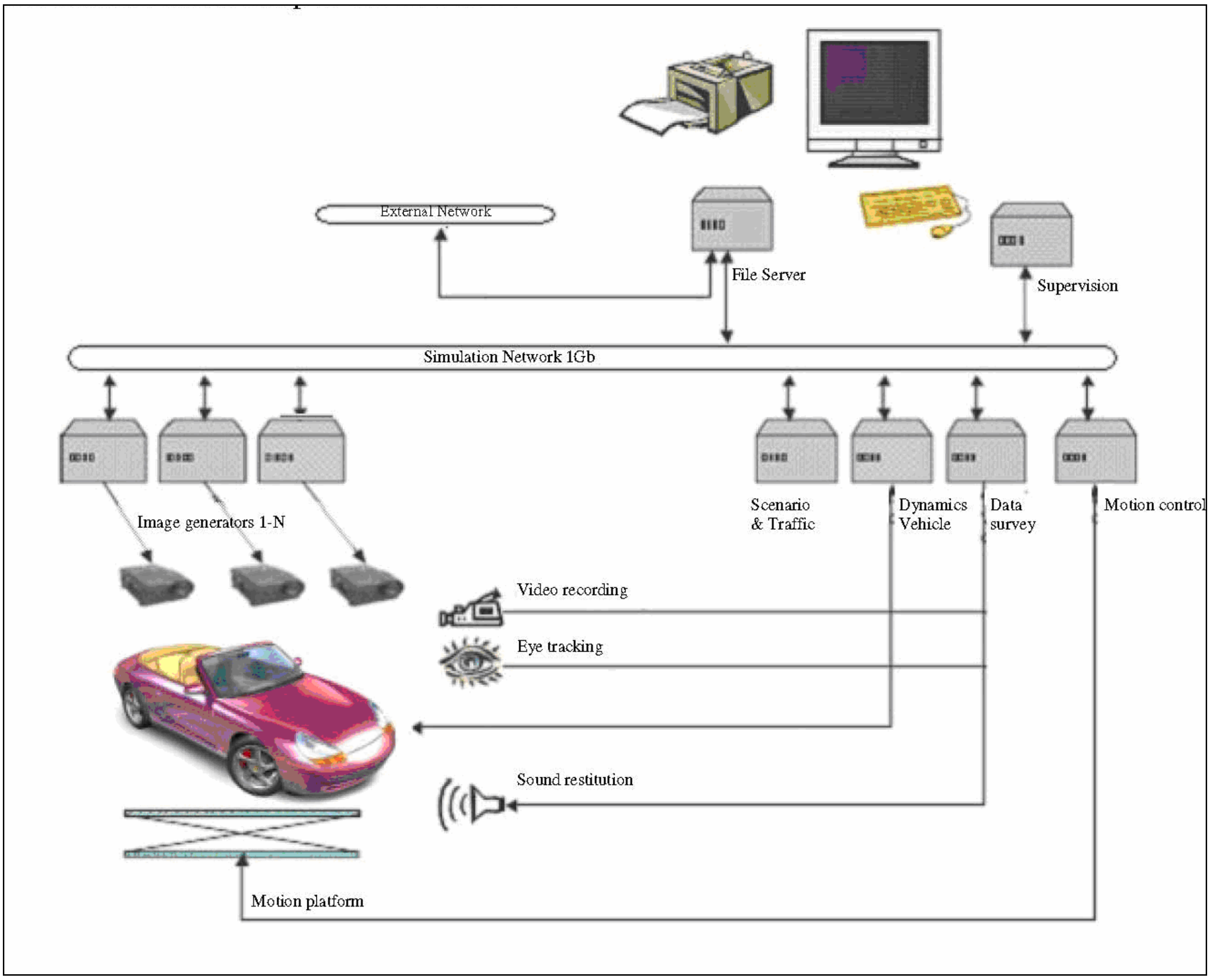


Fig. 10. Hardware setup for the simulator (Taken from (OKTAL 2006))

2.4.3 Software setup

The simulator software consists of different modules running in real time to perform special tasks. Fig. 11 shows the modules currently in use. All of the modules in the figure run in real time during an experiment except the Mice-module, which is used for creating different traffic scenarios before the experiments are conducted.

Certain modules make it possible to extend the use of the driving simulator. These additional modules make it possible to model:

- Variable light conditions, from daylight to darkness
- Time of year, from summer to winter
- Sight conditions, whether good visibility, fog, smoke or rain

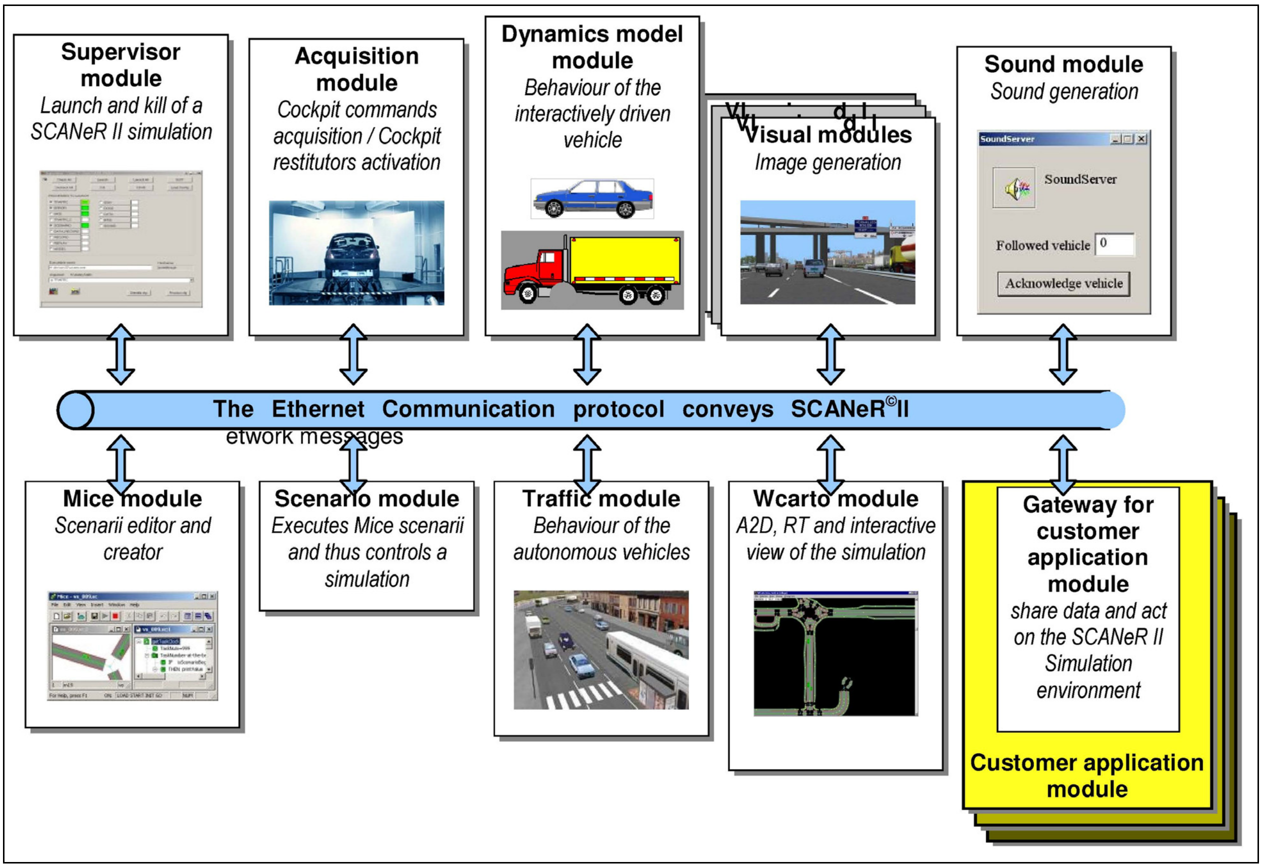


Fig. 11. Software modules from the driving simulator (Taken from (OKTAL 2006))

2.4.4 Description of databases

The simulator needs three different databases to work properly:

1. Terrain database
2. Model database
3. Scenario database

1) The Terrain database consists of three files:

- Road network
    - Describes the network for the autonomous vehicles
    - Describes traffic rules for the road network
- Road surface
    - Describes the road surface for the interactive vehicle
    - The road surface can be used to describe road friction for different road surfaces (For instance a road shoulder can have different friction than the rest of the road). Friction can be described using a coefficient from 0-1
- Visual database
    - Describes the visual content presented to the driver.

In addition it is possible, but not necessary, to have databases that describe infrastructure, collision boxes and animations.

2) The model database is described in two files. One file describes the vehicle characteristics, while the other describes the visual look of the vehicle. The model database can also be used to describe other objects, such as traffic signs or traffic cones. In the vehicle characteristics physical properties of other vehicle can be described.

3) The scenario database describes the action of the different vehicles. The traffic and behavioural parameters of the other vehicles are described in the scenario. The actions of the autonomous vehicles can be described in great detail or they can be modelled to reflect “intelligent” driving behaviour.

2.4.5 Measurement parameters

It is possible to record several parameters with the driving simulator at NTNU/SINTEF. Among the most important are:

- Position (x,y,z)
- Speed
- Lateral position on the road
- Distance to other vehicles
- Motor RPM
- Pressure on the brake pedal, throttle and clutch
- Parking brake usage
- Gear choice
- Steering wheel position
- The use of instrumentation in the car
- Use of lights: high beam-/low beam and indicator
- Use of emergency flasher, signal-horn and wipers
- Video recording of the driver
- Driver eye movements (using eye tracking equipment)
- Heart rate of the driver (using chest belt)

The car simulator uses three cameras to record what happens while driving. One camera captures the visual scene (forward direction), one camera captures the driver’s face, and one is mounted in the roof behind and to the right of the driver, capturing the instrumentation of the car and the driver’s hands. All cameras are mixed together using a video multiplexer, so that the result is one picture. Subtitles can be added to the video. The subtitles may contain data such as date, time, speed and frame.

Fig. 12. Sample video output with subtitling

2.4.6 Research using the NTNU/SINTEF driving simulator

The driving simulator is used both by NTNU and driver instructors education at Nord-Trøndelag University College for education and basic research. SINTEF has mainly used it for research topics in the following categories:

- Driving performance
- Driver assistance aystems
- Traffic regulation and road design
- Information technology
- Training methods
- Traffic medicine

A wide variety of research projects have been completed using the driving simulator. The main research projects that have been done at the time of the writing of this thesis are:

- ICT in road traffic
  (Rødseth et al. 2002) studied the use of information and communication technologies (ICT) in road traffic. The main purpose of the project was to establish routines and methods for collection, processing, and distribution of dynamic data in road traffic. The driving simulator was used to develop safe and user friendly ways to convey the information to the driver.
- Electronic data collection for freight transport in urban areas
  This project is presented in a report by (Wahl et al. 2002). The main purpose of the project was to test new methods to collect data about freight transport, where detailed data about trips were compared to freight information. The

simulator was used to evaluate the user interface for the driver in connection with the data collection.

- Road work warning evaluation
  This pilot study is presented by (Sakshaug 2002). The main purpose of this project was to look at how to present speed limits at road work sites. It specifically looked at the distance between signs and the steps in speed limit reduction. Four different sign plans were evaluated. No differences were found in speed between the sign plans.
- STARDUST – Toward Sustainable Town Development. A Research Development of Urban Sustainable Transport
  The project is presented on the Internet site http://www.trg.soton.ac.uk/stardust/. The objective of STARDUST project was to assess the extent to which ADAS (Advanced Driver Assistance Systems) and AVG (Automated Vehicle Guidance) systems can contribute to sustainable urban development, not only in terms of direct impacts on traffic conditions and environment but also in terms of impacts on social life, economic viability, safety, etc. The driving simulator was used to conduct human factors/driver behaviour studies.
- IMMORTAL - Impaired Motorists, Methods of Roadside Testing and Assessment for Licensing
  The project is presented on the Internet site http://www.immortal.or.at/. This research programme concerns the accident risk associated with different forms of driver impairment and the identification of 'tolerance levels' applied to licensing assessment and roadside impairment testing (including drug screening). The simulator was used to look at the effects of medication used by ADHD patients on their cognitive and driving performance.
- UPTUN - Cost-effective, Sustainable and Innovative Upgrading Methods for Fire Safety in Existing Tunnels
  The project is presented on the Internet site http://www.uptun.net/. The simulator was used to carry out experiments to find out how drivers perceive fires and react to them.
- CLARESCO – Car and truck lightning analysis.
  The project is presented on the Internet site http://www.claresco.net. The objectives of the CLARESCO project were to carry out an extensive assessment of advanced lighting technologies by studying advanced front-lighting systems for both **car** and **truck** lighting. The NTNU/SINTEF driving simulator was one of several driving simulators used in the project.
- The influence of pain killer medicine on driver performance. This is an ongoing research project.
- Design of roundabouts to ease bicycling. This is an ongoing research project.
- Finding design values for reaction time. This project is presented in Chapter 5.

These project examples show that the wide use of the driving simulator provides a great deal of new data on the simulator's performance, while the different uses show how versatile the simulator is for experimental research. At the same time the research can be seen in the context of other data sources, which illustrates how important it is to use the simulator as one of several research tools.

2.5 Data model creation

2.5.1 Creating new data models

Today it is possible to develop new terrain models based on several different sources. I have been involved in developing these models for the NTNU/SINTEF driving simulator. This section describes the development of these new models.

Visualizations of new road plans are becoming more and more common. Using a driving simulator to visualize plans creates new challenges. Among these challenges is the need to model vehicle behaviour and to make the visual database sufficiently sophisticated for real time display. At the same time there are several advantages to using a driving simulator:

- It is possible to experience the planned road in an in-vehicle environment.
- Traffic can be simulated (for instance traffic volume and car behaviour).
- The simulator allows the study of specific design implications regarding incidents and ease of understanding.
- Specific behaviour related to selected design criteria can be evaluated.

The use of visualization for road planning has traditionally been used to document road plans and to some extent for expert evaluation. The use of a driving simulator extends the application possibilities. The use of the driving simulator model for a new road can allow for:

- Documentation of road plans;
- Expert evaluation of the road plans;
- Experiments to be conducted before the road is built; and,
- The ability to conduct follow-up measurements on the road.

The creation of simulator models relies on a number of sources:

- Road plans created using the Novapoint software, information about the horizontal and vertical road curvature, terrain models, or paper drawings.
- Maps of different formats: digital maps (1:1000, 1:5000, 1:50 000, 1:250 000), raster maps, or paper maps.
- Approximate descriptions such as verbal or written description of the road
- Images, which are used mainly for textures.
- Graphics, which like images, are mainly used mainly for textures.

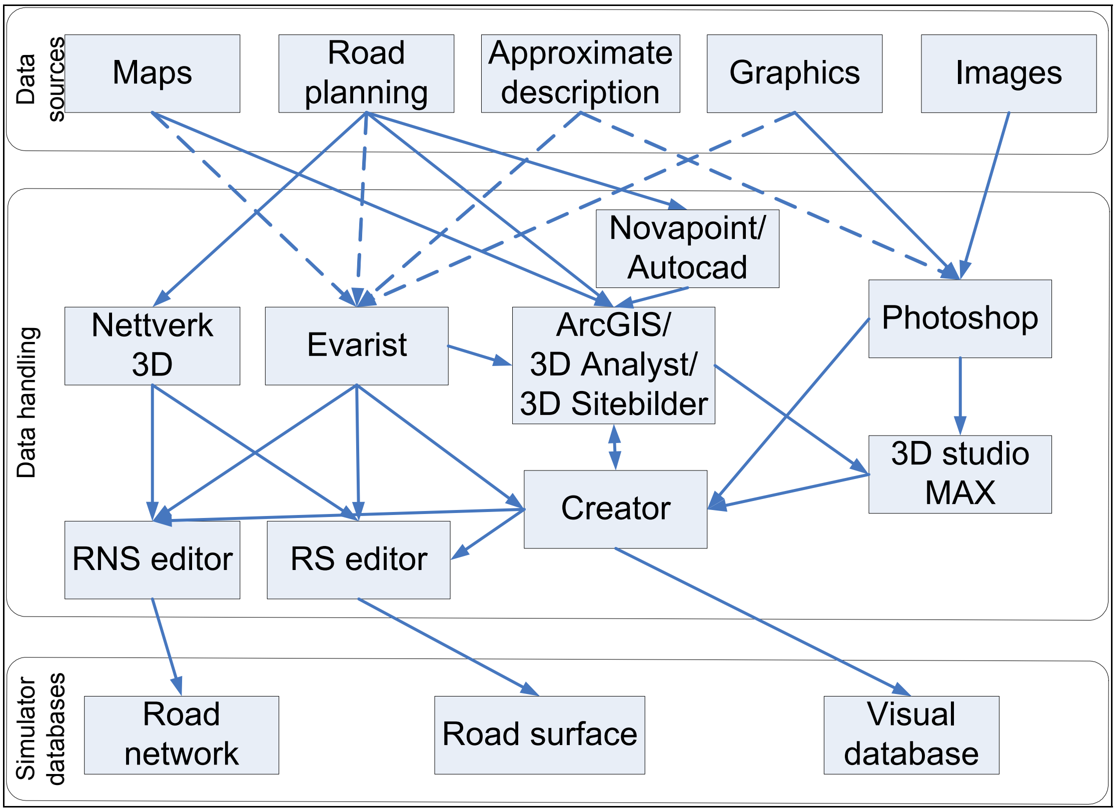


Fig. 13. Data flow showing the process for the creation of new simulator databases.

Fig. 13 shows the data flow from data sources all the way to simulator databases. The solid lines indicate direct data export/import, while dotted lines indicate manual data transfer. Different software packages must be used to create the different databases. The software used to create the databases are:

- Novapoint/Autocad - Used to study and make minor changes to road plans. Novapoint is the main tool used for creating road plans in Norway. Novapoint/Autocad is the best tool for generating ditches, cuts and fills.
- Nettverk 3D - Software developed at SINTEF to export road network and road surface information based on planned road data. Can import information about uniform road sections, but has limited capabilities related to junctions.
- Evarist - Software developed by Octal, the developer of the NTNU/SINTEF driving simulator, which is used to create larger road models. The software exports road network and road surface information, and a visual 3D model of the road.
- ArcGis - Used for combining several data sources. It is also used to combine data with different map projections. All new databases created for the driving simulator are based on UTM 32 or UTM 33 projections to make it easier to conduct direct comparisons with field studies using GPS data.
- 3D Analyst - An additional module of the ArcGis package from ESRI that is used to make 3D models.
- 3D Sitebuilder - Used to export the 3D model created with 3D Analyst in a format appropriate for visual presentation in the driving simulator.
- Photoshop - Used to manipulate textures used in the simulator to make objects look photorealistic.
- RNS editor – Used to create or modify road network models.

- RS Editor - Used to create or modify road surface modules.
- 3d Studio Max – Used to create special visual 3D models for the driving simulator. This is the main tool to used create road tunnels.
- Multigen Creator - The main tool for manipulating the visual 3D models used in the driving simulator. It is used both to create and to modify visual 3D models before they are used in the driving simulator.

My work has consisted of examining the possibilities and restrictions of each software package. Several of the software packages can accomplish the same results, but each has its strengths and weaknesses.

This work has been specifically carried out in connection with the creation of a road representation that is presented in Section 2.5.2 .

2.5.2 Data models created

There are more than 500 km of roads available for the simulator, including motorways, highways and rural roads. Most of the roads have been created by the simulator's manufacturer. However, I have been involved in the creation of five road models that are presented here.

E6 Støren – Soknedal (Sør-Trøndelag county)
Until 2004 there were no existing or planned Norwegian roads available for use as models in the driving simulator, except for some road segments through tunnels. In connection with research on reaction times, a 10-kilometre road was created for the simulator. This road consisted of existing roads and some that will be built. A map of the road is shown in Fig. 14.

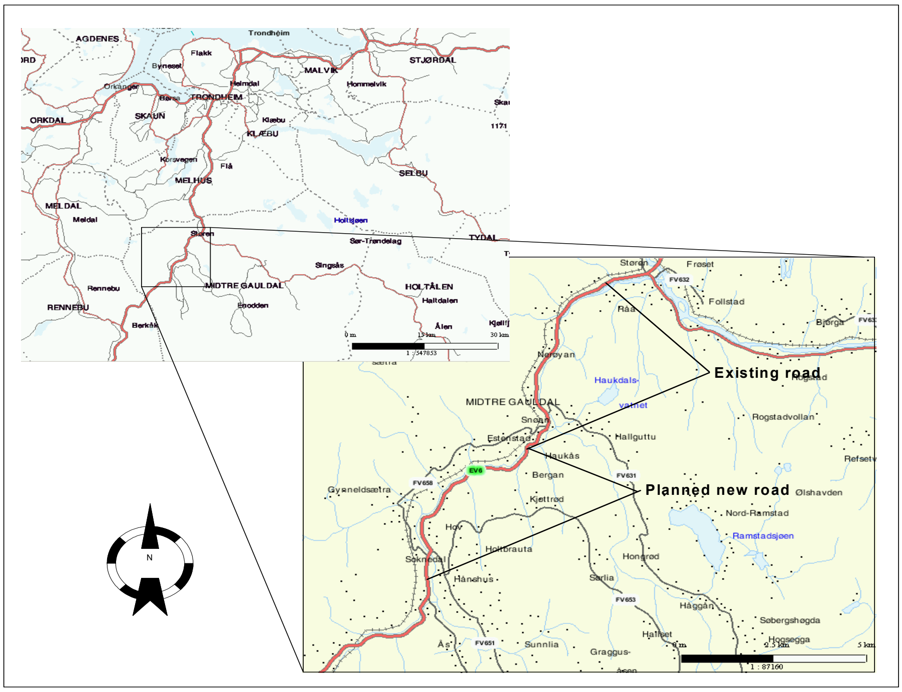


Fig. 14. Map of E6 Støren - Soknedal

The implementation of an existing road for the simulator makes it possible to conduct a direct comparison between observational behaviour on a real road and on a road created in the simulator.

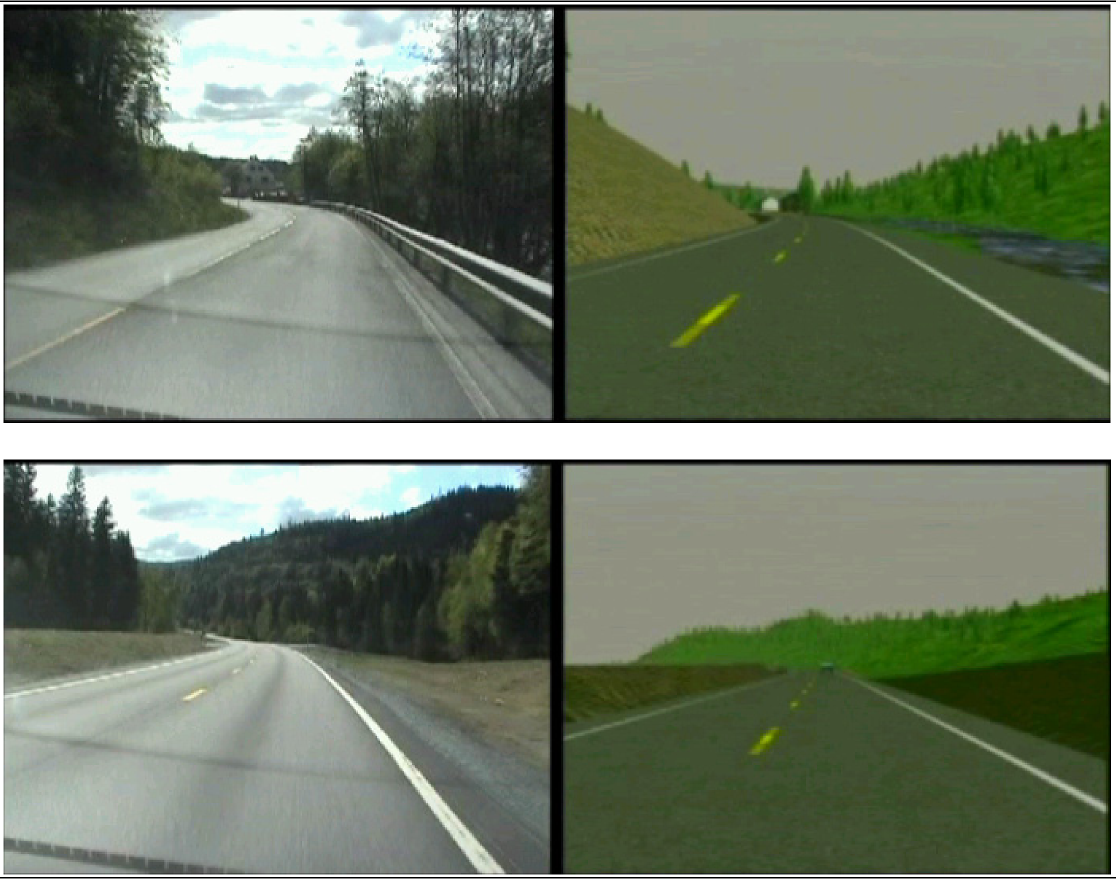


Fig. 15. Comparisons of an actual road segment and its simulator counterpart.

Fig. 16. Simulator photo of the E6 highway planned for Soknedal.

Fig. 17. Bird's-eye view of planned E6 highway at Soknedal.

The basic road is 8.5 metres wide. It was further developed to make it possible to use a 10-metre wide road in conjunction with the project described in Chapter 6.

E6 Lillehammer – Otta (Oppland county)

A total of 47 km of road is being planned for this section of the highway. Of this, 20 km of the planned road from Frya to Kvam is at the moment modelled for the driving simulator. The model is based on a very early version of the road plan. The final road design is planned for 2010. Two alternatives have been analysed: a two-lane road with a physical barrier between opposite lanes of travel, with the use of one extra lane for passing, and a narrow four-lane road.

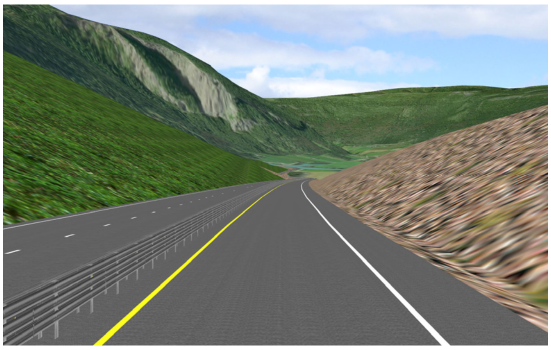


Fig. 18. Simulator model of the planned two-lane road with an extra lane for passing.

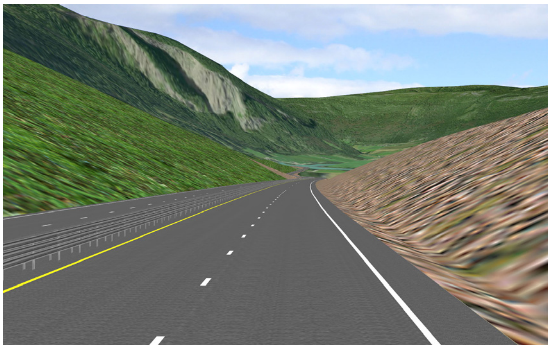


Fig. 19. Simulator model of the planned four-lane road.

Further development and refinement of the road model is planned. The model of the road in the driving simulator can be improved by increasing the detail and including two-level crossings. There is also interest in completing simulator models for the entire 47 km of planned road.

Fig. 20. Map showing the road between Ringebu and Otta.

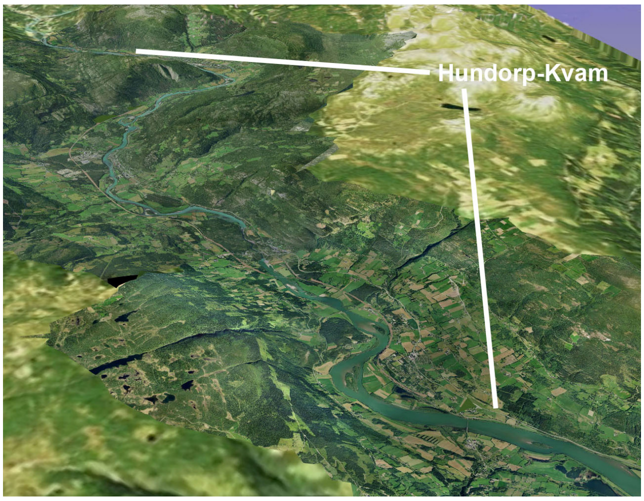


Fig. 21. 3D-model of the road model.

E18 Sky – Nøklegard (Vestfold and Telemark counties)

The road between Sky and Nøklegard has been planned at a relatively high level of detail. The road comprises four lanes, with a total width between 20 and 26 metres. The design includes 11 km of road, two picnic areas, 2 two-level crossings, six tunnels, and three bridges.

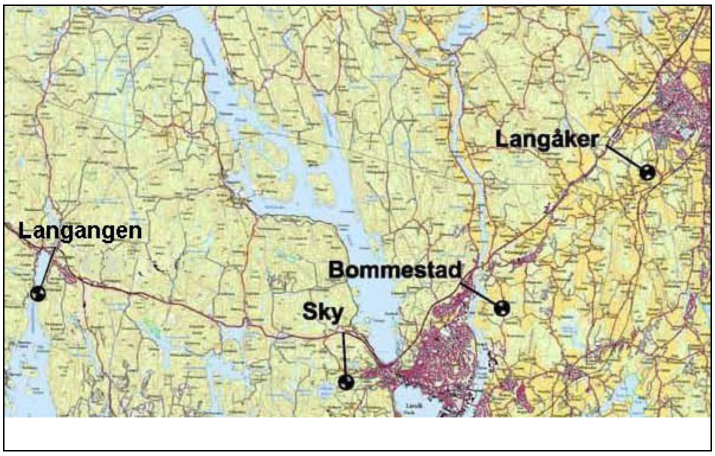


Fig. 22. Map showing the route of the E18 between Sky and Nøklegard (Langangen).

Tramway in roundabout (Oslo)

The purpose of this project was to study the same roundabout, but with different solutions for warning drivers of possible tramways in the roundabout. The roundabout model is based on a real world junction in Oslo. To accomplish this, the same roundabout was placed at a total of six different places in an imaginary road system. In total, seven possible signs and layout arrangements were designed to be placed at any one of the six junctions.

Fig. 23. Tramway in a roundabout.

Fig. 24. One of seven signing scenarios tested for tramway warning.

Road network in Trondheim

This model is based on both existing and planned roads. Because of the complexity of the model it has been based on six models that have been merged to form the total road network.

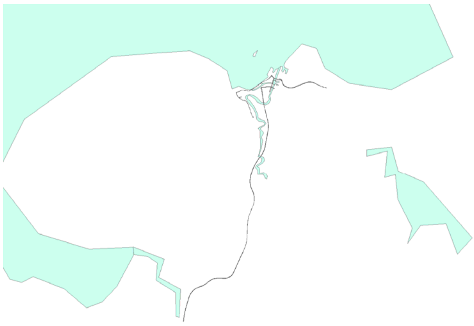


Fig. 25. The Trondheim road network available for use in the driving simulator

Fig. 26. Early version of Trondheim's Olav Tryggvason Street.

The Trondheim road network is comprised of the town centre, tunnels, and rural roads.

2.6 Summary

The NTNU/SINTEF driving simulator has been improved since it first was installed, with the development of new hardware, software and databases designed to improve the quality of the driving simulator. NTNU/SINTEF has the ability to develop both databases and to develop scenarios for conducting experiments. The two research institutes also have ample opportunity to influence software improvement, and to improve the hardware within the limits set by software developers.

3. Real-life traffic data collection methods

3.1 Overview

There are many ways to collect traffic data in real life traffic situations. I have differentiated between roadside data collection, and instrumented vehicles as two types of traffic data collection in this thesis.

3.2 Roadside data collection

Fig. 27. Roadside traffic data collection.

Roadside data collection is conducted at fixed points along a road or at a junction. The data collection can be either temporal or over a longer time span. One can also differentiate between single roadside equipment that is used by itself and collaborating roadside equipment that works together.

Single roadside equipment

There are several types of roadside equipment. Radar, loop detectors and video recordings are all common examples. Some of this equipment is especially suited for temporal logging because it is easily transported and mounted. Other types of equipment are better at conducting measurements over a longer time. This is specifically true where loops are buried in the road pavement to conduct measurements.

Fig. 28. Trailer with 15 meter pole for roadside video recordings.

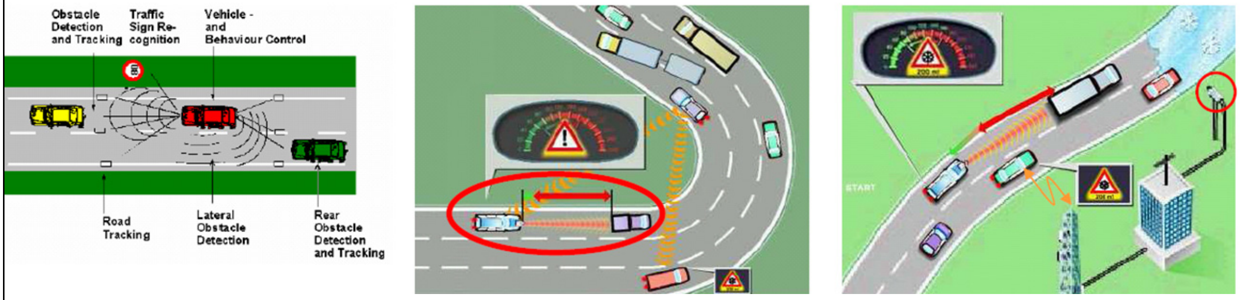
Fig. 29. Images from two video cameras mounted on top of the trailer's pole.

Collaborating roadside equipment

This is equipment where data from several roadside loggers are used in conjunction to give overall information about traffic. Number plate matching or the matching of toll tags are two examples of this type of coordinated measurement equipment. These types of measurements are often used to calculate travel time on specific roads.

3.3 Instrumented road

A great amount of measurement equipment is currently located on roads, such as detectors at signalled controlled intersections, automatic traffic control and video surveillance, but not all of this information is coordinated and presented in real time. With the instrumented road, SINTEF and Norwegian Public Roads Administration want to create a situation where this information is coordinated. This will also make it easier to introduce new equipment and to compare the results from new equipment with exiting equipment.



On-board systems +

E/E Architecture
Bus-systems (ex. CAN-bus)
Digital maps
Sensors

Vehicle-vehicle communication +

Comm.systems
Frequencies
Protocol

Vehicle-infrastructure communication

Comm.systems
Frequencies
Protocol
Services
Responsibilities
Cost sharing

Fig. 30. Overview of the instrumented road

An instrumented road will also make it possible to study the exchange of real time information between vehicles and the roadside, as is shown in Fig. 30. This will make it easier to test technologies such as driver assistance systems, which will be available in the near future.

3.4 Instrumented vehicle

3.4.1 Overview

There is a need to observe driver behaviour in real-world situations. An instrumented vehicle can be used both to register behaviour of the driver of the instrumented vehicle and of other road users. Using an instrumented vehicle to monitor other road user makes it possible to observe behaviour without the driver's knowledge in real traffic situations.

This thesis defines an instrumented vehicle as: "A vehicle equipped with measurement instruments that can be recorded and later analysed". The definition means that although advanced cars may contain many instruments that measure car performance to help the driver with the task of driving, they are not called instrumented vehicles. If the same instruments can be recorded, and the results can be analysed at a later stage, the vehicle can be called an instrumented vehicle. The difference between just being an advanced car and being an instrumented vehicle lies in the ability to record the data from the instruments.

3.4.2 Research

Instrumented vehicles may have various purposes and are equipped with a range of different instruments. The design ideology may also differ. With an instrumented vehicle, it is possible to either measure behaviour of the driver of the instrumented vehicle or drivers of adjacent vehicles. Usually the measuring instruments inside and outside of the instrumented vehicle are concealed. By concealing the fact that it is an instrumented vehicle it is easier to conduct research on the driver's actual behaviour without the research situation influencing driving performance.

An instrumented vehicle can also be fitted with special measurement equipment. (Hjälmdahl and Várhelyi 2000) wanted to study the workload for ISA drivers (Intelligent Speed Adaptation). They used two methods, a self-reported measure in the Raw Task Load Index (RTLX) and a secondary measure in the Peripheral Detection Task (PDT). The RTLX method is a simplified version compared to the NASA-TLX method. The PDT consists of responding to stimuli from two red lights mounted on the dashboard. When the light is lit the driver should respond by pressing a micro-switch placed on his thumb. Three variables are extracted from this; reaction time, percentage missed signals, and number of responses without stimuli.

It is also possible to formalize methodology so an accompanying researcher may record behaviour of the research subject. (Hjälmdahl and Várhelyi 2004) validated a method for driver assessment by in-car observations. They studied whether observers could be trained to observe safety variables and register a driver's behaviour in a correct and coherent way, and whether the drivers drove in their normal driving style, despite the presence of the observers. The variables studied by the observers were speed, speed adaptation, yielding behaviour, time gap, behaviour towards vulnerable road users at crossings, lane usage/ lane change, use of indicator and overtaking. In addition they studied in a descriptive way (qualitatively): conflicts, interaction, communication and special events. After training their observations were compared with a key representing a correct observation. The observers showed a high correlation with the key. To establish whether the test drivers drove normally during the in-car observations, comparisons of 238 spot-speed measurements were carried out. The drivers' speeds when driving their own private cars were compared with their speeds during the in-car observations. The analysis showed that the drivers drove in the same way when being observed as they did normally. It was concluded that in-car observation can be a reliable and valid method for observing driver behaviour. Most of the variables studied had well-documented relevance to traffic safety.

The instrumentation of a vehicle can include data loggers. Sometimes the term "probe" is used for vehicles that are equipped with only a GPS recorder. This makes it relatively easy to equip more vehicles and therefore obtain more data. These types of simple installations can also be done in vehicles that are used on a day-to-day basis, which thus avoids making the drivers aware that they are being recorded. At the same time, the main purpose of the data logging, continual data collection, is achieved. Some examples of this approach might be black boxes that record the last 5 minutes of car handling before there is an accident. (Lotan and Toledo 2005) described a preliminary research project where the purpose was evaluating the safety implications and benefits of an in-vehicle data recorder for young drivers. To improve the quality of the experience for young drivers during their mandatory accompanied driving period, an in-vehicle data recorder collected information, including more than 20 different manoeuvre types in raw measurements, which could then be and used to indicate overall trip safety.

When the driver of the instrumented vehicle is the subject of the research, an instrumented vehicle with continuous data collection makes it easier to gather more data as well as more accurate data about driving performance. On the other hand, these data loggers can make it more difficult to conduct research without the driver being influenced by the research situation.

3.4.3 Instrumented vehicle at KTH

The instrumented vehicle at KTH[4] was one of the most important influences during the early design stages of the NTNU/SINTEF instrumented vehicle. Transportation and logistics at KTH in Sweden have an instrumented vehicle developed by Volvo Technology (VTEC). The vehicle is a Volvo V70 with a 2.4 litre engine and a manual gear shift.

In addition to standard instruments, the vehicle is equipped with GPS to record x,y, and z coordinates and GPS speed, radar to detect up to 4 adjacent vehicles, and video cameras to record driver behaviour and traffic in front of the vehicle. By logging the standard instruments in the car it is also possible to record the following data:

- Distance travelled;
- Speed in the travelling direction;
- Acceleration in the travelling direction;
- Steering wheel angle;
- Brake pedal pressure;

---

[4] KTH = Kungliga Tekniska högskolan

- Engine RPM;
- Use of lights (turn signal indicator, brake lights, low/high beams, parking lamps, backup lamp, etc.);
- Yaw rate; and,
- Lateral acceleration.

The logging program was developed by Volvo. The data collection can be conducted on an ordinary laptop with a refresh rate of 50 Hz. The program can play back both logged data and digital video.

3.4.4 Instrumented vehicle at NTNU/SINTEF

The instrumented vehicle at NTNU/SINTEF is a Volvo V70 2.4s. It is possible to use it in many different research projects. Two purposes that are of special importance are:

- The study of an individual driver's behaviour in traffic through studies of driver behaviour regarding road geometry (Traffic signs, road markings, road design) and other drivers (vehicles, pedestrians, bicyclists).
- Extend and validate the use of the driving simulator through parallel studies of driver behaviour in real world and in the simulator.

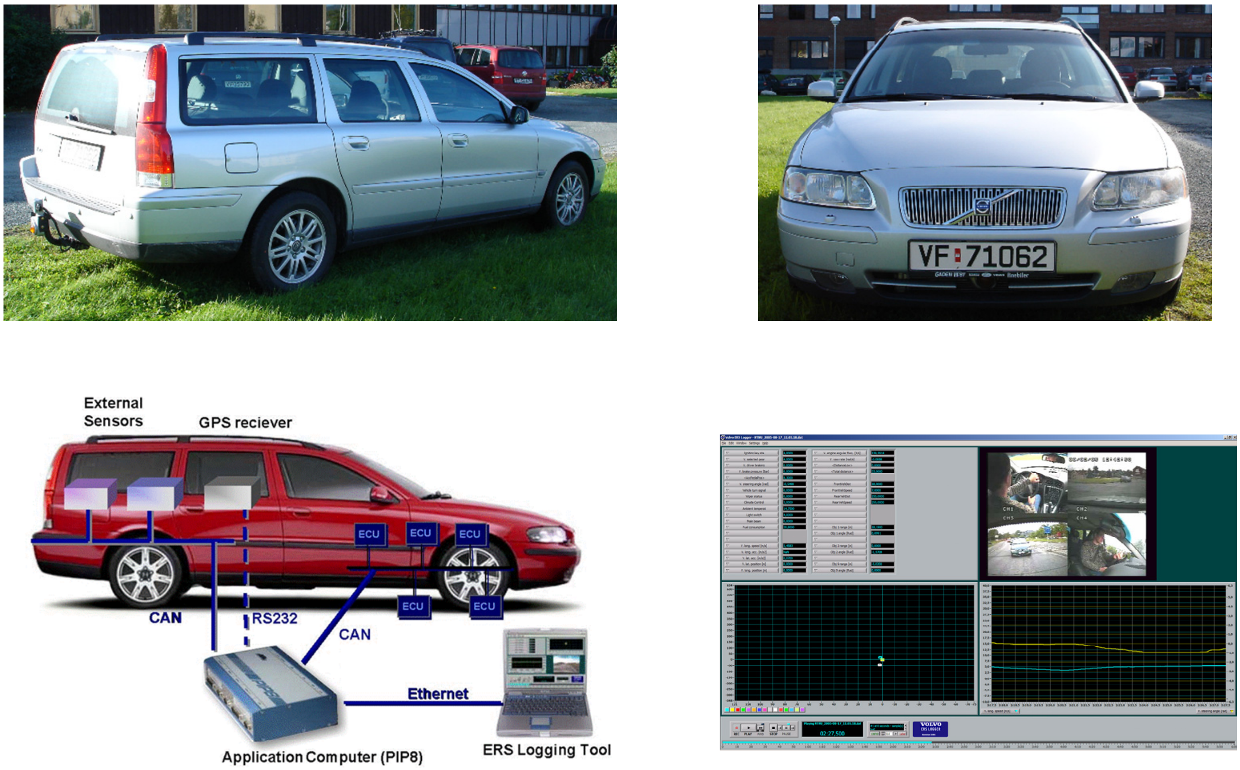


Fig. 31. An overview of the instrumented vehicle.

In choosing instruments for this vehicle, one important criterion has been the ability to compare results between the instrumented vehicle and the driving simulator.

Data sources in the car can be divided into two parts: the collection of information from standard sensors that are built into the car from the manufacturer, and the collection of data from extra sensors specially mounted on the instrumented car.

The basic setup makes it possible to record the following data from standard sensors in the car:

- Distance travelled
- Speed in the travelling direction;
- Acceleration in the travelling direction;
- Steering wheel angle;
- Wheel angle;
- Brake pedal pressure;
- Engine RPM;
- Use of lights (turn signal indicator, brake lights, low/high beam, parking lamps, backup lamps, etc.);
- Gear;
- Fuel consumption;
- Outside temperature;
- Windscreen wiper status (interval, speed, if possible automatic start);
- Climate control status;
- Ignition key position;
- Yaw rate;
- Lateral acceleration;
- Wheel locking; and,
- Wheel spinning.

The additional sensors that were installed at the time of delivery were:

- Video camera and video recording equipment to record the driver, the instruments in the vehicle, and the road ahead of the driver.
- Measurement of position with GPS equipment.
- Measurement of distance to adjacent vehicles by radar

Photos of radars and cameras are shown in Fig. 32-Fig. 35.

Fig. 32. Rear-facing radar.

Fig. 33. Front-facing radar.

Fig. 34. Rear-facing camera.

Fig. 35. Front-facing camera.

The logging of standard car sensors is made possible with the use of the standard vehicle network CAN-bus. In addition the radar sensors also send their signal over the CAN-bus. The GPS uses standard NMEA0182 signals and sends its information over a RS232 connection.

The CAN-bus signal and RS232 signal are all collected by a Volvo application PC (PIP8). The application PC sends the data over a standard 100 mbit network cable to a logging PC running an ERS logging tool.

The video signal from the four cameras are sent directly to the logging PC. The layout of the system is shown in Fig. 36 and Fig. 37.

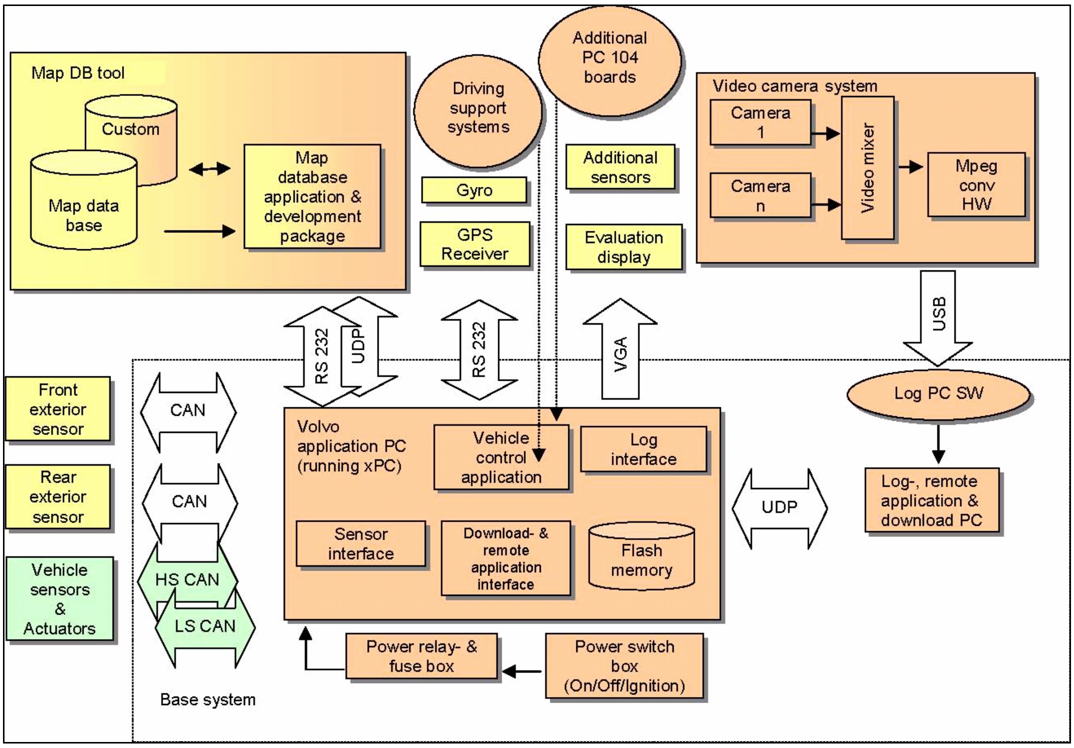


Fig. 36. The instrumented vehicle system description.

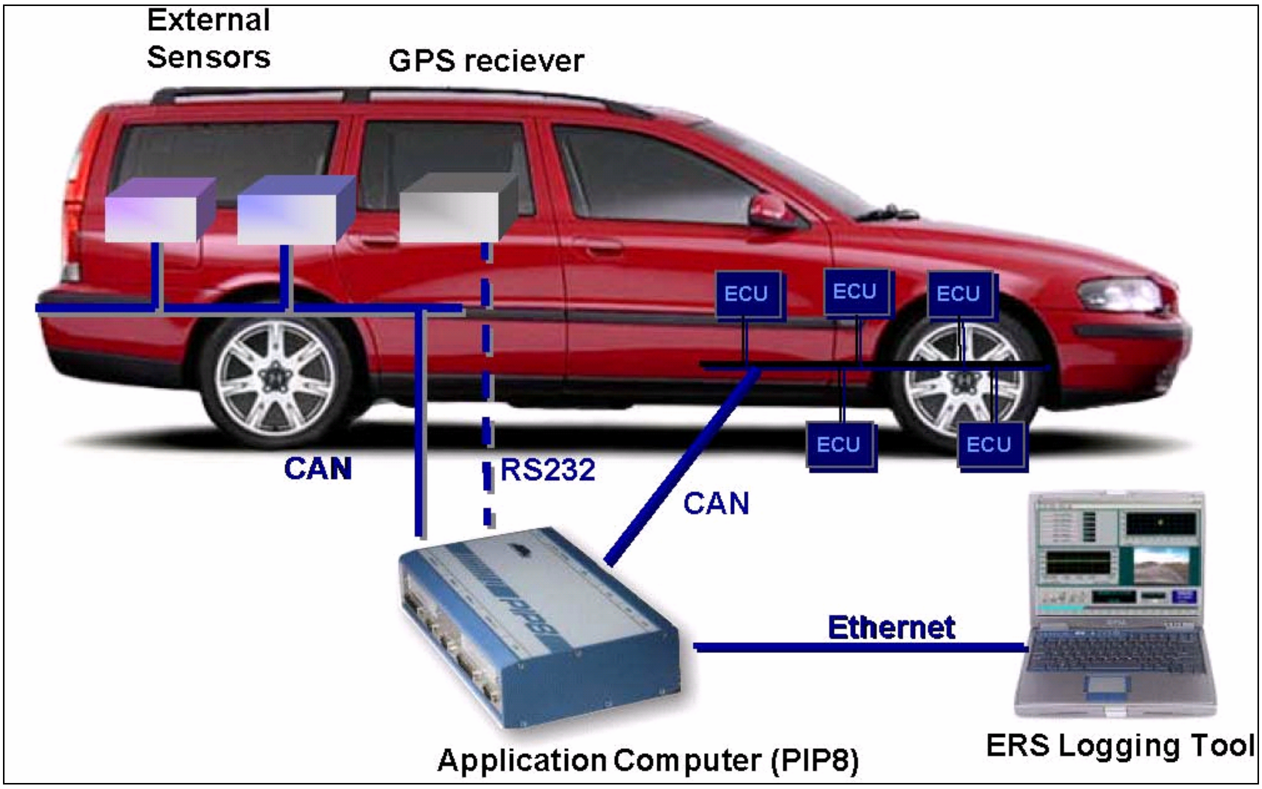
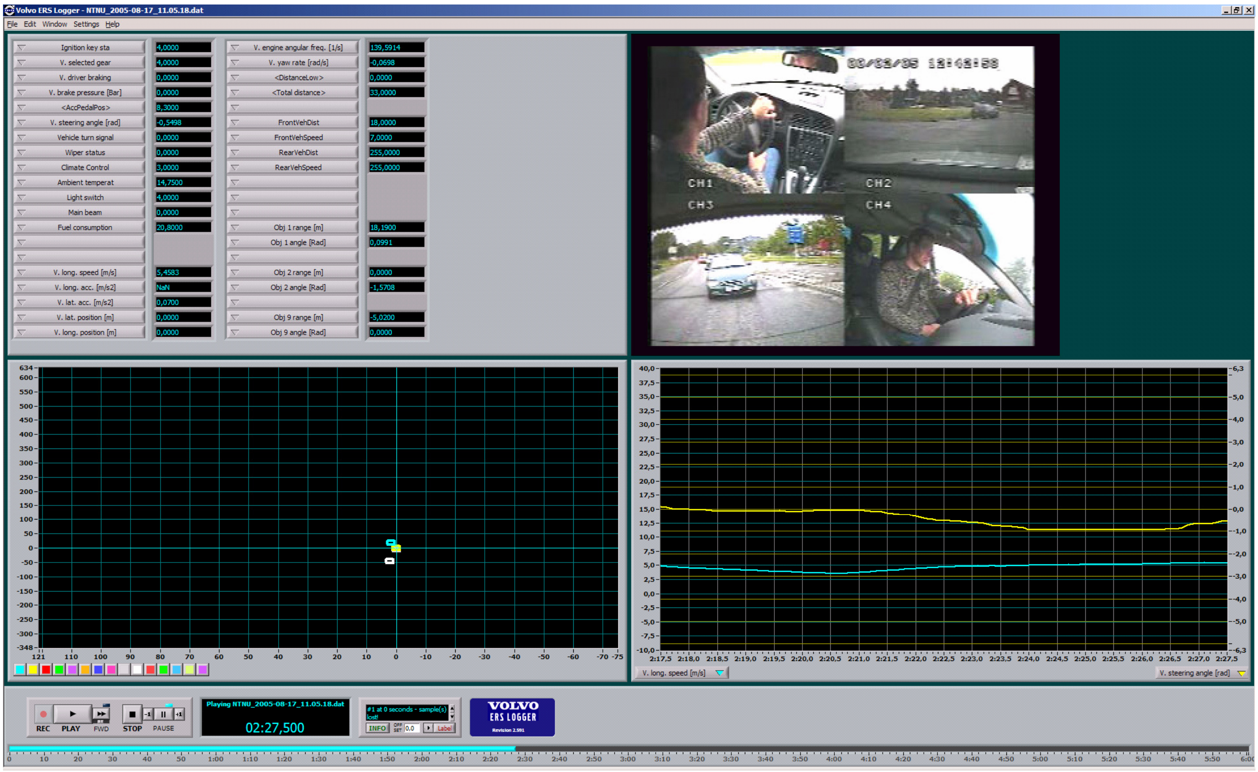


Fig. 37. The instrumented vehicle system.

The ERS logging tool can be used for data collection, data analysis and data conversion. When it is used for logging it has to be connected to the application PC using a network cable and to the video system with a USB cable.


Volvo ERS Logger - NTNU_2005-08-17_11.05.18.dat

Ignition key sta	4,0000	V. engine angular freq. [1/s]	139,5914
V. selected gear	4,0000	V. yaw rate [rad/s]	-0,0698
V. driver braking	0,0000	<DistanceLow>	0,0000
V. brake pressure [Bar]	0,0000	<Total distance>	33,0000
<AccPedalPos>	8,3000		
V. steering angle [rad]	-0,5498	FrontVehDist	18,0000
Vehicle turn signal	0,0000	FrontVehSpeed	7,0000
Wiper status	0,0000	RearVehDist	255,0000
Climate Control	3,0000	RearVehSpeed	255,0000
Ambient temperat	14,7500		
Light switch	4,0000		
Main beam	0,0000		
Fuel consumption	20,8000	Obj 1 range [m]	18,1900
		Obj 1 angle [Rad]	0,0991
V. long. speed [m/s]	5,4583	Obj 2 range [m]	0,0000
V. long. acc. [m/s2]	NaN	Obj 2 angle [Rad]	-1,5708
V. lat. acc. [m/s2]	0,0700		
V. lat. position [m]	0,0000	Obj 9 range [m]	-5,0200
V. long. position [m]	0,0000	Obj 9 angle [Rad]	0,0000

Playing NTNU_2005-08-17_11.05.18.dat
02:27,500

#1 at 0 seconds - sample(s) lost!

Fig. 38. ERS logging tool.

When conducting an analysis or data export, the logging PC does not have to be connected to the application PC and video system. This means that after data collection in the field, the researcher can take the logging PC to the office and conduct the data analysis and export.

3.5 Advantages and disadvantages of traffic data collection tools

All the traffic data collection tools have strengths and weaknesses. Different factors will influence the validity of the research depending on the traffic data collection method.

3.5.1 Advantages and disadvantages of a driving simulator

The use of driving simulators has several advantages compared to field research:

- It is possible to make detailed measurements of the driver's behaviour.
- It is possible to make measurements under controlled experimental settings.
- It is safe to undertake data collection in the simulator.
- Data recording can often be cheaper.
- It is possible to recreate traffic situations.

A driving simulator makes it possible to record much detailed information about driving performance. The information can be specially tailored to the specific research problem.

Controlled experiments are also possible in a driving simulator. Another advantage is the ability to conduct randomized experiments where drivers are randomly assigned different driving conditions.

The driving simulator also makes it possible to do research that is too dangerous undertake in real life. This may be research involving drug use, or research designs to determine the maximum stress that is imposed on a driver by travel information systems, as one example. It is also possible to conduct data collection that demands that the test driver violate traffic regulations.

The driving simulator also makes it possible to test traffic solutions without having to implement them in real life. This ability may be used in the study of different design solutions for a new road.

The driving simulator enables certain situations to be created that may occur infrequently in real traffic. These situations may be repeated and presented to all or specific drivers.

The driving simulator has some important disadvantages. It is probably impossible to create a natural situation surrounding the research. The driver will be aware he/she is being studied, and this will most likely influence performance.

It is also difficult to know how realistic the driver's experience is in the driving simulator as compared to the real world. The driver will be influenced by different sense impressions such as sight, sound, movement, and perhaps even smell. All of this may be difficult to create realistically in the driving simulator, but this is something that is constantly being improved.

Another disadvantage is that it is difficult to make large quantities of measurements due to cost. Making measurements of individual drivers in the driving simulator is relatively time consuming compared to making measurements using roadside equipment.

3.5.2 Advantages and disadvantages of an instrumented vehicle

Instrumented vehicles have different types of advantages and disadvantages, depending upon the aim of the research.

Research conducted on the driver of the vehicle has the advantage of allowing for a great deal of detailed data to can be, recorded depending on how the vehicle is instrumented. On the other hand, it is more difficult to collect a large number of measurements. The greatest advantage compared to a driving simulator is that the research can be conducted in real traffic. At the same time, the research is being conducted on a driver that is aware he/she is being studied, and thus will be influenced by the situation.

When several instrumented vehicles are used as probes in real traffic, the number of measurements increases. The effect of the driver's knowing about the research is also reduced if it is a vehicle in ordinary use that is utilized. The amount of information is often reduced because the vehicle has simpler instrumentation. In certain circumstances this can be an advantage since it reduces the need for protection of privacy.

Research conducted on drivers who are not aware that they are being studied has the advantage that the results will not be influenced by the driver adjusting to the research situation. This research is primarily conducted using rear-facing radar to study time gaps or speed adoption. Front-facing radar can be used to study speed choice made by other vehicles. The number of measurements may be limited, but this approach means there is no need to organize test drivers.

3.5.3 Advantages and disadvantages of roadside equipment

Roadside equipment is relatively cheap and easy to use. It usually results in a large number of measurements, making it possible to find even small statistical correlations.

The disadvantage is that it is difficult to conduct in-depth measurements of single drivers and vehicles. It is difficult to control all the effects that contribute to possible changes in data.

3.6 Summary

An alternative to driving simulator data collection is to use either an instrumented vehicle or roadside data measurement equipment. Both have their advantages and disadvantages. Roadside data measurement equipment can quantify information from a great number of vehicles, but the measurements are usually related to one point. Instrumented vehicles can conduct measurements over a road stretch, but the number of vehicles assessed is usually limited.

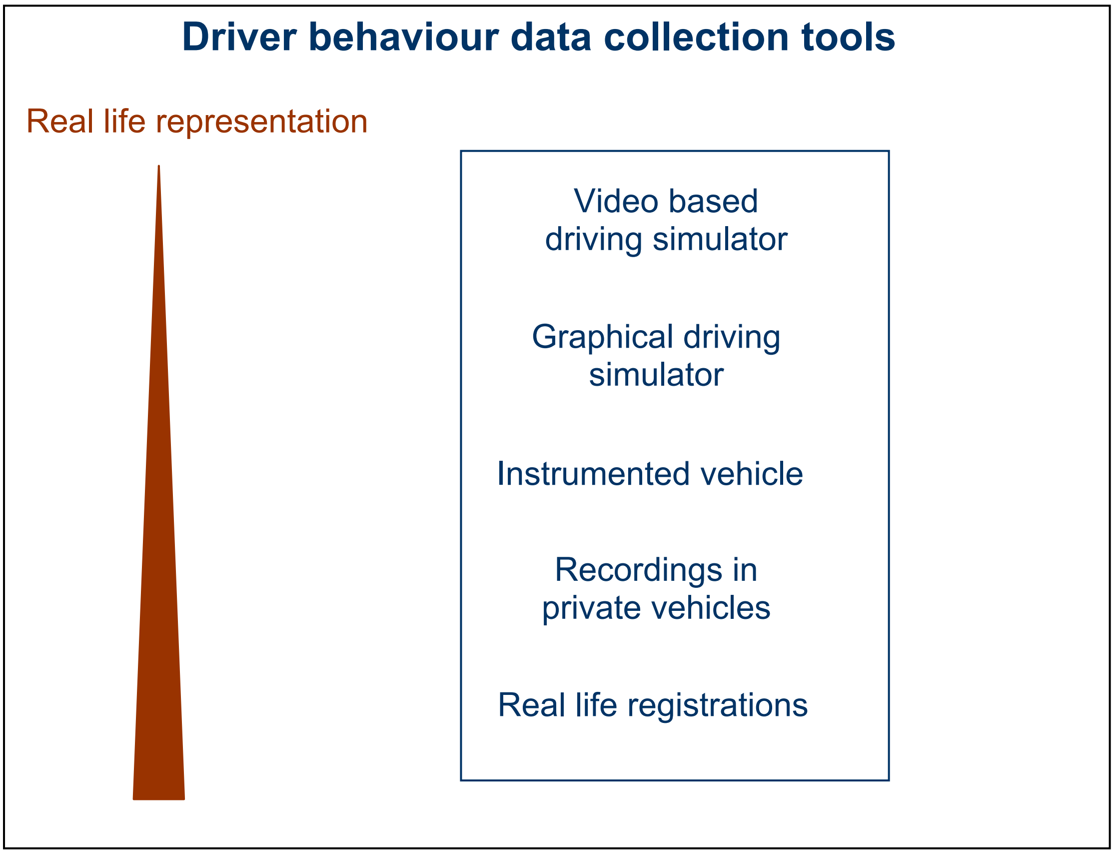


Fig. 39. Driver behaviour data collection tools research.

In the figure above the main tools for behavioural research in road traffic is summarized.

4. Methods for driving simulator validation

4.1 General description of Validation

4.1.1 Introduction

As we have seen in previous sections, the use of a driving simulator in many cases has advantages in comparison with both roadside measurements and moving vehicle measurements. The choice of a driving simulator to conduct research should hopefully provide results that are an accurate and correct representation of how drivers react in real traffic. To obtain this information, research conducted in a driving simulator must be validated.

In the Oxford dictionary of English(Soanes and Stevenson 2003) , the definition of validation is:

> **Validation**: “check or prove the validity or accuracy of: all analytical methods should be validated in respect of accuracy. demonstrate or support the truth or value of: acclaim was seen as a means of validating one's existence. make or declare legally valid.”

(Shadish et al. 2001) describe the term validity as “refer[ing] to the approximate truth of an inference”. They argue that the term knowledge claim can be used in place of inference, and treat the terms interchangeably.

Evidence of validity may come from other sources of information, such as from past findings and theories. Validity judgement cannot be absolute. (Shadish et al. 2001) make the following argument for this:

- We can never be certain that all of the many inferences drawn from a single experiment are true.
- We cannot be certain that other inferences have been conclusively falsified.

4.1.2 Validation in computer science and simulation

The term validation is often used in conjunction with the term verification. In computer science the terms verification and validation (V&V) are often used. Sometimes the term verification, validation and evaluation (VV&E) are used. The terms are defined in (Wentworth et al. 1995) as: “Verification of an expert system is the task of determining that the system is built according to its specifications. Validation is the process of determining that the system actually fulfils the purpose for which it was intended. Evaluation reflects the acceptance of the system by the end users and its performance in the field.”

The terms verification and validation are also often used by the military with the term accreditation in regards to models and simulations (M&S). In the U.S. (Department Of The Army 2000), the definition of Verification, Validation & Accreditation (VV&A) is: “Verification is the process of determining if the M&S accurately represent the developer’s conceptual description and specifications and meets the needs stated in the requirements document. Validation is the process of determining the extent to which the M&S adequately represents the real-world from the perspective of its intended use. This process ranges from single modules to the entire system. Accreditation is an official determination that the M&S are acceptable for its intended purpose."

4.1.3 Validation in behavioural research

Validity is often divided into groups depending on how the accuracy of the information is represented. Cozby (2004) describes three types of validity: construct validity, internal validity, and external validity.

(Shadish et al. 2001) also include the term statistical conclusion validity. In my thesis, I use the following definitions of the types of validity, partly based on Cozby (2004) and (Shadish et al. 2001):

- Statistical conclusion validity: The degree to which the use of the appropriate statistics used to conclude whether the presumed independent and dependant variables covary. The validity of inferences about the correlation (co-variation) between treatment and outcome.
- Internal validity: The degree to which the result of an experiment can be attributed to the manipulation of the independent variable rather than to some other, uncontrolled variable.
- Construct validity: The certainty to which a measurement device accurately measures the theoretical construct it is designed to measure.
- External validity: The degree to which the experiment's results can be generalized to different persons, setting, treatment variables and measurement variables.

I will describe this model further in Section 4.5 .

4.2 Current validation of driving simulators

It is not possible to undertake the complete validation of a driving simulator once and for all. According to (Kaptein et al. 1996), it is meaningless to refer to the validity of a research instrument as such. Validity is only defined relative to a specific research question.

(Godley 1999) differentiates between two levels of validity:

- Behavioural validity or predictive validity, which concerns the correspondence between the simulator and the real world, in the way the human operator behaves.
- Physical validity, which concerns the physical correspondence between the simulator and its real world counterpart.

4.2.1 Behavioural validation

A fundamental question for this thesis is data collection about behaviour in traffic. Whether a simulator is a valid instrument for data collection for a specific driving task depends primarily on the information used to perform the driving task. How accurately this information is presented in a driving simulator may differ between driving tasks.

Although physical validity is an influence, ultimately it is the behavioural validity that is important in the collection of traffic data. Physical validity may, on the other hand, help strengthen behavioural validity. Behavioural validity may have different levels of accuracy. According to (Kaptein et al. 1996), validity can be divided into separate parts:

- Absolute and relative validity – The driving simulator has absolute validity with regard to the research question if the absolute size of the effect is comparable to the absolute effect in reality. The driving simulator has relative validity with regard to the question if the difference is in the same direction and relative size of the effect of the measure as in reality.

- Internal and external validity – Internal validity refers to the relationship between the manipulation and the obtained effect with no alternative explanation of the effect. External validity refers to the extent to which the results can be generalized to other situations.

- Validation of realism – This refers to what extent the research situation appears realistic to the driver. The realism can, for example, influence the motivation of the driver, and therefore the results.

- Statistical validity – Statistical validity refers to statistical tests that can be conducted on the data.

An in-depth look at behavioural validation is undertaken in Section 4.3 , where I have divided the validation research into six parts:

- Direct comparison with real life data;
- A comparison of the driving simulator with physiological tests and a questionnaire;
- Expert testing;
- Validation compared to specific driver characteristics;
- Stability over time and driver characteristics; and
- Driving training.

4.2.2 Physical validation

Although physical validation is not the focus of this thesis, it is important to undertake this kind of validation. This is especially true when the focus of the research is to examine vehicle handling. Examples of physical validation of a driving simulator can be found in (Chrstos and Grygier 1997) and (Salaani and Heydinger 2000), which involve validation of the physical models for a 1994 Ford Taurus and 1997 Jeep Cherokee. The vehicles models are used in the National Advanced Driving Simulator presented in Section 2.1.2 .

The validation of the 1994 Ford Taurus involved measurements of the steering, brake, and throttle controls. A series of test manoeuvres was conducted on three different test tracks to evaluate the physical model. When possible, each test was run 10 times. It was noted that at the time of the study that there was very little publicly available test data on instrumented vehicle handling and braking. Therefore, the paper presented in-depth results from the measurements to make the test data available to other researchers.

The 1997 Jeep Cherokee vehicle handling and dynamics were evaluated and results were compared with experimental field testing. The Jeep evaluation covered vehicle directional dynamics that included steady state, transient and frequency responses, and vehicle longitudinal dynamics that included acceleration and braking.

A driving simulator will always be a mathematical and artificial model of the real world, and it will always be possible to improve it. Specific parts of a vehicle model may also be tested. An example of this can be found in (Nordmark and Åström 2001), where they validated a new physical model for tire friction in the VTI driving simulator. Although they only studied tire friction, the amount of testing was extensive since the model was intended to be valid for a broad range of surfaces, from ice to dry asphalt. The researchers found very good correspondence between the validation of the new models compared to field experiments in open loop conditions. At the same time, validation of the whole simulator was also performed, with drivers making the same slalom manoeuvre between cones in the simulator as well as on a real track. No significant differences between the driving behaviour were found in the objective measures, but the subjective evaluations showed a more mixed picture.

4.3 Literature review of behavioural validation of driving simulators

There have been many research projects undertaken were one important part has been to validate the results from the driving simulators, particularly in recent years. Not all of the cited research has been conducted with the stated purpose of validation; they are included here because they have tried to validate the results.

(Hoskins and El-Gindy 2006) undertook a literature review of driving simulator validation studies, but the surveys reviewed were all conducted between 1987 and 2000. Thus, no validation research within the last six years was presented. The researchers concluded that difference in driver behaviour in simulated and real environments can be a useful method for validating driving simulators, but care must be taken to avoid generalizing the results to other driving characteristics or other subsets of the populations. As stated earlier, I have found a number of studies conducted after 2000 that address driver simulator validation. These are presented here, because I feel they support the validity of using a driving simulator in the context of the research project.

4.3.1 Direct comparison with real life data

Most of the studies that have been conducted have used the mean speed as the one and only parameter in validating the driving simulator. Some studies have employed additional parameters. There has been substantial growth in the amount of research that provides a direct comparison of reiterated data between real life and a driving simulator.

Driving simulator validation for speed research

(Godley et al. 2002) conducted a study of the behavioural validity of the Monash University Accident Research Centre driving simulator. This is a mid level driving simulator. Twenty-four drivers drove the instrumented vehicle and 20 drivers drove using the driving simulator. The participants drove on three roads containing transverse rumble stripes. The three pairs of sites were approaches to stop sign intersections, right curves, and left curves; all involved deceleration.

The drivers reacted to the rumble strips in a similar ways in both the instrumented vehicle and in the driving simulator. However, the drivers generally drove faster in the instrumented vehicle. This is an example of relative validity.

Driver performance in the EPSRC Driving Simulator (LADS)

(Carsten et al. 1997) compared speed and lateral position between the driving simulator and the road. Video cameras were mounted at 21 different locations along an 8 km road section. Each location had two cameras, one for speed and one for lane position measurement. Data was collected for 100 vehicles passing through in free-flow conditions (greater than 7 s headway). The same stretch of road layout was recreated in the simulator.

Independent t-tests were performed between mean speeds across the various locations on the real road and the mean speed of 100 drivers (50 male, 50 female) in the simulator. The mean speeds did not differ reliably. The researchers claimed their findings confirmed a high degree of absolute correspondence between simulator and real-life driving speeds.

Similar tests were performed between mean lateral positions. There was no absolute relationship between the two, but relatively the graphs were the same shape, but offset by, on average, 0.4 m. Further validation studies performed after upgrading the field of view visible within the LADS improved this discrepancy, but an absolute validation remains to be achieved.

In another study in the LADS, (Jamson 1999) looked at a claim that novice drivers are unable to negotiate curves as accurately as expert drivers. The results in the driving simulator were similar to on-road studies, in that novices were significantly more likely to weave in lane, to wander out of lane more often, and to have shorter time to line crossings. Jamson stated that these findings provided a useful behavioural validation for the Leeds Driving Simulator.

Driver reaction time in crash avoidance research: Validation of a driving simulator study on a test track

(McGehee et al. 2000) conducted a series of experiments in the Iowa Driving Simulator and on a test track. These trials, which were designed to be analogous, were conducted to examine driver reaction time. The study in the driving simulator was conducted with 60 females and 60 males aged 25 to 55. The drive lasted for approximately 15 minutes and ended with an incident at an intersection.

The study on the test track was conducted with 192 test drivers aged between 25 and 55. The drive took approximately 15 minutes and required the driver to complete 3.5 laps on the test track. On the last lap, a full-size foam core photograph mock-up of a car was propelled 6 feet into the subject’s lane.

A comparison of the total brake reaction time between the studies was equal at the 95th percentile. The same was true with respect to time to initial steering. Time to throttle release was different, but this was explained by methodological differences.

Validation of a Driving Simulator for Work Zone Design

(Bella 2005) conducted a research project aimed at calibrating and validating the interactive fixed-base driving simulator owned by CRISS (Inter-University Research Center for Road Safety, a joint venture between three leading Italian universities) to enable its use for design and verification of the effectiveness of temporary traffic signs on highways. A survey of speed measurements on highways next to a work zone of medium duration was conducted. The same situation was reconstructed in virtual reality using the driving simulator. Speed measurements were conducted with a laser speed meter in the transition area, the activity area, and the termination area, and speeds were shot in the advance warning area with a camera from an overpass. Speed data from the field and the simulator were analysed to determine whether drivers responded differently in the simulator compared with their response during the real driving experience. The speeds observed in the real situation and those measured with the simulator did not differ statistically.

Effects of cognitive and visual load in real and simulated driving

(Östlund et al. 2006) conducted a driving simulator study and a field experiment to study the effects of visual and cognitive loads on driving performance. One of the study's objectives was to validate the use of a driving simulator as a tool to study the effects of distractions. The field experiments were conducted in an instrumented Volvo S80 and conducted on a highway. Twenty-four individuals participated in the field experiments. The driving simulator studies were conducted in the VTI Driving simulator II, the predecessor to the current driving simulator described in Section 2.1.2 . Forty-eight individuals participated in the driving simulator study. The visual load was imposed by using a visual detection task and the cognitive load was imposed by an auditory memory task. The driving simulator validity was found to be "very high" and the secondary tasks affected the performance very similarly in the driving simulator and on the real world. An exception was that the lower relative risk in the driving simulator seemed to result in lower stress levels and higher travel speeds.

Driving behaviour in a real and a simulated road tunnel - A validation study

(Tornros 1998) conducted a study to validate driving behaviour in a simulated road tunnel. The speed and lateral position of 20 subjects were measured in a real tunnel and in the same tunnel implemented in the VTI driving simulator. In both situations a left-hand steered and manually geared passenger car was used. Driving speed was higher in the simulated tunnel than in the real tunnel. Elimination of speed information from the speedometer caused a similar small speed increase in both situations. Additionally, the difference in speed between driving lanes was similar in both cases. The effects on speed variation were similar to that found for speed level. Regarding lateral position, the subjects positioned the car somewhat further away from the nearest tunnel wall in the real tunnel than in the simulated tunnel. In both situations, the distance to the nearest wall was greater when it was located to the left of the driver than on the opposite side. Lateral position deviation was about the same when the road was straight, but in a curved section it was somewhat greater for the simulated tunnel. It was concluded that behavioural validity in absolute terms was not quite satisfactory, especially regarding choice of speed, whereas relative validity was good for both speed and lateral position.

The validity of last-second braking and steering judgments in advanced driving simulators

In research reported in (Greenberg et al. 2006), timing judgements made under a variety of last-second braking and steering conditions were compared across three testing environments: closed course testing in actual vehicles, simulation testing at the University of Iowa’s National Advanced Driving Simulator, and simulation testing at Ford Motor Company’s VIRTTEX simulator. The kinematics conditions fell into three classes: deceleration events, constant speed events and stationary events. Seventy-two test participants drove the steering and braking scenarios in each test environment. The subject pool was different for each test environment, although demographically similar. During the braking trials, subjects were instructed to brake to avoid colliding with a lead vehicle. In the steering trials, they were instructed to steer around a lead vehicle to avoid a crash. The driver's judgements about when to steer or brake were characterized by the time-to-collision calculated at the moment of decision. The deceleration trials showed, in general, the best agreement between the simulators and the closed course testing. Constant speed and stationary trials agreed less well. Although minor differences between the two advanced simulators were observed, the overall pattern of correspondence with the closed course data was similar.

Road-to-lab: validation of the static load test for predicting on-road driving performance while using advanced in-vehicle information and communication devices

(Young et al. 2005) evaluated the Static Load Test for its ability to predict on-road driver performance while using in-vehicle devices. In this test, participants performed various in-vehicle tasks in a lab while viewing a videotaped road scene on a monitor, tapping a brake pedal when a central or peripheral light was observed. This would not be considered a driving simulator today, but the vehicles still presented similar handling characteristics as found in a simple driving simulator. The purpose of the validation study was not to determine if the situation in the Static Load Test could be compared to real world driving, but to see if a model could be made that described the real world based on the test.

For the on-road comparison test, the device, tasks, and lights were the same, as were the participants who drove the vehicle while performing the tasks and responding to the lights. In both the lab and road tests, ten driver performance variables were measured. The goal was to produce a linear model to predict an on-road variable from the lab data with low residual error, high percent variance explained, and few errors in classifying tasks as meeting or not meeting on-road driver performance criteria. Separate test data from a replicated Static Load Test at an independent lab were used to further validate the models. The researchers claimed that the results indicated that a simple, inexpensive, and low-fidelity Static Load Test can accurately predict a number of on-road driver performance variables suitable for assessing the safety and ease-of-use of advanced in-vehicle devices while driving.

Comparative analyses of driver behaviour on the track and in a virtual environment

(Guzek and Jurecki 2006) present examples of tests on driver behaviour in certain prearranged pre-accident situations, consisting mainly of the sudden occurrence of an obstacle in front of the driver. The tests were carried out under the conditions that would be experiences during an experiment on the road, which was the Kielce Track, as well as in the autoPW driving simulator at the Warsaw University of Technology. The driver's assignment was to drive along the indicated route at a predetermined speed. At a predetermined distance from the cross road, a vehicle emerged from the right side and travelled across the cross road, stopping after it reached a specified point. The tests were carried out for the following parameters: driving speed: 40, 50, 60 km/h; and distances from the vehicle at which the driver would notice the obstacle: 10; 20; 30; 40 and 50m. In order to avoid the drivers adopting routine behaviours during successive attempts, the parameters were changed at random and in a way unknown to the driver. The tests also allowed for "free passages", or test periods during which no obstacle was introduced. The tests were conducted for a group of 30 drivers. The recorded parameters during the tests were: reaction times to turn, braking with the service brake and release of the accelerator pedal, which were measured as the time from the moment when the obstacle became visible until the time when the driver started a given manoeuvre. The researchers concluded that that there was a linear dependence between reaction time measured in the driving simulator and the real world.

4.3.2 Comparison of driving simulator with physiological tests and questionnaires

Often the use of driving simulator is just one of several options for testing drivers. This sub-chapter reviews research where the use of a driving simulator was compared to administering either physiological tests or questionnaires to the driver. The use of alterative measurement methods provides good insights into the pros and cons of using a driving simulator

Cognitive abilities related to driving performance in a simulator and crashing on the road

(Anderson et al. 2005) studied the cognitive abilities related to driving performance in a simulator and in road crashes. The purpose was to examine the relationships between performance on standardized neuropsychological measures of cognitive abilities, simulated driving performance, and state crash records involving drivers with cognitive decline due to ageing and dementia. Participants were 202 experienced older adult drivers age 55 years and older: 70 had mild dementia due to probable early Alzheimer's disease and 132 had no neurological disease. All completed neuropsychological tests and drove contemporaneously in a driving simulator. The participants' State Department of Transportation driving records were monitored for up to two years after testing. The simulator composite score, reflecting overall driving ability, was significantly correlated with overall cognitive ability. Drivers who crashed during an intersection incursion scenario performed significantly worse on the composite measure of cognitive function than those who successfully steered around the incurring vehicle. Crashers had specific cognitive deficits on measures of visuomotor abilities and attention. Memory test performances for both verbal information and visual material were associated with subsequent on-road crashes. They claim that these findings provide support for the validity of driving simulation as a safe means of evaluating a range of driving responses that cannot be tested on the road, and suggest that relatively simple and inexpensive neuropsychological tests of specific cognitive abilities could be used to help evaluate older drivers' risk of unsafe driving.

Using self-reported data to assess the validity of driving simulation data

(Reimer et al. 2006) used self-reported driving behaviours from a written questionnaire to assess the measurement validity of data derived from a driving simulation. The issue of validity concerned the extent to which measures from the experimental context map onto constructs of interest. Following a description of the experimental methods and setting, an argument for the face validity of the data was presented. Convergent validity was assessed by regressing behaviours observed in the driving simulator on self-reported measures of driving behaviours. Significant relationships were found across six measures: accidents, speeding, velocity, passing, weaving between traffic, and behaviour at stop signs. The researchers concluded that although the relationship between self-reported behaviours and observed responses in the simulator fell short of perfect correspondence, the data collected from the driving simulator were valid measures of the behaviours of interest.

4.3.3 Expert testing

Section 2.5 introduced the concept of using expert evaluation of planned roads. In a similar way, the use of a driving simulator can be validated through the use of expert testing. This is similar to using questionnaires as described in the previous section, but also introduces the advantage of an expert into the testing scenario.

Validation process of the ULTIMATE driving simulator

(Dagdelen et al. 2006) describes the ULTIMA simulator at Renault. The ULTIMA simulator is a high-performance dynamic simulator featuring a large X-Y table and a hexapod motion system, as well as a high-frequency motion seat, designed for vehicle dynamics engineering applications. The vehicle dynamics and the different feedback systems (cabin motion, steering wheel, pedals and sound) were calibrated. The research was conducted in 2004, and took into account intrinsic technical limitations of driving simulation (real-time scheduling, transport delays, actuator linearity, actuator limits, etc.) as well as motion cueing and perception constraints. The second step was running expert Renault drivers through the simulator for a subjective comparison with the actual vehicle reference (Laguna). Driving scenarios were exact replicas of the different official Renault test tracks. Evaluation was based on a series of standardized tests, which were limited here to standard driving, excluding extreme manoeuvres. During each test session, the drivers' inputs and the vehicle trajectory were recorded. After each test session, the drivers filled out an evaluation questionnaire and rated several vehicle dynamics criteria, such as lateral/longitudinal acceleration levels, roll/yaw coordination in turns, pitch during acceleration/braking, and steering stability, among others. They claim that the level of fidelity attained already makes the simulator a good candidate for advanced driver behaviour studies, in particular with regards to electronic driver aid systems.

Validation of Renault's dynamic simulator for Adaptive Cruise Control experiments

An earlier driving simulator at Renault addressed the development of ACC (Adaptive Cruise Control) algorithms in a study by (Reymond et al. 2000). The driving simulator can reproduce paradigm scenarios, such as approaching, following, and overtaking a lead vehicle on a highway with a high degree of realism, reproducibility and measurability. Different ACC function parameters were evaluated by standard drivers under these conditions. The researchers stated that in the Renault motion-based driving simulator, drivers could perceive the acceleration onsets caused by activation/deactivation of the ACC, and could therefore better assess the subjective driving comfort of the system than in a fixed-based simulator. The paper discussed requirements for the motion rendering in terms of physical validity (reproduction of simulated vehicle accelerations) and perceptual validity (realism of perceived car motion, minimization of sensory conflicts). Expert drivers experimented with the behaviour of the simulated vehicle during highway driving tasks, such as lane following and overtaking. The non-linear motion cueing filter, after integration in the simulator, produced a driving sensation that was deemed superior by test drivers to the initial configuration based on a classical filter.

4.3.4 Validation compared to specific driver characteristics

As I have previously described, a driving simulator must be validated relative to the research that will be conducted. Often the purpose is so specialized that special considerations have to be taken into consideration.

Comparative sensitivity of a simulated driving task during prolonged wakefulness

The objectives of (Arnedt et al. 2005) were to compare the sensitivity of simulated driving in the York Driving Simulator to self-report measures, nocturnal sleep latency tests (SLTs), and an auditory vigilance task and urban and motorway driving. Healthy males 18 to 35 years maintained wakefulness for one night and were tested at 2400, 0230, 0500 and 0730 h.

The primary objective of the first study was to evaluate the sensitivity of a face valid measure of driving ability relative to measures that are widely used and known to be sensitive to the effects of sleep loss. The second study was conducted to clarify an important confounding factor in the first study, as well as to examine more closely the degree to which task characteristics (in particular, monotony) affected simulated driving performance in a manner consistent with other performance measures.

The researchers stated that the findings from both investigations suggested that the York Driving Simulator was sensitive to the decrements in performance evident with increasing time awake, and that it appeared to have sensitivity under these conditions that was at least comparable to an auditory vigilance task. Moreover, increasing the monotony of the task produced an even more marked deterioration in driving performance. The researchers found that the simulated driving task offered advantages over self-report, physiological, and other performance measures in the assessment of driving performance, such as providing a more face valid measure of driving and using a highly practised skill. The performance impairments in the York Driving Simulator following prolonged wakefulness were equivalent to the decrements produced by levels of alcohol intoxication at which it is illegal to operate a motor vehicle. The findings suggested that this type of driving task may have clinical utility in the assessment of driving ability in patients who are at an increased risk for sleepiness-related accidents.

Validation study of the use of a driving simulator to measure road driving performance of older drivers

(Lee 2002) used a low level driving simulator called STISIM to study the driving performance of elderly drivers. Older drivers from a number of residential communities took part in the research. A “simulated driving index” and a “road assessment index” were calculated to compare real world driving and driving in the simulator. A higher “simulated driving index” or “road assessment index” score indicated better overall driving performance in the corresponding setting.

In this study, a positive correlation was found between the “simulated driving index” and “road assessment index”.

Validation of a driving simulator by measuring the visual attention skill of older adult drivers

The purpose of (Lee et al. 2003) was to validate a laboratory-based driving simulator as an off-road screening tool for older adult drivers, by measuring their visual attention skill, and to determine how the visual attention skill changed across time in a 45-minute simulated driving test. One hundred and twenty-nine older drivers from residential communities took part in the study. A range of driving scenarios was devised and implemented in a simulator setting to assess the driving skills of the participants. Visual attention skill was assessed by the participant's reaction times to a sequence of 14 visual stimuli during the primary task of sustained driving. The visual attention skill of older drivers was found to decline with age, whereas the effect of gender was not significant. Participants increased their speed of reaction for the first half of the test, but then slowed down during the second half. That visual attention skill declined with age was consistent with the literature, and validated the driving simulator as an effective screening tool for older adult drivers. The researchers concluded that with rapid advancements in computer technology, the driving simulator will likely play an important role in assisting occupational therapists with off-road assessment of older drivers.

Predictive validity of driving-simulator assessments following traumatic brain injury: a preliminary study

(Lew et al. 2005) evaluated whether a driving simulator and road test evaluations can predict long-term driving performance. The researchers conducted a study with 11 patients with moderate to severe traumatic brain injury. Sixteen healthy subjects were also tested to provide normative values on the simulator at baseline. At their initial evaluation (time-1), the subjects' driving skills were measured during a 30-minute simulator trial using an automated 12-measure Simulator Performance Index (SPI), while a trained observer also rated their performance using a Driving Performance Inventory (DPI). In addition, patients were evaluated on the road by a certified driving evaluator. Ten months later (time-2), family members observed patients driving for at least 3 hours over 4 weeks and rated their driving performance using the DPI. At time- 1, patients were significantly impaired on automated SPI measures of driving skill, including: speed and steering control, accidents, and vigilance to a divided-attention task. These simulator indices significantly predicted the following aspects of observed driving performance at time-2: handling of automobile controls, regulation of vehicle speed and direction, higher-order judgment and self-control, as well as a trend-level association with car accidents. Automated measures of simulator skill (SPI) were more sensitive and accurate than observational measures of simulator skill (DPI) in predicting actual driving performance. The road test results at time-1 showed no significant relation to driving performance at time-2. It was concluded that simulator-based assessment of patients with brain injuries can provide ecologically valid measures that, in some cases, may be more sensitive than a traditional road test as predictors of long- term driving performance in the community.

Impact of internal versus external cueing on driving performance in people with Parkinson's disease

(Stolwyk et al. 2005) examined the impact of impaired internal cueing on specific driving behaviours. A simulator measured the driving behaviour of 18 current drivers in the mild-to-moderate stages of PD (Parkinson's disease) and 18 matched controls. Participants navigated through different driving conditions where the opportunity to use internal and external cues was manipulated. Internally guided driving behaviour was measured by creating two different driving conditions. In the first, internal cueing was possible. Participants used a map to memorize the road sequence before the start of each trial. The road sequence was easy to remember because it was repetitive in nature, and was exactly the same in each simulation. No participant exhibited any difficulties memorizing the road sequence, and all were able to verbally recall the road sequence before and after each trial. In the second situation, participants could use the advance information from the map to internally cue driving behaviour for upcoming road obstacles. People with PD exhibited difficulties using internal cues to regulate driving behaviour around traffic signals and curves. Instead of using internal cues, participants with PD were more reliant on external cues to regulate driving behaviour. To investigate how external cues impacted driving performance, the presence/absence of external cues, which warned of upcoming obstacles, were manipulated. Participants with PD were less able to adapt their driving behaviour to suit driving conditions. Because all participants with PD were current drivers in the mild-to-moderate stages of the disease, the findings challenged the widely held assumption that cognitive difficulties only impact driving performance in the moderate-to-severe stages of PD.

Validation study of elderly car drivers in a simulator

In (Hakamies-Blomqvist et al. 2000), the way elderly test subjects drove along the same route in an instrumented vehicle and a driving simulator were compared. Both interviews and test data were collected. The research was conducted with an early version of the high-level VTI driving simulator.

The results showed that the test subjects drove roughly the same way in the simulator and in the instrumented vehicle. Validation was weaker with persons who had difficulties in driving. The main differences between the simulator and the real world were found at speed on 70 km/h sections, lateral position on carriageways, differences in variation in speed and steering wheel angle, and the frequency of braking.

According to the results of the study, the more reason there is to test, the less satisfactory is the simulator as a test environment for elderly people. The best validity was achieved with those elderly test subjects who had no difficulties in driving. They concluded that a simulator of the type studied here is particularly well suited for studies of elderly drivers in which the issue is relative in nature. The simulator did not give the same realistic picture of the finest details of driving behaviour. These shortcomings are particularly important in view of the legal protection provided to drivers if such testing is carried out for purposes of exclusion, since the validity of the simulator is lower when people who have difficulties in driving are studied.

Evaluating driving performance of cognitively impaired and healthy older adults: A pilot study comparing on-road testing and driving simulation

(Freund et al. 2002) compared the on-road and simulated driving performance of nine older adults (five men and four women) aged 67 to 78. Of these participants, four were classified as cognitively impaired and five were control subjects. The subjects completed a 30-minute on-road test along with a 30-minute driving simulation through an urban course, programmed to require execution of manoeuvres that allowed an assessment of the ability to drive and perform executive function tasks in a multitask environment. Conditions that revealed errors of discrimination were emphasized. The driving simulator used was a commercially available product, the low level STISIM Drive, developed by Systems Technology. A dual-brake equipped automobile was used for all on-road evaluations, and an occupational therapist dually certified as a driving rehabilitation specialist and commercial driving instructor who was blind to the results of the driving simulation evaluated performance. Performance measures for both driving tasks included hazardous or potentially catastrophic errors, traffic violations, and rule violations. The mean scores for simulated and on-road performances were significantly correlated. The results showed a strong association between the two driving tests. There was a strong association between hazardous and lethal errors committed in the simulator and failing the road test. Subjects who failed the on-road test committed a mean of 5.4 hazardous errors and a mean of 4 lethal errors while using the driving simulator compared with no errors in those categories by subjects who passed the on-road test.

Despite the small sample size, there was a correlation between simulated driving performance and on-road driving performance by cognitively impaired and healthy older adults. The researchers claimed that high fidelity driving simulation is a method that is sufficiently sensitive to objectively evaluate driving performance and may be a valid alternative to on-road testing.

4.3.5 Stability over time and driver characteristics

The results from driving simulator studies must not only be able to be compared to results from the real world. They also have to be consistent over time, and compared to different driving simulator setups.

Driver distraction: The effects of concurrent in-vehicle tasks, road environment complexity and age on driving performance

(Horberry et al. 2006b) presented the findings of a simulator study that examined the effects of distraction on the driving performance of drivers in three age groups. There were two in-vehicle distracter tasks: operating the vehicle entertainment system and conducting a simulated hands-free mobile phone conversation. The effect of visual clutter was examined by requiring participants to drive in simple and complex road environments. Overall measures of driving performance were collected, together with responses to roadway hazards and subjective measures of driver perceived workload. The two in-vehicle distraction tasks degraded overall driving performance, degraded responses to hazards and increased subjective workload. The performance decrements that occurred as a result of in-vehicle distraction were observed in both the simple and complex highway environments and for drivers in different age groups. One key difference was that older drivers travelled at lower mean speeds in the complex highway environment compared with younger drivers. The conclusions of the research were that both in-vehicle tasks impaired several aspects of driving performance, with the entertainment system distracter having the greatest negative impact on performance, and that these findings were relatively stable across different driver age groups and different environmental complexities.

Test-retest reliability of standard deviation of lane position as assessed on a PC-based driving simulator

To examine the test-retest reliability of the standard deviation of lane position (SDLP), (Marcotte et al. 2003) tested subjects at a three-month retest interval and a year or longer retest interval. SDLP is a frequently used metric for assessing driving ability that measure how much subjects “swerve” within their driving lane. This measurement has been used with individuals under the influence of alcohol, illicit drugs, and prescribed medications in both on-road and simulator studies. Test-retest reliability is critical if one is to measure change in individuals over time. SDLP was assessed in both groups using an interactive PC-based driving simulator that consisted of a monitor, steering wheel, and brake/accelerator pedals. Participants were required to maintain lane position while holding a constant speed (55 mph) and responding to divided attention tasks in the corner of the monitor. The researchers claimed that SDLP was a reliable measure for periods ranging from months to years when assessed in cognitively stable subjects. As such, this may serve as a useful tool in tracking the effects of neurological disorders and pharmacologic treatments on driving abilities.

4.3.6 Driving training

Simulators are in general often used for training. It is therefore relevant to study the use of driving simulators in this context. The main purpose of flight simulators is for training, and usually they are used for training for special situations that are difficult or cannot be represented in real life. The validity of using driving simulators for training will be dependent on the stated purpose of the training.

The perceptions of emergency vehicle drivers using simulation in driver training
(Lindsey and Barron 2005) assessed the perception of adding a driving simulator to a traditional training program for emergency service providers. The number of accidents over the past decade involving emergency vehicles in the United States was a major concern. The sample population consisted of Emergency Medical Technician students attending the National EMS Academy in Lafayette, LA. The group self-scheduled the day they would attend the driving portion of the class. This resulted in 52 participants in the control group and 50 participants in the treatment group. The treatment group used a driving simulator prior to driving the competency course. Surveys were used to assess the emergency vehicle operators' perceptions of using a driving simulator as part of an emergency vehicle training course. The simulator allowed the treatment group to understand the course prior to actually driving the course. The control group thought the simulator would have afforded them the opportunity to learn the course before actually driving the course. Both groups thought the simulator should be a part of the driver training course, but did not see the simulator replacing actual driving experience.

The development and evaluation of a high-fidelity simulator training program for snowplow operators
(Drews et al. 2005) reported the results of a pilot training program incorporating high-fidelity simulation developed for snowplow operators. Ratings of user acceptance of the training were very high, with drivers of all levels of experience indicating that the training helped them prepare for several issues critical to the safe and efficient operation of a snowplow. In the six-month period following training, the odds of getting in an accident were lower for the group of drivers who received training compared with a matched control group who did not receive it. In addition, the data indicated that fuel efficiency was greater for the trained drivers than for the control group.

Development and validation of virtual driving simulator for the spinal injury patient
(Ku et al. 2002) developed a virtual reality (VR) driving simulator in order to safely evaluate and improve the driving ability of spinal injury patients. The simulator was composed of an actual car, a beam projector, and a large screen. For the interface of the driving simulator, an actual car was adapted and then connected to a computer. The car was equipped with hand control driving devices especially adapted for spinal injury patients. The virtual environment consisted of 18 sections (e.g., a speed-limited road, a straight road, a curved road, a left turn) and each section was linked naturally to the next. The subjects selected for this trial were 10 normal drivers with valid driving licenses and 15 patients with thoracic or lumbar cord injuries who had prior driving experience. For evaluation, five driving skills were measured, including mean speed, steering stability, centreline violations, traffic signal violations, and driving time in various road conditions such as straight and curved roads. The normal subjects

manipulated the gas pedal and the brake with their feet, while the patients manipulated a hand control with their hands. After they finished driving the whole course, the participants answered questions such as, “How realistic did the virtual reality driving simulator seem to you?” and “How much was your fear reduced?” In this study, the difference in manipulation method did not seem to influence relative performance in the VR driving simulator, although the researchers claimed that training to improve the use of hand controls in the VR driving simulator would be useful to reduce the fear that the patients feel while driving.

4.4 Previous validation of the NTNU/SINTEF simulator

(Moe 1995) conducted a research project to examine the test participant’s view of the validity of the ordinary car in the driving simulator. In all, 17 professional drivers and 20 ordinary drivers took part in the validation. The validation was conducted using a questionnaire, which the test subjects answered after they had tried the simulator. The validation was divided into physical, operational and physiological experience regarding the perceived realism of the simulator.

The physical realism of the simulator is described by how realistic the vehicle is with regard to instruments, interior and equipment. The physical realism of the vehicle was considered high, since a complete vehicle was used with all the equipment found in a real vehicle.

The operational realism is how the test participants experience the operation and handling of the car. There was a significant difference between how the professional drivers and the ordinary drivers perceived this. On a scale from 1 to 7 where 1 is not realistic and 7 is very realistic, the professional drivers answered an average of 3 while the average of the ordinary drivers’ answers was 4.

The psychological realism is how the test participants experience the different elements of the driving process and the overall driving process. Again the professional drivers had a significantly lower experience of this realism than the ordinary drivers. The professional drivers assessed the physiological realism to be between 2.5 and 3 while the ordinary drivers assessed it to be just above 4.

A great deal of research has been conducted using the NTNU/SINTEF graphical driving simulator. Up to now, there has been no recording of real traffic data intended for validating data from the driving simulator. The behavioural validity of the driving simulator has been assumed based on the level of physical validity found by (Moe 1995) and by comparing the results from the simulator with existing knowledge.

4.5 Threats to driver simulator validation

4.5.1 What is to be validated?

This section focuses primarily on how to improve the validity of traffic data collected in the driving simulator. The focus of this thesis is on behaviour validity. Since there is no one single “truth” against which we can compare driving simulator research, I will also provide some discussion regarding alternative methods for behavioural research. The complete understanding of any problem requires using a variety of methodological approaches. The use of a driving simulator is just one of a number of possible alternatives.

Validating a driving simulator is difficult and must be seen in relation to the research topic. Usually the purpose of conducting research in a driving simulator is to be able to gain an understanding of driving in real traffic. To do this one has to be able to generalize the results from the driving simulator to the real world driving situation. This is probably the one topic related to driving simulator validation that is the most discussed. I will discuss in subsequent sections several more topics related to driving simulator validation that I feel are very important.

There are several advantages of using driving simulators measurements that make it difficult to compare results with real world traffic data. One such example is the possibility of collecting data in situations that would be dangerous in real life, thus prohibiting real life measurements. One has to rely on data that is not directly used in the research to validate the data collected. For instance it is difficult to know how valid the driving behaviour of a drunken driver is in the simulator as compared to his or her actual behaviour in real life. However, one can use data collected from sober persons to increase the validity of the research on drunk drivers.

As described earlier, it is not possible to describe the simulator as an absolute valid research tool. However, it is possible to increase its validity as a research tool. According to (Godley 1999) the accumulated evidence from different driving simulators and a range of driving tasks does add weight to the validity of simulator research. This means that even though a validation is targeted at validating specific data collection from the simulator, it may be useful for later validation of similar or other traffic data collection in the driving simulator.

To look at the validity of using a driving simulator in research I will use the topology proposed by (Shadish et al. 2001). This topology relates validity to threats to validity. These threats are divided into 4 types.

In Chapter 5 to 7, I will select some appropriate threats to describe the validity of using the driving simulator for research about a specified subject.

In section 4.5.7 I propose standard set parameters that should be recorded to make it possible to compare different research projects, whether conducted in a driving simulator or other comparable research. The ability to compare different experiments and to compare results with other research methods is an important tool in reducing the threats to validity that result from using the driving simulator as a research tool.

4.5.2 The threats to validation concept

A series of criteria can be established to improve the validity of the simulator as a research tool and the research projects that are conducted using this tool. The criteria for validation in this thesis is based on the validity system proposed by (Shadish et al. 2001) briefly described in section 4.1.3 , along with the adaptation by (Elvik 1999) related to validation of evaluation research. The validity of a study is related to four types of validities:

- Statistical conclusion validity;
- Internal validity;
- Construct validity; and
- External validity.

I have adapted and described the relevance of these measures to driving simulator research in Sections 4.5.3 to 4.5.6 .

4.5.3 Statistical conclusion validity

(Elvik 1999) states that "Statistical conclusion validity refers to the numerical accuracy, reliability and representativeness of results of a study or a set of studies."

(Shadish et al. 2001) defines nine threats to statistical conclusion validity:

Low statistical power

"An insufficiently powered experiment may incorrectly conclude that the relationship between the treatment and outcome is not significant."

Experiments in a driving simulator are very costly and time consuming. The number of test subjects is therefore often limited. NTNU/SINTEF usually uses between 20 and 30 subjects, unless the purpose of the experiments is to compare groups, in which case 20-30 subjects in each group. As shown in section 2.2 some researchers use as few as 10 test subjects for experiments in driving simulators. This is very different from non-experimental evaluation research, where research usually is based on larger data samples. An example of this is the evaluation of traffic safety measurements that can be measured in accidents in road traffic, which can be described as the number of accidents/million passenger kilometres. We will never acquire a data source of million passenger kilometres using a driving simulator.

Violated assumptions of statistical tests

"Violations of statistical test assumptions can lead to either overestimating or underestimating the size and significance of an effect."

Statistical tests are based on assumptions about distribution. It is often convenient to rely on standard normal distribution. At the same time, some tests are more robust in situations with violations of normality, such as the t-test, particularly if the group sample sizes are large and about equal in size. This must be remembered when dealing with driving simulator studies whether the results are compared within the study or compared to evaluation research.

Statistical tests also require independent observations. In a driving simulator this may be a problem if there are reasons for not having perfect random selections of test subjects.

Fishing and the error rate problem

"Repeated tests for significant relationships, if uncorrected for the number of tests, can artificially inflate statistical significance. "

The use of a driving simulator can generate a lot of data. Section 2.1.5 presents the more than 14 possible parameters that can be measured in the NTNU/SINTEF driving simulator. These data can be recorded 20 times/second. Since it is relatively easy to divide test subjects into groups, and the simulator provides in-depth information about road conditions, it becomes clear that there is the potential for fishing for significant results using data collected in a driving simulator.

Unreliability of measures

"Measurement error weakens the relationship between two variables and strengthens or weakens the relationships among three or more variables."

The reliability of measures is very important in all research. Some make a distinction between reliability and validity, but I will follow the description presented by (Shadish et al. 2001) and will use the term as one basis for validity. It can often be assumed that variables can be reliably registered in a driving simulator, but this may not always be true.

In the case of recording reaction time, as described later in Chapter 5, it was discovered that the time of the start of the incident was not accurately measured using automatic measurements. To get reliable recordings, the start of each incident had to be manually recorded using the recorded video.

Often the use of manual measurements would lead to less reliable measurements, especially if different measurement personnel are used. An example of this is given in (Hjälmdahl and Várhelyi 2004), where the reliability of observations made by different observers was compared with a key representing a correct observation. The research is described in further detail in section 3.4.2 .

Restriction of range

"The reduced range of a variable usually weakens the relationship between it and another variable."

In a driving simulator, this can be particularly important when studying special groups of drivers. One example of this is when studying the effect of information technology on workload. It would be likely that a group of drivers older than 80 years would be at one end of the scale while emergency vehicle drivers would perhaps be at the other end. If another project planned to study the effect of driver education on either one of these two groups, it would be impractical to use the same measures and the same scale for such studies.

Unreliability of treatment implementation
“If a treatment that is intended to be implemented in a standardized manner is implemented only partially for some respondents, effects may be underestimated compared with full implementation.”

In a driving simulator, this might involve situations where the physical attributes of the simulator limit the implementation of a measure. An example of this might be the lack of motion system in a driving simulator. Even without motion system, it is possible to conduct research of the effect on rumble strips, by using sound, but the lack of motion would probably affect the results from the treatment.

Extraneous variance in the experimental setting:
“Some features of an experimental setting may inflate error, making detection of an effect more difficult.”

A driving simulator usually makes it possible to control experimental settings, but this might not be true in a field setting, such as in an instrumented car. Noise, temperature, weather and other conditions that are not easily controlled might therefore influence the results. Nevertheless, this must be taken into consideration when conducting experiments in a driving simulator, so that experimental settings do not change during an experiment. An example of this might be the need to turn on the ventilation of the vehicle because a test subject becomes sick. A change in the use of ventilation can influence driver performance.

Heterogeneity of units:
“Increased variability in the outcome variable within conditions increases error variance, making detection of a relationship more difficult.”

It is possible that the units used to measure an outcome are not heterogeneous over the whole measurement range. This is particularly important when calculating an index to compare driver performance. If some variable are of particular importance for one group of drivers, while other variables are more important for other drivers, the index scale might not be appropriate to compare changes between the groups. This might be the case in research that was conducted by (Lee 2002) and presented in section 4.3.4 .

Another problem will arise if ordinal measurements are treated as scale measurements. An ordinal scale might use the labels: very low, low, medium, high, and very high. Ordinal variables allow the description of the order of values, but make it impossible to determine if the distances between the variables are equal. This situation might arise if researchers employ personnel to assess driving performance levels, for example.

Inaccurate effect size estimation:
“Some statistics systematically overestimate or underestimate the size of an effect.”

There are different estimates used to measure effects, such as mean or median. In situations where there are extreme values or outliers, the use of mean effects to compare two variables may not provide the correct estimate.

4.5.4 Internal validity

(Elvik 1999) states that “Internal validity refers to the possibility of inferring a causal relationship between the measure that is being evaluated and the dependent variables this measure is intended to influence.”

(Shadish et al. 2001) defines nine threats to internal validity:

Ambiguous temporal precedence
“Lack of clarity about which variable occurred first may yield confusion about which variable is the cause and which is the effect.”

This threat to the validity of the research is important in observational studies. Experiments are conducted because it is possible to know which factor was deliberately manipulated before another was measured. In observational studies it is often difficult to be certain which variable is the cause and which variable is the effect.

Selection
“Systematic differences over conditions in respondent characteristics that could also cause the observed effect.”

The selection of test subjects is important to the outcome of the study. This threat to validity is of great concern related to surveys and every effort must be made to minimize this threat. Ideally, test subjects should be selected on a randomized basis within the group intended for study. This is often not possible, partly because test subjects cannot be randomly be asked and forced to take part in an experiment in a driving simulator. Different methods have to be used to identify test subjects.

The test subject’s background and interests will influence his or her desire to take part in driving simulator studies. At the NTNU/SINTEF driving simulator, we have recruited a pool of test subjects who have expressed interest in taking part in driving simulator experiments. This pool is getting bigger and more diverse all the time, but I believe that test subjects who are part of this pool have more interest in computers and cars than the ordinary driver. Many validation studies related to the generalization of results from driving simulator experiments presented in section 4.3 have also been conducted with similar groups.

Often test subjects have to be recruited outside of this self-selected pool, and often more or less pressure has to be used to encourage these non self-selected individuals to take part in the experiment. An example of this is studies of older people or more specifically, older female drivers. Special care has to be taken when results from these groups are compared to reference groups.

History
“Events occurring concurrently with treatment could cause the observed effect.”

During any one experiment conducted in a driving simulator, history will not likely affect the outcome. But if the experiment is compared to earlier experiments, history may have an effect. For example, experiments in a driving simulator are often related to driving support systems. Driver support systems such as navigation are rapidly becoming more common in the real world. Test subjects with experience using such systems might influence the outcome of driving simulator experiments.

Maturation:
"Naturally occurring changes over time could be confused with a treatment effect."

Maturation, like history, is related to change over time, but relates more to the natural changes in test subjects over time. This change must be taken into consideration if one wants to follow test subjects over time. One example would be if the purpose of the experiment was to follow the different effects of driving training in a driving simulator over a specific time period and related to different age groups. A hypothesis might be that effects such as maturity would influence a 16-year-old while this might not have any effect on 50-year-old test subjects. This would have to be taken into consideration in the design of the study.

Regression
"When units are selected for their extreme scores, they will often have less extreme scores on other variables, an occurrence that can be confused with a treatment effect."

This is a threat to validity that has attracted a great deal of attention in evaluation studies, such as before and after studies. This artefact may be found when treatments are undertaken because of an unusually high number of accidents in the "before" period. Such cases would often have high values because of natural variance in accidents. This would lead to an expected natural decline in accidents, even without the introduction of the treatment. This effect is called the "regression to the mean" effect.

The effects of regression would also be found in experiments in a driving simulator if test subjects were selected because they had an extreme score on a test. A retest of the same subjects can lead to a less extreme score, even without the introduction of a treatment in a driving simulator. One example might be in an experimental study of how a treatment affects the choice of speed. After the initial tests, the test subjects could be divided into three groups based on their speed at a certain place: those who drove fastest, slowest or an average speed. The effect of regression would lead to the probable effect that those who drove slowest would on average drive faster, those who drove fastest would drive on average a little slower, and those who kept to the mean speed would show the least change in mean speed.

Attrition
"Loss of respondents in various treatments or measurements can produce artificial effects, if that loss is systematically correlated with conditions."

Often this threat is called mortality. The study of older people is one area where the use of driving simulators is popular today. Repeated experiments on patients with Alzheimer's disease might be an example of experiments where those who are most severely affected by the disease would be expected to have greater attrition.

Another important threat is the influence of test subjects who become sick during simulator experiments. At the NTNU/SINTEF driving simulator, test subjects are told to stop the experiment if they start to feel sick. In our experience certain people are more prone to driving simulator sickness. To reduce this threat to validity, we try to include test subjects in our subject pool who are least likely to get driving simulator sickness.

Testing
"Exposure to a test can affect scores on subsequent exposures to that test, an occurrence that can be confused with a treatment effect."

The act of testing can lead to a change in the test subject; subjects might learn how to react to certain situations or learn to anticipate future situations. An example of these effects can be found in Chapter 5, where the study of reaction times is described, and the test subject reaction times to subsequent trials were found to be dramatically reduced compared to the first trial.

Instrumentation
"The nature of a measurement may change over time or conditions in a way that could be confused with a treatment effect."

Usually the instrumentation in a driving simulator is not changed during the course of one experiment, but there are situations that could have an effect. Some instrumentation, such as the eye tracker presented in section 2.4.2 must be calibrated for each person, and is very sensitive to movement of the cameras. One camera is placed in the door of the vehicle, which is opened and closed every time a test subjects enters the vehicle.

If experiments are conducted at different times, there is also a danger that changes in the driving simulator may cause changes in results. As has been described in section 2.4.1 the NTNU/SINTEF driving simulator is upgraded on a regular basis.

Additive and interactive effects of threats to internal validity
"The impact of a threat can be added to that of another threat or may depend on the level of another threat."

Several threats to validity may affect the study at the same time. These effects might be either additive or multiplicative, and can be difficult to predict and measure, but should be taken into consideration when planning the experiment.

4.5.5 Construct validity

(Elvik 1999) used the term theoretical validity instead of construct validity. (Shadish et al. 2001) described construct validity as: "Threats to construct validity concern the match between study operations and the construct used to describe those operations." They define fourteen threats to construct validity:

Inadequate explication of constructs
"Failure to adequately explain a construct may lead to incorrect inferences about the relationship between operation and construct."

As I have shown in Section 2.2.1 there are several different options for describing and measuring workload. These different options might lead to either a too general or too specific interpretation of the results. For example, if lateral position is used as a sole indicator for workload, it probably is an insufficient variable to describe workload, which would therefore lead to a overly general interpretation of this measure. At the same time, lateral position might be influenced by sleep deprivation, making the interpretation overly specific, or even incorrect, if conclusions about workload are drawn based on this variable.

Incorrect experiment constructs might also lead to the wrong conclusions. For example, if the study concerns the use of cell phone while driving and a Nokia phone is used, test subjects who own a Nokia phone would probably perform better than those who own a different brand, because individuals who own a Nokia are already familiar its operation.

Construct confounding
"Operations usually involve more than one construct, and failure to describe all the constructs may result in incomplete construct inferences."

This is an important consideration when conducting research on different groups of people. For example, in studying the effect of a year-long loss of a driving licence on driving performance, there is no simple way of selecting drivers for the study before the individuals have actually lost their licence. Instead, the study would compare subjects who have not lost their driving license with those who have lost their driving licence for more than a year. It is likely that the lost-license group would on average drive faster than the group that still has a licence. The reason for this is that most drivers who have lost their licences have done so because of speeding.

Mono-operation bias
"Any one operationalization of a construct both under-represents the construct of interest and measures irrelevant constructs, complicating inference."

This threat to validity describes the possibility that measuring only one construct will result in a lack of different outcomes when varying the construct. An example of this can be found in Chapter 5 where studies of reaction time have been presented. Different situations, such as braking by the vehicle in front, or a vehicle unexpectedly entering the road, are tested to study the effect on reaction time.

Mono-method bias
"When all operationalizations use the same method (e.g., self-report), that method is part of the construct actually studied."

Sometimes there must be several methods to record a response. The study of reaction time described in Chapter 5 included several methods for measuring a reaction: release of throttle, use of braking pedal, or use of steering wheel for evasive manoeuvre. The reaction first detected is used. If only one method of detection had been used, this would certainly affect the outcome of the study.

Confounding constructs with levels of constructs
"Inferences about the constructs that best represent study operations may fail to describe the limited levels of the construct that were actually studied."

If treatments are implemented at levels that are too low, this might cause no effect. For example, in a study of the effect of alcohol on driving, the use of only one level of intoxication may lead to the wrong conclusions. If the test is only on the effect of blood-alcohol levels of 0.2 per thousand, one might not find any effect. The use of several treatments levels would increase the validity of the result. Chapter 6 contains an example of this, in the study of the effect of road and lane width on lateral position.

Treatment sensitive factorial structure
"The structure of a measure may change as a result of treatment, but this change may be hidden if the same scoring is always used."

This threat to validity is of particular importance in surveys if the subjects learn to see the possible responses in a different way, and thus in effect change the scale. An example would be to see if a campaign to reduce speed has any effect. A survey can be used before and after the campaign, where one question might ask if the subject considers his or her speeding to be dangerous. If the focus of the campaign is the effect of speed on accidents, this might lead subjects to respond that their unchanged driving speed is more dangerous in response to a survey conducted after the campaign.

Reactive self report changes
"Self reports can be affected by participant motivation to be in a treatment condition, motivation that can change after the assignment is made."

The possible subjects for a study might have strategic reasons for either taking part or not taking part in an experiment. For example, if a treatment is to be given to those who have a low score, the subjects might score intentionally low to get a specified treatment. After receiving treatment they have no incentive to have a low performance, and therefore their improvement in performance is registered to be larger than it really is.

Test subjects may ask if the result of the driving simulation study can result in the loss of their driving licence. This might be of special relevance when studying older people who are afraid they will lose their licence if they perform poorly, which could result in attracting only those who have high opinion of their driving skills for the survey. Another tactical reason for not taking part in an experiment is that a test subject may fear affecting the general judgement of their group as drivers if the individual performs poorly. I have heard of older people who do not want to participate in a driving simulator study because they are afraid that if they perform poorly, it might lead to a negative view of older drivers.

Reactivity to the experimental situation
"Participant responses reflect not just treatments and measures but also the participants' perceptions of the experimental situation, so that those perceptions are become a part of the treatment construct actually tested."

This threat to validity includes the possibility of a placebo effect. This is of particular importance in drug testing, where the mere act of being given a pill may cause an improvement in a subject, even if the pill contains only sugar. This is also an important factor when conducting research in the driving simulator about the effect of drugs.

Knowledge about the purpose of the experiment can influence the subjects to perform a certain way. Whenever possible and ethical, the true nature of the experiment conducted is hidden from test subjects at the NTNU/SINTEF driving simulator. This was the case in the research about the influence of lane width reported in Chapter 6. Although the subjects became aware of some of the changes, some changes were almost undetectable unless one knew about them before the study was conducted.

Experimenter expectancies

"The experimenter can influence participant responses by conveying expectations about desired responses, making those expectations a part of the treatment construct actually tested."

This threat is closely related to the previous threat, but unlike the previous threat, where the expectations of the test subject could influence the outcome, this threat deals with the influence of the researcher's expectations. This is a very real threat in some simulator studies, since the test subjects have to be given an introduction to the simulator and information about the driving situation.

Novelty and disruption effects

"Participants may respond unusually well to a novel innovation or unusually poorly to one that disrupts their routine, a response that must then be included as part of the treatment constructs description."

One problem with using driving simulator experiments might be the difficulty of studying long-term effects. In real world observational studies, it is possible to conduct consecutive studies to study behavioural effects of a measure. Often the purpose of a driving simulator study is to evaluate the effect of a new measure. An example of this can be seen in Chapter 6, where the effect of the introduction of a visual central reserve is reported. Both the immediate effect and long-term effects have been evaluated.

Compensatory equalization

"When a treatment provides desirable goods or services, administrators, staff, or constituents may provide compensatory goods or services to those not receiving treatment, and this action must then be included as a part of the treatment construct description."

This threat addresses the compensation that test subjects receive either because they are perceived as being treated favourably or unfavourably. In evaluation studies, these might include measures that are used in a control group that are not brought to the attention of the researcher.

In relation to driving simulator studies, this might include special favours to special groups of test subjects. For example, in a study examining if smokers perform better under stressful situations, smokers might be allowed leave the study area to smoke. This special treatment might lead to a more relaxed attitude.

Compensatory rivalry
"Participants not receiving a treatment may be motivated to demonstrate that they can do as well as those receiving the treatment, which must then be included as part of the treatment construct description."

In a comparison of groups, some test subjects might feel as if they are in a competitive situation and will want to perform better than they would in a non-competition situation. In the threat related to reactive self-report changes, I described situations where drivers were reluctant to take part in an experiment because their thought they would perform poorly. In a similar vein, those who do participate in an experiment may train in real traffic either before the experiments or between experiments in the driving simulator to perform better than they normally would have done.

Resentful demoralization
"Participants not receiving a desirable treatment may be so resentful or demoralized that they may respond more negatively than otherwise, and this resentful demoralization must then be included as part of the treatment construct description."

Just as some test subjects might try to perform better because they are taking part in a competition, some test subjects may in some cases perform more poorly because they are demoralized. This might be the case when the experiment is perceived as their last chance to receive a benefit. In a driving simulator experiment, this might occur in experiments where test subjects are on the borderline of losing/regaining their driving licence on the basis of physical characteristics. One possible outcome could be lower scores if they feel they have performed so poorly in early tests that no matter the outcome of the current test, they will not be able to retain their driving licence.

Treatment diffusion
"Participants may be affected by a situation to which they were not assigned, making construct descriptions of both conditions more difficult."

This threat to validity is most important in evaluation research. A control group might unconsciously be influenced by measures introduced to influence the experiment group. This might result if an awareness campaign is conducted in one part of the country while another part of the country is used as a control group. If the campaign is covered in the media, the control group can also be affected by the awareness drive.

4.5.6 External validity

(Elvik 1999) states that "External validity refers to the possibility of generalising the results of a set of studies to other contexts and settings than those in which each of studies in the set was made." External validity is related to extent to which the results can be used on different individuals, settings, treatments and outcomes. When designing experiments, it is not possible to use multiple designs, multiple treatments and variances in groups of test subjects. The external validity of an experiment is therefore closely related to the construct of the experiment.

(Shadish et al. 2001) defines five threats to external validity:

Interaction of the causal relationship with units: An effect found with certain kinds of units might not hold if other kinds of units had been studied.

This threat to external validity has special importance related to driving simulators because of the possibility that the results might change if different test groups had been used. This is the case for age, gender, physical capabilities, culture, nationality and other characteristics. The driving simulator makes it possible to conduct specialized experiments, which means in some circumstances it might be appropriate to use test subjects who best represent the group being tested, even if this results in a substantial extra cost. At the NTNU/SINTEF driving simulator, we have used Danish test subjects for testing the best design of a Danish roundabout with respect to bicyclists, and we have used persons from Chinese culture to test the design of Chinese tunnels.

Interaction of the causal relationship over treatment variations
"An effect found in one treatment variation might not hold in other variations of that treatment, or when that treatment is combined with other treatments, or when only part of that treatment is used. "

It is not possible to conduct experiments using every feasible combination of treatments. A combination of treatments may give better results in some cases, while other combinations can have an effect of nullification.

Interaction of the causal relationship with outcomes
"An effect found from one kind of outcome observation may not hold if other outcome observations were used. "

In traffic safety research, this can be exemplified by looking at different level of traffic injuries. If the number of killed is used to measure the outcome of the measure, is it possible to predict the outcome of injured, or damage to vehicles?

Interactions of the causal relationship with settings
"An effect found in one kind of setting may not hold if other kinds of settings were to be used. "

This might be the question that most are concerned with: how well do the results found in the setting of driving simulator generalize to the setting of the real world. As shown in Section 4.3 field experiments and observational studies are often used to decrease the threat to this validity.

Context-dependent mediation
"An explanatory mediator of a causal relationship in one context may not serve as a mediator in another context. "

This threat could be related to a contextual change in setting, test subjects or the nature of the treatment.

4.5.7 Standard set of parameters to be recorded

To reduce the threats to validity both in the current experiment and possibly in future experiments, it is important to include data which makes it possible to compare some or all of the results with other studies, both in the simulator and in the real world.

All of the following points may affect the results of the studies and thus are also important when it comes to validity of the data:

- Basic data to be collected;
- The physical characteristics of the driving simulator;
- The virtual representation of the road;
- The gender, age, nationality and driving experience of the test drivers; and,
- Instructions given to the test participants.

Some basic driving parameters should be recorded in all research conducted in the driving simulator. The most basic parameters are speed and lateral position. These are probably the easiest parameters to compare to real world situations.

Other parameters, such as the throttle and brake pedal pressure, steering wheel angle, and use of signal lights could also serve as a basis for improving the validity of research. A driver's physical influences can be measured, if applicable, with equipment such as an eye tracker or a heart rate monitor.

The physical characteristics of the driving simulator usually change over time. At the NTNU/SINTEF graphical simulator, much of the computer hardware has changed since the first installation. A description of the state of the driving simulator is important.

The NTNU/SINTEF driving simulator has the ability to change the virtual road. These may be simple changes such as road markings, but this could significantly change the traffic data recorded. Information about the virtual road used is important.

Many experiments have been conducted at the NTNU/SINTEF driving simulator, which means that quite a few people have tried it. When new research is attempted, this group of people is often called upon, because they have previous experience with the driving simulator and do not need as much training in the simulator before participating in the research. These test subjects also tend to be less prone to simulator sickness.

In addition to the driving in the simulator, the handling and presentation of the research to the test subject is important for the validity of the data.

Although it is impossible to conduct a "be-all and end-all" validation of the driving simulator with respect to traffic data, it is possible to build up a database of experience from previous validations. To be certain the research is conducted correctly, one usually should undertake a basic validation for speed. The validation should both be undertaken with a comparison of the data to the real world and to earlier research in the driving simulator. Doing this will reduce the likelihood of errors in the design of the research.

4.6 Summary

In choosing to use a driving simulator to conduct research, the expectation is that the results will be an accurate and correct representation of how individuals drive in real traffic. To obtain this knowledge, research conducted using the driving simulator needs to be validated to see if drivers behave the same way in a driving simulator as in real life.

The term validity can be described as "refer[ing] to the approximate truth of an inference". The evidence of validity may come from other sources of knowledge, such as from past findings and theories. Validity judgement cannot be absolute.

Much of the data collected in driving simulators has been compared to real life data measurements. I have divided the research related to validity into different blocks:

- Direct comparison with real life data;
- Comparison of driving simulator with physiological tests and questionnaires;
- Expert testing;
- Validation compared to specific driver characteristics;
- Stability over time and driver characteristics; and,
- Driver training.

Some of projects discussed have related their findings to specific modules that I have described: the driver input module, the sound, visual and motion module, or the dynamic module of the interactive car. The traffic module for autonomous vehicles is very seldom discussed, while the accuracy and possibilities of the data measurement module usually are not discussed.

I have focused on different threats to validity in the use of the driving simulator as a research tool. Four specific types have been discussed with respect to the three research topics addressed in this thesis:

- Statistical conclusion validity;
- Internal validity;
- Construct validity; and,
- External validity.

The rather comprehensive and varied presentation of aspects related to validity in chapter 4, will hopefully provide a solid platform for planning of and evaluation of driving simulator based research.

5. Validation case 1 – Reaction time

5.1 Background

5.1.1 Challenges of the case study

In connection with the revision of the Norwegian Standards for geometric design of roads and streets, SINTEF was asked to investigate driver reaction time in road traffic. The project was financed by the Public Roads Administration in Norway. I was the project manager. The project results have been presented in (Engen and Giæver 2004).

The purpose of this chapter is twofold:

- To present the SINTEF project and extend the discussion about the validity of the results about reaction times.
- To study how the validity of using the NTNU/SINTEF driving simulator for reaction time studies was affected by the research.

5.1.2 The purpose of the SINTEF project

Reaction time is a basic parameter used to calculate standard geometric features of roads, such as when calculating sight distances at road intersections and curves.

Today, the standard design reaction time in Norway is set to 2.0 seconds. This same value is used in all traffic situations. The project goal was to update the "reaction time" parameter in connection with the revision of the Norwegian Standards for the geometrics of roads and streets. The assignment was to evaluate the possibility of using different values for reaction time in different situations.

Using different methods to find reaction times is beneficial in aspects of validation. At the same time this approach introduces new threats to the validity of the research. This will be further discussed in this chapter.

5.1.3 Validity of using a driving simulator for reaction time studies

This case study posed a situation where it was difficult to conduct a direct comparison between the driving simulator and the real world. Thus it was important to evaluate how best to validate the driving simulator studies, even when a direct comparison between real life and the driving simulator was not being made.

The focus of this discussion is to highlight the challenges in using driving simulators for this type of research, and to see if related results from different research methods could improve the validity of driving simulators as a research tool.

5.1.4 Data collection methods

Even though there have been many studies of reaction time, few researchers have tried to study reaction time in road traffic. The studies that have been undertaken in road traffic have limitations, because it is difficult to introduce situations where a fast reaction is required. It is possible to overcome this in driving simulators, but few such studies have been done.

Different types of data collection methods were used to improve the validity of the research:

- A literature review;
- Tests in the NTNU/SINTEF driving simulator;
- A few selected measurements in real traffic; and,
- Further analysis of earlier driving simulator research.

Literature review

The literature review documented earlier research and results about reaction times. There was a special focus on research that might substantiate differences in reaction time in different situations (sections 5.2 and 5.3).

Tests in the driving simulator

This study used the new E6 Støren– Soknedal road, described in section 2.5.2 . The road is about 10 km, and is a priority road in a rural environment. Sections of the road have already been built, which makes it possible at a later stage to make direct comparisons between traffic on the real road and in the driving simulator. We created different driving situations for the measurement of reaction times on this road, and a scenario to measure the reaction time for drivers in the different situations. Two sets of studies were conducted: one experiment using 31 test subjects in a strictly controlled setting, and one experiment using students allowed to test the driving simulator. The scenario, data models and the driving simulator remained the same, but the situation around the simulator was different. This made it possible to look at how factors such as introduction to the simulator, information before driving, and training would influence results (sections 5.4 and 5.5).

Measurements in real traffic

Measurements of reaction times in real traffic were also made. In addition to providing extra information about reaction time, the measurements also formed a possible basis for validation (sections 5.4 and 5.5).

Further analysis of earlier tests in the NTNU/SINTEF driving simulator

Both the graphical and video-based driving simulators have previously been used in a great deal of research. Previous tests from the graphical driving simulator that were not specifically targeted at measuring reaction times were difficult to use for calculating reaction times. On the other hand, many tests have been conducted in the video-based simulator to determine reaction times. These results can be used both to give an additional basis for a design reaction time, and to validate the driving simulator (sections 5.5.4).

5.2 Defining reaction time

Reaction time can be defined in different ways. Reaction time is often called perception-response time or just response time, and is often divided into different parts. (McGehee et al. 2000) divided reaction time into four stages:

- Detection - Starts when a subject comes into the driver's field of view and ends when the driver is conscious that something is present.
- Identification – Collection of adequate information such that a decision can be made.
- Decision – The decision about a possible action to be carried through.
- Response – The brain sends the instructions to the muscles, which carry out the selected task.

The simplest form of reaction time reduces detection, identification and decision to a minimum, for example, pressing a key when exposed to a stimulus. According to (McGehee et al. 2000), reaction times typically range from 0.16 to 0.25 seconds. By increasing the complexity of any of the four stages, as in a road traffic situation, the reaction time will increase.

(Olson and Farber 2003) argued that the reaction time is not normally distributed, but is biased. There are two reasons for this:

1. There is a limit to how fast a human can react, but there is no definitive limit for how slow a human can react.
2. Most of the reaction times will be relatively fast.

This means that the reaction time will not have the same mean time and median time.

This definition is very important related to the explanation for the research construct. The SINTEF project used the definition from (Boff et al. 1988): "The time from the onset of a stimulus to the beginning of the subject's response to the stimulus by a simple motor act." The time of "onset of stimulus" and "beginning of subject's response" can be difficult to determine.

The project used the following definition for "onset of stimulus": When a person registers or should have registered an object or an incident that leads to the need for an action. For example, this may be a person lying on the road beyond a curve, or a vehicle that was formerly stopped suddenly drives into a junction in front of you.

The following definition was used for the "beginning of subject's response": The time when the person starts the primary action as a result of the onset of the stimulus. Often there are several sequential actions that result from the stimulus, such as: release throttle, then to move the foot to the brake pedal, and then activate the brake pedal. Here the primary action is to activate the brake pedal. Often the term "brake reaction time" is used for this specific situation.

5.3 Literature

5.3.1 Studies of reaction time in road traffic

Traditionally, research on reaction times has been conducted on roads with real traffic. Since 2000, more research has been conducted using closed circuits and simulators.

The research can to a great degree be divided into four types of research.

Different tests in real traffic have been carried out to measure reaction time. Several studies have been carried out by recording the brake light of random vehicles exposed to different kind of stimuli. Such tests have been done both by using the brake light of a vehicle in front of the studied vehicle or at places with traffic signals. In (Wortman and Matthias 1983), the reaction time was measured at different junctions. The mean reaction time varied between the different junctions from 1.09 and 1.55 seconds, while the 85th percentile varied between 1.5 and 2.1 seconds.

In other studies, subjects have been asked to respond to a specific signal. For example, in (Johansson and Rumar 1971), drivers who agreed to participate in a study would use their brake pedal when exposed to a sound at an unknown place. The mean reaction time was 0.6 seconds, while the 85th percentile was about 1.0 seconds. A few drivers took as long as 2.0 seconds to respond. The (American Association of State Highway and Transportation Officials 2001) sets the standard reaction time for the USA at 2.5 seconds. This number is partly based on the results from (Johansson and Rumar 1971). When an unexpected event occurs, the reaction time increases by 35%. Therefore, based on (Johansson and Rumar 1971), reaction times may be as high as 2.7 seconds.

(McGehee et al. 2000) carried out a study of reaction times both in a driving simulator and on a test track. Both studies were implemented as an unexpected intersection incursion. On the test track, a foam core photographic mock-up car was used. In the simulator the mean time from incursion start to throttle release was 0.96 seconds, while it was 1.28 seconds on the test track.

Older drivers have an increased accident risk at complex intersections for a variety of manoeuvres. To study this, (Edwards et al. 2003) measured perception-response time in a driving simulator. The University of Calgary Driving Simulator was used to test healthy older drivers (65-83) and younger drivers (19-22). Critical scenarios included the sudden appearance of a pedestrian at an intersection, a last-second yellow light, an unexpected change during a left turn, and a vehicle violating a stoplight. Older drivers had significantly higher perception response times than younger drivers for the latter three of the four intersection scenarios. Analysis of specific manoeuvres also revealed qualitative response differences between young and old groups. Surprisingly, more older drivers ran the yellow light than younger drivers.

(Owens and Lehman 2001) studied the effects of age and distraction on reaction times in a driving simulator experiment. Participants operated a driving simulator while intermittently answering pre-recorded questions of various difficulty, or dialling specified numbers into a cellular telephone. Two road hazards were presented at unpredictable times and locations, including red brake lights and a red pedestrian-shape of approximately the same area as the brake lights. Targets were presented in two different locations: directly in front of the driver at the bottom of the screen, and off to the side of the road. The results showed a significant overall increase in reaction time for older subjects, as well as a strong dependency with the dialling task condition. There were no significant differences from the control for either easy or difficult verbal response conditions. In addition, stimuli on the side of the road took significantly longer to respond to, especially when combined with the dialling task. The data suggest a strong link between age, visual task load, stimulus location, and increased reaction time to unexpected stimuli.

(Hugemann 2002) discussed driver reaction times in German road traffic. Reaction time has been considered a constant at least since the « Verkehrsgerichtstag (DVG) » (German Council on Jurisdiction in Traffic) gave its recommendations in the early 1980s. The council concluded that even in most simple situations, reaction times of 1.5 seconds should be accepted as quite normal.

(Summala et al. 1998) studied the connection between driving experience and perception of braking by the lead car when the driver is looking at in-car targets. Perception of the lead car's braking was measured on-road, when subjects with various levels of driving experience were looking at a digital display located on the lower part of the windscreen, at the speedometer level, or mid-console. The results indicated that the detection of the lead car's brake lights, in daylight, is substantially impaired when a following driver is looking at the speedometer area, and that brake lights do not contribute to detection at all when he/she is looking at a target mid-console. Driving experience did not influence performance in detecting a closing headway in peripheral vision, in contrast to the improvement in lane-keeping found in an earlier study. The researchers suggested that such differential ability in using peripheral vision for lane and distance keeping may mislead experienced drivers when they follow another vehicle and perform certain in-car tasks.

(Muttart 2005) wanted to quantify driver response times based upon research and real life data by building upon previous research to identify the variables that significantly influenced driver response times, and to determine the size of that influence. The goal for this research was to explain why seemingly analogous published studies have come to very different driver response time results. The research divided analogous driver response situations into one of four groups: (1) lead vehicles that were stopped or moving slowly, (2) being cut off (when a vehicle changes lanes and pulls into the path of the responding driver), (3) path intrusions, or (4) known lights, icons or sounds. The findings showed that research on measured response times in analogous situations can be used to estimate the mean response time for a particular situation, if adjustments are made to account for methodological differences between the studies. Non-analogous studies are poor predictors of driver response (An anticipated light stimulus response cannot accurately predict the response time to a path intrusion or lead vehicle). Mean driver response times can be predicted within 0.4 seconds without accounting for individual difference. Therefore, the research concluded that external validity can be obtained regardless of the testing method (closed track, simulator or road), as long as the subject is unaware of either the stimulus or the appropriate response. The research concluded that having a subject respond to multiple events does not (by itself) suggest that drivers will respond significantly faster.

In (Green 2000), different literature was analysed to estimate reaction times under specific conditions. This literature review calculated the reaction time in different situations. The main conclusions were:

- High expectancy and little uncertainty - the fastest reaction time is between 0.7 and 0.75 seconds.
- Normal but common signals such as brake lights – expected reaction time 1.25 seconds.
- Surprise intrusions – mean 1.5 seconds.
- Urgency – faster reaction time when aroused to shorter time to collision.
- Extreme emergency – longer reaction time especially if there are competing response alternatives.
- Age – older people seems to respond 0.1 to 0.3 seconds more slowly than younger drivers.
- High cognitive load (complex roadway, cell phone, etc.) – slows reaction time.

As shown in the literature review, several different reaction times have been found in different research. Much as one would expect, it is not possible to conclude that one reaction time for road traffic is the true one. Road traffic is in itself complex and it is not possible to create one construct to describe it. Additionally, the construct of defining the measured variable of reaction time is complicated, and the variable of driver characteristics may also influence the reaction time.

5.3.2 Current standards in different countries

In terms of existing guidelines, we have mainly relied on information found in (Fambro et al. 1997), supplemented with information from (Meek 1990), (Olson and Farber 2003) and (Vegdirektoratet 1992) for the Nordic countries.

Reaction time used in different countries:

2.0 s	Norway, Austria, Great Britain, France, Switzerland, Germany, Greece, Denmark (Brake reaction time)
2.5 s	USA, Canada, South Africa, Denmark (Orientation in cross sections)

I have only found background data about the design reaction time for USA. According to the (American Association of State Highway and Transportation Officials 2001), the design reaction time for the USA has been based on three reports:

- A study by (Johansson and Rumar 1971), presented in the previous section.
- A report from MIT from 1935, which stated the mean reaction time was found to be 0.64 seconds, while 5% of the drivers ha a reaction time of 1.0 seconds or more. The highest reaction time was more than 1.5 seconds.
- A study by Norman from 1953, referred to by (Fambro et al. 1997), in which the reaction time varied between 0.4 seconds and 1.7 seconds.

Based on this, it is concluded that reaction times may be as high as 1.5 seconds, and 1.0 seconds should be added to any reaction time to allow for the unexpectedness of incidents in road traffic.

(Fambro et al. 1997) conducted three studies carried out in conjunction with a revision of the formula to calculate stopping distance. They found a mean reaction time of 1.1 seconds, while the 95% percentile was about 2.0 seconds. Based on their own studies and on a literature review, they concluded that the 90% percentile and 95% in unexpected situations would be respectively 2.0 seconds and 2.5 seconds. Partly because 2.5 seconds was the established design reaction time, they recommended keeping 2.5 seconds as the design reaction time.

5.4 Methods for measuring reaction time

The purpose of the SINTEF research projects presented in (Engen and Giæver 2004) was to determine a design reaction time. The reaction time is, among other factors, an important parameter when calculating sight distances. One variable that is important to measure is the mean reaction time, but this measurement cannot be used directly to determine a design reaction time because the design reaction time should include a larger percentage of drivers.

Norway and many other countries use the 85th percentile as a design value for different situations, and when determining design reaction time. This means that 85% of the drivers have a reaction time less than the design reaction time. The 95th percentile could also be used, as has been done in select earlier studies. The primary focus will still be the 85th percentile, to maintain its relationship to other design parameters. The omission of the 95th percentile is partly related to the number of recordings and test subjects, and partly related to the difficulty of defining a measurement as missing or as having a long reaction time.

5.4.1 Driving simulator

The reaction time experiment conducted in the driving simulator involved the creation of eight different situations, constructed so that test subjects would have an ideal speed. The situations and the intended speed are presented here in the order they were presented to the drivers:

1. A car suddenly drives out from a side road. Speed around 80 km/h.
2. A lorry ahead brakes when the test driver is less than 20 metres away. Speed around 60 km/h.
3. A bus drives out from a bus bay, in the same direction as the test driver. Speed around 80 km/h.
4. A bus drives out from a bus bay, in the opposite direction from the test driver. Speed around 80 km/h.
5. A car in front starts overtaking, after the test driver has started overtaking the same car. The test car has to yield to avoid a critical situation. Speed around 70 km/h.
6. A bus drives out from a bus bay, in the opposite direction as the test driver. Speed level around 80 km/h. (Similar to situation 4).
7. A car driving in the opposite direction turns left at a T-junction in front of the test driver. The test driver has to brake to avoid a critical situation. Speed around 70 km/h.
8. A car travelling in the opposite direction overtakes another car and approaches in the wrong lane of travel, towards the test vehicle. The test car has to brake and steer away to avoid a head-on collision. Speed around 80 km/h.

Fig. 40. Front projection in the driving simulator of situation 4.

Fig. 41. Front projection in the driving simulator of situation 7.

Fig. 42. Front projection in the driving simulator of situation 8.

Fig. 43. Video image showing what happened during the reaction time test drive.

Each test subject was recorded on DVD and the data were recorded in a data file. The data file contained more data than it was possible to show on the video image. When the experiments were designed, the plan was to indicate with a label that the start of the incident that should have lead to a reaction had begun. But this label could not be used because it was not always possible to detect the incident start as initiated by the test subject. The incident start therefore had to be manually recorded from the DVD video, which in turn meant that the incident stop (start using break pedal) also had to be recorded from the video image. The data record was checked to see if the steering wheel had been used in an avoidance manoeuvre before the use of braking pedal. This allowed greater control over each incident and reduced the threat to validity (instrumentation) by using the same measurement method for both the start and the stop of the incident.

Fig. 43 shows a picture from the video film used for analyzing data. The video image was divided into four frames:

- The traffic situation, i.e. the front projection in the driving simulator.
- The face of the test driver.
- The instrument panel of the car.
- Documentation of the behaviour of the test driver.

The example Fig. 43 shows the following information about the behaviour of the test driver:

- The subject has reached metre mark 22479 of the road.
- The test driver ID is 104.
- The speed at the moment is 82.6 km/h.
- The pressure on the brake pedal is 0.65 (on a scale between 0 and 1).
- There is no pressure on the throttle (0.00 on a scale between 0 and 1).
- The last incident was situation 16.

The reaction time was defined as the time between the occurrence of the specific traffic situation and the activation of the brake pedal by the test driver. Normally the test driver both releases the throttle and activates the brake pedal in the reaction time, but in some cases the test driver released the throttle before the traffic situation occurs.

Two different sets of experiments were conducted. The driving simulator, the physical characteristics and the simulated traffic were the same in both sets. The only differences were the test subjects themselves and their introduction to the experiment:

- Driving simulator experiment 1: In all 31 test subjects drove the fixed routes in the driving simulator. These participants were taken from the NTNU/SINTEF pool of experienced driving simulator drivers. There were 8 women and 23 men. The oldest individual was 58 years, and only 5 individuals were older than 50. The test subjects were told that the purpose of the experiment was to study the road design. They were told to drive as they usually would drive.
- Driving simulator experiment 2: This test was not planned as a part of the project originally, but was made available as part of a new students' introduction to the university. These tests were done after the initial tests. The subjects were not given any introduction as to how they were expected to drive. While all the participants in the SINTEF research project drove back and forth on the road, the students drove only one way. In all there were 160 students who drove in the simulator, but only 60 students were selected for the data collection. Those students who obviously did not drive as they normally would were extracted from the data set; in particular, these were students who drove with a mean speed of more than 110 km/h.

5.4.2 Field trials

Reaction times were observed at three junctions as a supplement and as a basis for the validation of results from the driving simulator. The junctions are located in Trondheim:

- Kongsvegen/Saupstadringen,
- Fjellseterveien/Gamle Oslovei, and
- Haakon VII's gate/Leangen allé.

There are traffic lights at all the junctions. Reaction times were measured in connection with braking when the traffic light changes to red and the release of the brakes when the traffic signal turns green. The junctions and approaches where the measurements were made are shown in Fig. 44 - Fig. 46.

There were also attempts to make measurements related to braking when catching up with a queue. It turned out to be difficult to observe this situation from a fixed point within an acceptable time limit. These measurements were not completed.

The measurements were done using video recordings of the junctions. The video recordings were taken from the same position as the positions where the photos in Fig. 44 - Fig. 46 were taken. From these sites, it was possible to both record the traffic signal and the use of braking lights. In all, 6 hours of recordings were done.

The measurements of reaction times for a signal change from green to red at the junctions were conducted using the following method:

1. The first vehicle approaching the junction after the traffic signal changed was recorded. Only vehicles perceived by the person making the measurements, to be so close to the junction that one could expect an immediate response to the change of traffic lights were recorded. Only one person was used in making measurements in order to eliminate the effect of different perception of different recording personnel (instrumentation) might have on selecting situations. Vehicles further away were ignored. Vehicles that had used their brakes before the signal change were ignored.
2. Vehicles that passed the stop line at the junction were also ignored.
3. The reaction time was measured as the time between when the light signal shifts from green to yellow until the braking lights were activated.

Measurements of reaction time for a signal change from red to green were conducted in the following way:

1. Vehicles that stand at the stop line when the traffic signal changes from green to red were watched.
2. Only those vehicles that had activated the brake lights when the traffic signal changed from red light were watched. The places for measurements were chosen with one criterion: that there was a rise so that the drivers were encouraged to use their brakes.
3. The reaction time was measured as the time from when the traffic signal turned from red to red and yellow until the brake lights were deactivated.

Fig. 44. The junction at Kongsvegen/Saupstadringen.

Fig. 45. The junction at Fjellseterveien/Gamle Oslovei.

Fig. 46. The junction at Haakon VII's gate/Leangen allé

5.4.3 Earlier studies in the video simulator

The tests carried out in the video-based driving simulator have been described in detail in Section 2.3 . As there are ongoing new tests of drivers in the video-based driving simulator, the amount of data is therefore continually increasing. At the time of the data analysis, there were 258 different test drivers in the database. Fifty-eight of these are considered a control group with test subjects without any known dysfunctions. These drivers are labelled "reference people". The rest are labelled "patients". These drivers have a potential cognitive or perceptual dysfunction, such as brain damage or visual problems.

In all, test drivers have driven 501 scenarios with a total of 20385 responses.

If a test driver does not respond within 4 seconds after a symbol is shown on screen, the symbol is removed and an error is recorded for the reaction time. These measurements are not used in the calculation of reaction time. The "reference person" had a mean of 4% errors, while the patients had a mean of 11% errors.

5.5 Results

5.5.1 Driving simulator experiment 1

Mean reaction time

Fig. 47 shows the mean measured reaction times and the 95% confidence interval of mean to the different traffic situations.

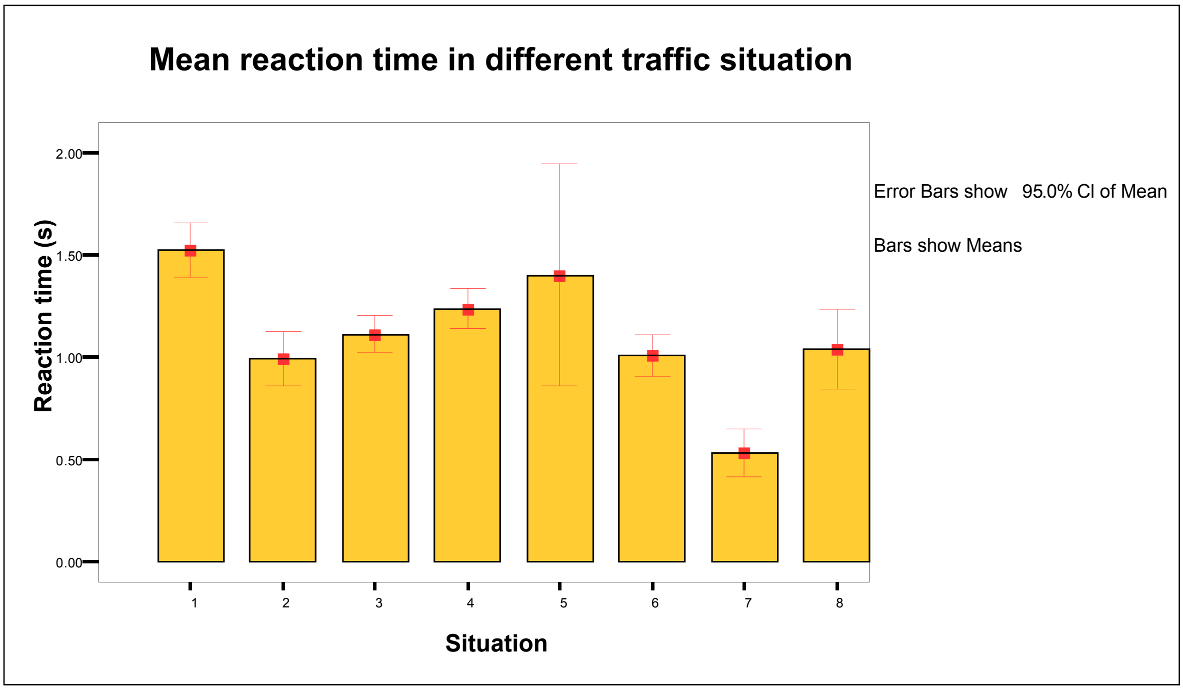


Fig. 47. Mean reaction time for different traffic situations.

The results from the measurements of reaction time are not fully compliant with normal distribution. This means that the reaction times do not have the same mean time and median time. Some tests, such as t-tests and ANOVA, work well even if the distribution is only approximately normal. Therefore, these tests have been used instead of relying on less powerful nonparametric tests.

The longest mean reaction time is seen in traffic situation 1, which is the first situation the test subjects encountered during the experiments.

In traffic situation 2, test drivers followed a lorry that was driving relatively slowly. The data showed that several test drivers had already released the throttle before the braking situation occurred.

The reaction time was relatively long in traffic situation 5, where the test driver had started overtaking and was accelerating, and suddenly he/she had to brake because the vehicle in front had also started overtaking. Situation 5 had the largest confidence interval. Notice that situations 5 and 6 had few measured situations, as is shown in Fig. 49.

The shortest mean reaction time took place in traffic situation 7, where one car on the main road and one car on the side road were waiting to make left turns. Just as in situation 2, most of the test drivers had also released the throttle when the traffic situation occurred. The speed limit on the main road was 80 km/h, but most of the test drivers had a speed around 70 km/h.

In total, the mean reaction time varied from 0.53 to 1.53 seconds in the different traffic situations. The maximum individual reaction time was 2.4 seconds and occurred in traffic situation 1. The minimum measured individual reaction time was 0.15 seconds in traffic situation 7, but here the test driver had released the throttle before the traffic situation occurred.

There were no indications of differences in reaction time by gender. There was a slight indication that the reaction time was shorter among drivers between 40 and 58 years, but the result was not significant.

Situation 1 had a higher reaction time than the rest of the situations except situation 5. The difference of the mean is p<0.005 in all other situations.

85th percentile reaction time
Fig. 48 shows the 85th percentile of the reaction time in different traffic situations.

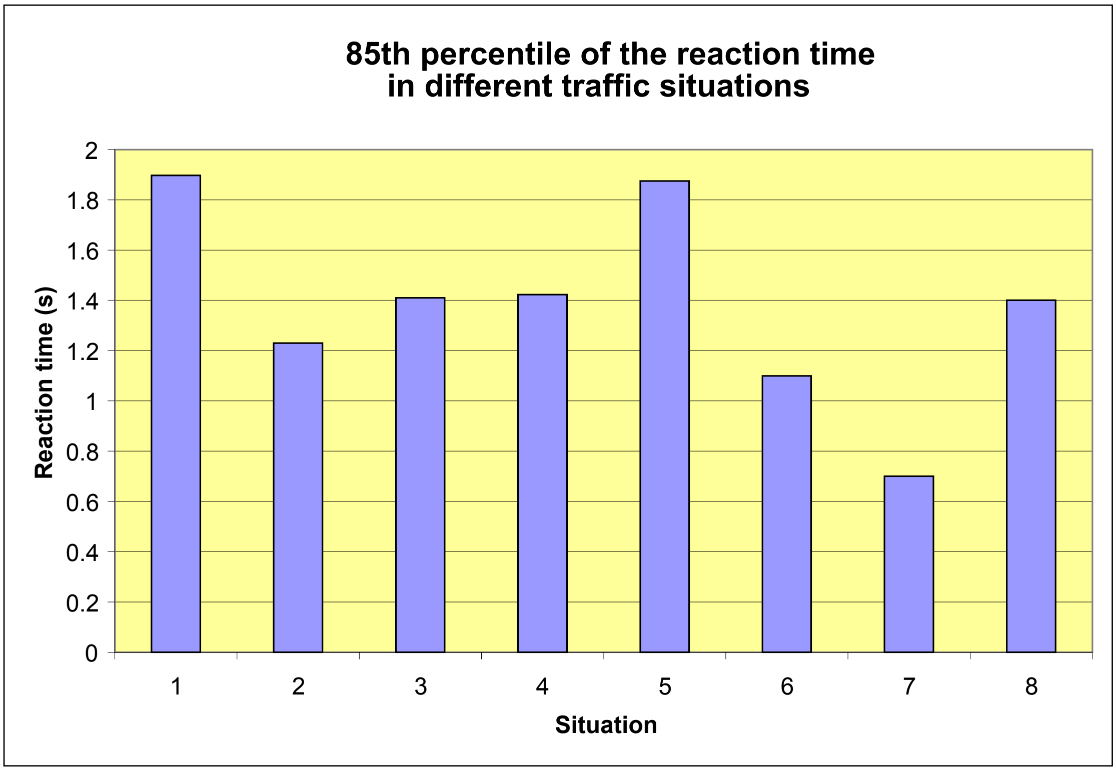


Fig. 48. 85th percentile of the reaction time to different traffic situations.

In total, the 85th percentile varied from 0.7 to 1.9 seconds in the different traffic situations.

5.5.2 Driving simulator experiment 2

There were more missing values for reaction times from driving simulator experiment 2, which involved students, than in driving simulator experiment 1.

Fig. 49. Percentage of drivers for which a reaction time was recorded.

The reaction time recorded in experiment 2 differed from those recorded in experiment 1, but it was not possible to tell if it generally was higher or lower. The results for the four sites with the highest numbers of measured reaction times are presented in Fig. 50.

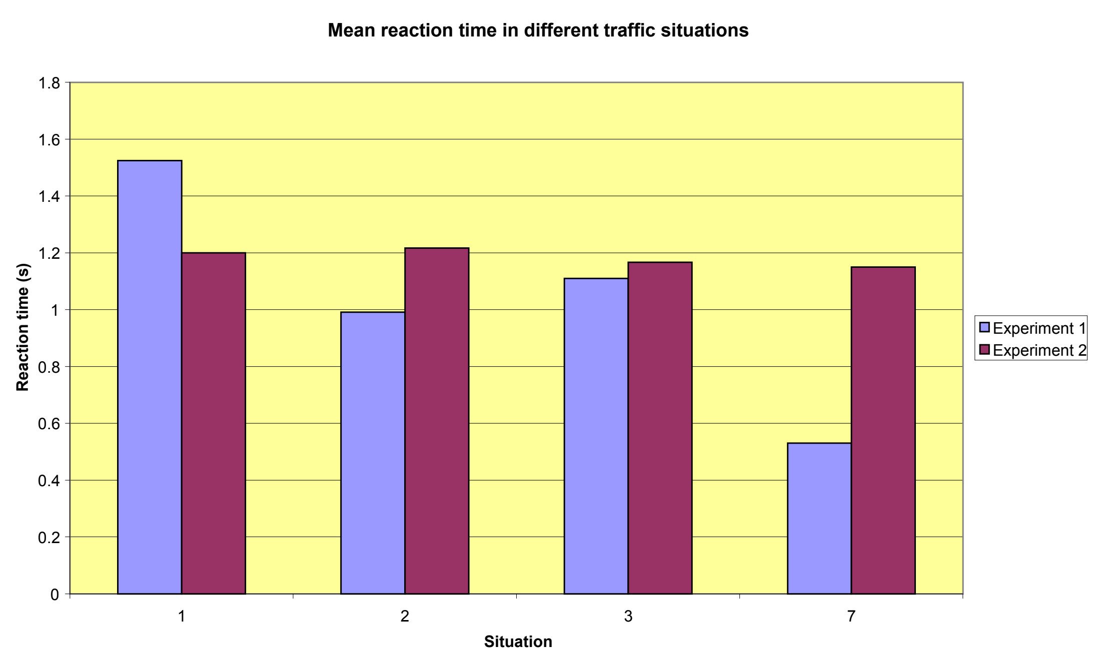


Fig. 50. Mean reaction times in different traffic situations.

5.5.3 Field observations

There were 41 measurements of reaction time when the traffic signal shifted from green to yellow. The mean reaction time at the three junctions varied between 1.2 and 1.3 seconds.

All together there were 214 measurements of reaction time when the traffic signal shifted from red to green. The mean reaction time varied between 0.9 and 1.0 seconds at the three junctions.

Fig. 51 and Fig. 52 show the distribution of measured reaction times at the three junctions. The 85th percentile for all three junctions together was 1.6 seconds when the traffic signal changed from green to red while it was 1.4 seconds when the traffic signal changed from red to green.

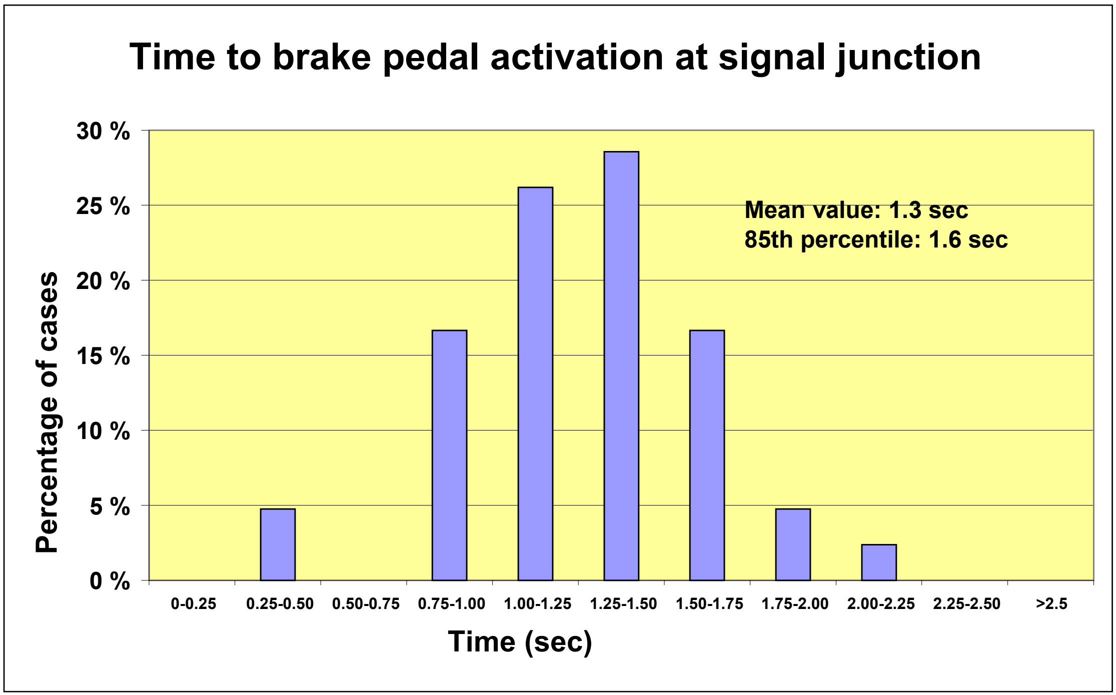


Fig. 51. Reaction time when the traffic signal turned from green to red

The reaction times when the traffic lights changed from red to green were shorter than the reaction times when the traffic lights changed from green to red.

There were situations in field research where the reaction time was longer than 2.5 seconds; this occurred when the lights changed from red to green.

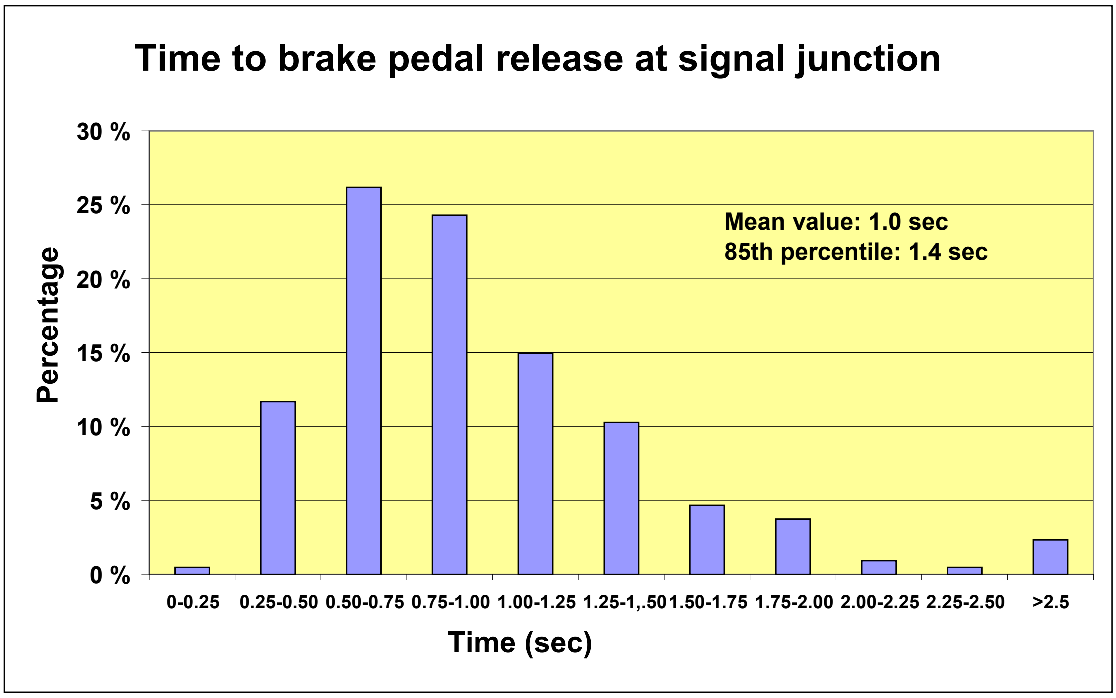


Fig. 52. Reaction time when the traffic signal turned from red to green.

5.5.4 Earlier studies in the NTNU/SINTEF video simulator

The mean reaction time was lower than both the graphical driving simulator and the measurements in junctions. The reaction time is shown in Fig. 53.

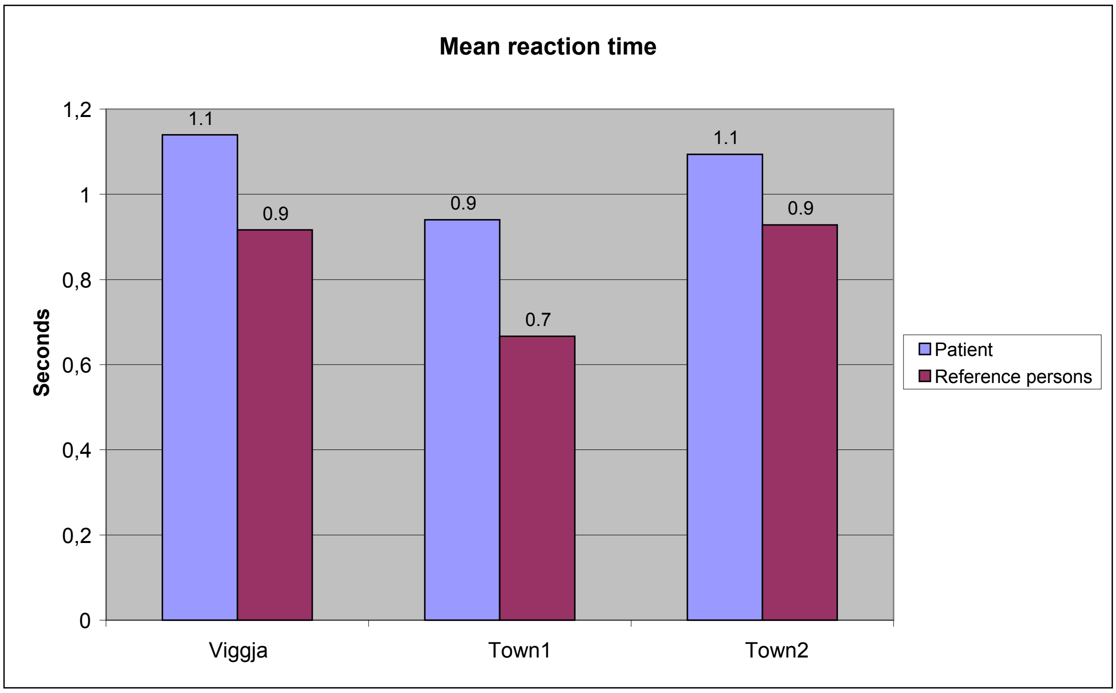


Fig. 53. Mean reaction time in the video-based driving simulator.

The data also contain information about gender, age and type of environment. None of these variables had a significant effect ($p<0.05$) on the reaction time.

The mean reaction time for the reference test subjects was less than 1.0 second for all scenarios. The patients had a mean reaction time between 0.2 and 0.3 seconds higher than the reference subjects.

The reference test drivers did not respond to incidents within 4 seconds in 4% of the cases.

In our study we have focused on a normal situation, and the 85th percentile related to all drivers, not just reference subjects. The results for the patients may be regarded as a maximum reaction time for people on the verge of losing their licence due to brain damage and visual defects. Fig. 54 shows the 85th percentile of reaction time for the three scenarios.

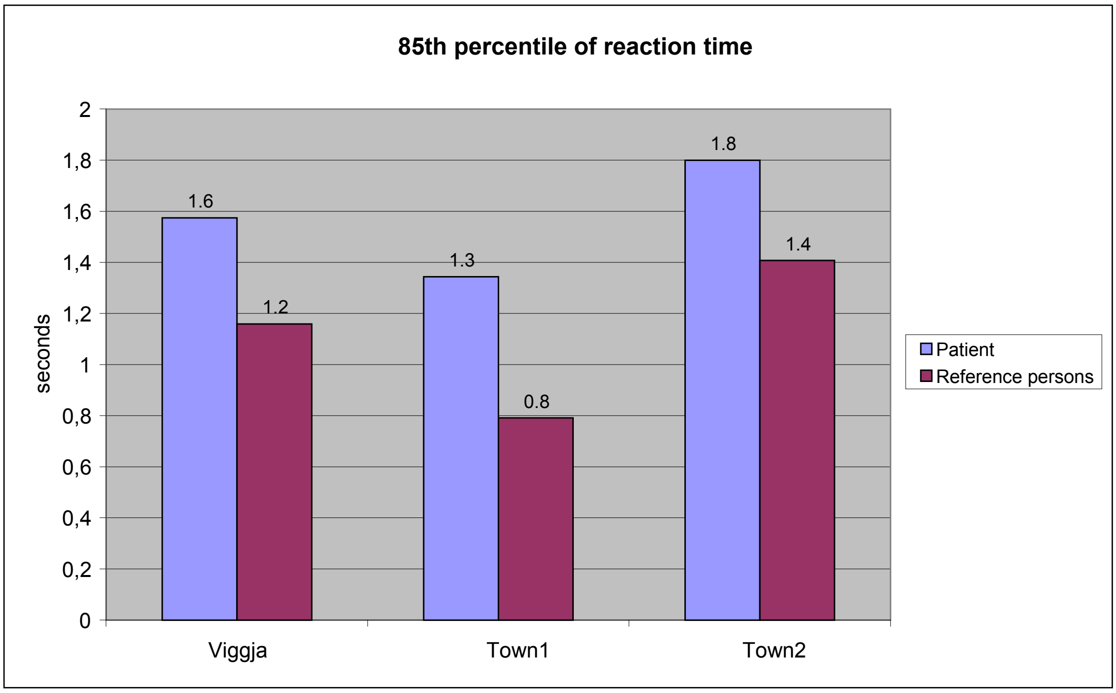


Fig. 54. 85th percentile of reaction time in the video-based driving simulator.

The 85th percentile was between 0.1 and 0.5 seconds above the mean reaction time.

5.6 Discussion

5.6.1 Design reaction time

This discussion about design reaction time relies on experiment 1 in the driving simulator, results from the field observations and the video-based simulator and the literature review. Because the purpose of experiment 2 in the driving simulator was to study the effect of not having the same standard introduction related to test subjects as a normal experiment, the results from that experiment have been omitted from this discussion.

Mean reaction time

In the driving simulator experiment 1 the mean reaction time varied from 0.53 to 1.53 seconds in the different traffic situations, but except for one situation, all the mean reaction times were above 1.0 seconds. Thus it is possible to state that the mean reaction time was between 1.0 and 1.5 seconds in most of the situations.

The mean reaction time of between 1.0 and 1.5 seconds is also supported by the literature review and the field studies at the junctions. The exception was that the video driving simulator results showed mean reaction times in the different situations of between 0.7 and 0.9 seconds.

Maximum reaction time

There are larger differences related to maximum reaction times between the data sources. In the video-based driving simulator, the normal test drivers did not respond to incidents within 4 seconds in 4% of the cases. This is probably because they did not see the stimuli.

In the field observations, there were situations where the reaction time was longer than 2.5 seconds, but this was only when the signal changed from red to green. This was probably because the drivers were not attentive and did not consider the situation dangerous.

The maximum individual reaction time measured in driving experiment 1 was 2.4 seconds and occurred in traffic situation 1. All 85th percentile values were below 2.0 seconds.

In the video simulator, the 85th percentile is between 0.1 and 0.5 seconds above the mean reaction time. In driving simulator experiment 1, the 85th percentile was between 0.3 and 0.6 seconds higher than the mean reaction time. This seems to correspond with results in the literature, where both the mean reaction time and the 85th percentile have been presented. In the field observations, the 85th percentile was between 0.3 and 0.4 seconds higher.

Expecting an incident

The longest reaction time can be found in situation 1. The drivers saw a vehicle that was waiting to enter the road at a junction. This was the first situation the subjects were introduced to and most likely the most unexpected. A similar situation was presented in situation 7, but here most of the test subjects were aware that something might happen, and the result was the shortest mean reaction time. The speed limit on the main road was 80 km/h, but most of the test drivers drove around 70 km/h. Most of the test drivers had also released the throttle when the traffic situation occurred, probably in a preparation for a possible incident.

In situations where the driver expected an incident and the reaction can be measured by the first movement, the mean reaction time could be as low as 0.5 seconds. This finding is supported by the literature review and the following points:

- In situation 7 in the simulator studies, the test driver saw two vehicles waiting in a junction. The test drivers seemed to prepare for an incident. Some drivers also released the throttle and most lowered their speed.
- In the video simulator where one of the main tasks was to react as quickly as possible to stimuli, the reaction time was generally lower than for the other registration methods.

The minimum measured individual reaction time measured in the driving simulator experiment 1 was 0.15 seconds in traffic situation 7, but here the test driver had decided to brake before the traffic situation occurred.

In traffic situation 5, the reaction time was relatively long. The test driver had started overtaking and was focused on accelerating the vehicle, and suddenly he/she had to brake because the vehicle in front had also started overtaking. The longer reaction time can be explained by a situation that was more unexpected.

Complex situations and alertness

The 85th percentile seemed to be affected more when scenarios were more complex. This can be seen in the video simulator, where the increase in reaction time was larger in town 1 and town 2 situation. This means that more drivers will have a longer reaction time when they are in a complex driving situation.

The video-based reaction time when the traffic lights changed from red to green was shorter than the reaction time when the traffic lights changed from green to red. This was mainly because we recorded the use of the brake pedal. When the traffic signal turned from green to red, the driver had to release the throttle and activate the brake pedal, while the driver only had to release the brake pedal when the traffic signal shifted from red to green.

The Viggja and the town 1 scenarios had about the same complexity regarding stimuli. In the Viggja scenario there was a “choice reaction” while there was a “simple reaction” in the town 1 scenario. The reaction type was also simpler in the town 1 scenario where the test drivers only had to use their voice, while the test drivers in Viggja had to push a button. These two issues probably account for the difference of 0.2 seconds in mean reaction time between the two scenarios.

While the response to stimuli was the same in the town 1 and town 2 scenarios, the Town 2 scenario was more complex with more possible positions of the stimulus on the screen. The stimuli were also more difficult to notice. This is probably the reason for the 0.2-second difference in reaction time between the two scenarios.

Generalization of the results
There are many factors that may influence reaction times. We can see this in both the driving simulator studies and in the studies from the video-based driving simulator. This can lead to the conclusion that there should be different design reaction times in different traffic situations. For instance, varying design reaction times are used for different road standards in Sweden

One can also assess the possibility of using different design reaction time in towns and on rural roads, but it is difficult to see a relationship between reaction type and area type. Theories have been put forward that drivers are more alert in towns than on rural roads and therefore have lower a reaction time in urban traffic. At the same time the traffic situation is more complex in towns, which might lead to longer reaction times. This work found no basis for making the assumption that there is a different reaction type depending on the area type.

We have not conducted any measurements using a graphical driving simulator in a town situation. It is possible that more extensive research could show a relationship between different traffic situations and reaction times.

5.6.2 The validity of using the driving simulator as a research tool for reaction times

This section discusses research on reaction time with respect to key threats presented in section 4.5 .

Statistical conclusion validity
All the data sources used included a maximum measured reaction time. This results in a restriction of range. The exact maximum was different in different situations. The maximum may even be influenced by the construct ("Treatment sensitive factorial structure"**)**. In the video simulator, the maximum was explicitly set to 4.0 seconds. This threat is not restricted to the driving simulator, but is also very important in other methods of data collection related to reaction time.

In situation 5 in the driving simulator experiment 1, there is a large deviation in the results, with extraneous variance in the experimental setting. One probable cause is the problem of determining the exact time of the start of the incident and which action to take. This also leads to a high level of missing data and a wide confidence interval of the mean. Determining a significant different mean reaction time for this situation therefore became impossible.

Internal validity
Perfect randomization is not feasible in the driving simulator experiments. In driving simulator experiment 1, the results were tested to see if gender and age had a statistical effect, but no significant differences were found. This threat to validity was therefore reduced in driving simulator experiment 1.

There are a lot of missing values in both the graphical driving simulator experiments and in the video-based simulator. The field observations would probably also have similar missing values, but the incidents could not be controlled in the same manner as in the simulators, and thus information about missing values could not be collected. This threat to validity is closely related to the threat of loss of responds (attrition effect).

The results from experiment 2 in the driving simulator resulted in more missing reaction time values. The most probable cause for this was that drivers did not drive as they would have in the real world. Since the scenarios assumed that drivers would drive as they would in the real world, tests of students did not result in the anticipated results.

It is very difficult to design an experiment related to reaction time where the test subject do not learn the purpose of the experiment relatively quickly. The first situations would naturally be more unexpected and thus result in a longer reaction time. The reaction times in situation 1 were significantly longer than all but one other situation. This is related to the threat of testing the testing itself.

Construct validity
This research has meant using different methods to determine reaction times. The different sources resulted in similar data about reaction times. This reduces the level of mono-method bias, and demonstrates the advantage of using driving simulator experiments as a new method of observational studies.

The difference between mean reaction time and the 85th percentile is not always constant. This might be because complex situations result in larger standard deviations, along with the skewness of the distribution. The use of mean reaction time with the addition of a safety factor would therefore not give the same result as using the 85th percentile. Using different construct levels would therefore reduce the threat of confounding constructs with construct levels.

When creating situations that would lead to a reaction in the driving simulator experiments, the main concern was the initiation of the reaction. Sometime the test subjects had not reacted to the incident until it was too late. This often occurred related to speeding. This can be seen in experiment 2 in the driving simulator. Instead of recording a high reaction time, the value was recorded as missing. This leads to a threat to treatment sensitive factorial structure.

External validity
We have tried to generalize the results of the driving simulator study to town situations by examining the complexity and expectations in encountering situations. We did not find that it would be possible to draw any conclusions about this.

When comparing the data from the driving simulator with expected outcomes from rural roads we get similar results in certain situations. It seems that the driving simulator results are more influenced by the sequence of incidents than by the capabilities of the driving simulator as a research tool.

The use of the driving simulator can be viewed as a research setting. Using different sources of data it was found that the results from driving simulator experiment 1 could be considered similar to those found in real life observations, the video simulator and in the literature review. The driving simulator for reaction time studies can therefore be recommended for future use in this type of research.

5.7 Conclusion

5.7.1 Recommended value for reaction time

Reaction time varies from situation to situation and between drivers. Because there is no upper limit to reaction time, it is difficult to use maximum reaction time as a design reaction time. Based on our survey, we could not differentiate the reaction time based on different variables such as place, situation or other characteristics.

The mean reaction time was between 1.0 and 1.5 seconds in most situations. This is supported by the literature review, experiment 1 in the driving simulator studies, and in the field studies at the junctions.

If the design reaction time is to be based on the most unexpected situation, the reaction time should not be set lower than 2.0 seconds, based on our observations.

In our report we recommended using the 85th percentile as a basis for the design reaction time. Results from the literature review, field studies and driving simulator experiment 1 all seem to place the 85th percentile reaction time between 1.5 and 2.0 seconds. To include as many situations as possible, 2.0 seconds should be used as a design reaction time. This is the same design reaction time as is used today.

5.7.2 Use of the driving simulator

The results from the project have a greater validity because we used four different sources of information for reaction times. The validity of using driving simulators for this type of research has also improved. This requires that the driving simulator research be conducted in a manner similar to experiment 1 and not experiment 2. This means that the test subjects are focused on driving normally.

6. Validation case 2 – Speed and lateral position

6.1 Background

The main purpose of this chapter is to compare roadside measurements and data from a driving simulator. Speed is easy to measure both in the real world and in a driving simulator. Lateral position is a little more difficult to measure in real life, but is still relatively simple to measure. Section 4.3 describes many research projects where speed and lateral position were used to compare real world data and driving simulator data.

Several SINTEF projects have studied the effect of road, lane and shoulder width on speed and lateral position. Information about these projects can be found in (Sakshaug et al. 2004), (Giæver and Engen 2005a), (Giæver and Engen 2005b), and (Giæver and Engen 2005c). Data from these projects are described and further analysed in this chapter. In addition, extra roadside measurements have been done.

6.2 What is true speed and lateral position?

A vehicle's mean speed and lateral position are influenced by the road geometry. When comparing measurements from real life and the driving simulator, the emphasis is often on having comparable road geometry. This is of course important, but speed and lateral position are also influenced by other factors that change over time as well.

Much information is available about factors that influence speed. There is not as much research on lateral position as on speed. I will therefore focus my general discussion on speed measurements.

The mean speed of vehicles is influenced by many factors such as weather, road width, asphalt colour, time of day, and traffic volume.

The mean speed even changes over time. Fig. 55 shows how the mean speed has changed from 1993 to 2002 on the main road network in Norway, partly because of the above -mentioned factors, and partly because of vehicle developments.

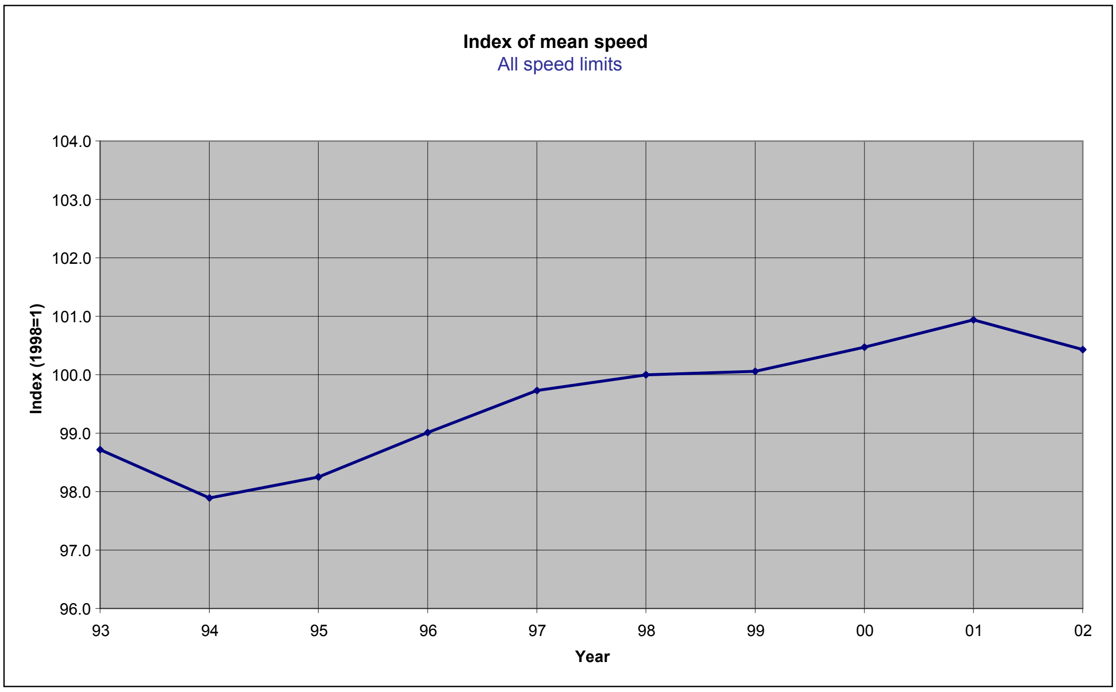


Fig. 55. Index of mean speed in Norway (1998 is 100).

These changes and their effects on mean speed make it difficult to state one “true” mean speed of a certain spot.

Often characteristics such as asphalt colour, time of day and weather are treated as constants in driving simulator studies. Some can be tested, such as weather and lighting, but often not much attention is paid to these characteristics. This makes it very difficult to know exactly what real world speed you would like to compare to the measured speed in the driving simulator.

Mean speed together with traffic density and traffic volume are very important in a macro view of traffic flow. These three factors are closely related. This means that the measured mean speed is influenced by traffic volume and traffic density.

6.3 Literature on speed and lateral measurements in driving simulators

The use of a driving simulator makes it relatively inexpensive to test new regulations and new road designs. Some of these studies are presented here.

(Ihs 2006) used a driving simulator to study the effect of centre line and guideposts on traffic safety, accessibility and comfort. The effect of removing the centre line on roads that had a width of less than 6.5 metres was studied. A road with a width of 8.0 metres was used to study the effects of guideposts. Speed and lateral position were measured in both situations. On the road section that was 8.0 metres wide, the speed increase was 2.0 km/h when guide posts were used and the road markings met requirements for visibility. When the road markings were poor, the speed increase was 10 km/h. On the 6.5-metre wide road, the speed was 6.0-km/h higher when there was a centre line than without a centre line.

(Horberry et al. 2006a) studied the possible safety benefits of enhanced road markings compared to standard road markings. He compared the relative effectiveness under simulated wet night driving conditions of commonly used highway markings with enhanced road-marking system with improved retro-reflectivity. Participants completed drives with both types of markings. A secondary mental arithmetic task was also employed in half of the conditions. Both objective and subjective data were collected. The results indicated that participants were better able to maintain lane position and speed with the enhanced markings than with the standard markings. A similar pattern was found for the subjective measures: workload was rated as lower for the test with enhanced markings; likewise, subjects reported the drives as being easier and were more confident in being able to drive safely when the roads displayed the enhanced markings.

(Anund et al. 2005) used a driving simulator to test placement and design of milled rumble strips on the centre line and the shoulder marking. Four different physical designs of milled rumble strips and two shoulder placements were tested on a road 9.0 metres wide. Driving behaviour, measured as lateral position, speed, and steering wheel angle, were recorded. The researchers also recorded physiological data such as brain activity (EEG), eye activity (EOG) and muscle activity (EMG). Additionally, the drivers were asked to rate their level of sleepiness every 5 minutes by using the Karolinska Sleepiness Scale (KSS). The participants in the experiment were regular shift workers driving during morning hours after a full night shift. Speed and lateral position were used as basic measurements. It was concluded that that rumble strips had a clear alert effect and induced correct averting actions. There seemed to be no risk associated with using rumble strips.

(Helmers and Törnros 2006) studied the effect of clothoids on driver safety margins on speed and lateral acceleration. The study included a test to see if base training in a driving simulator could influence the driving performance. The research found that drivers had higher speeds in curves with clothoids, but there were no difference in lateral acceleration, and training in the driving simulator had no effect on driving performance.

(Törnros and Wallman 2003) studied the effect of pavement ruts on driving performance. This was a pilot study and the purpose was twofold: to examine if the driving simulator was suited for this kind of study, and to describe to what extent the road surface influenced driving behaviour in the simulator. The ruts were on a level with the surrounding road surface, but the noise and vibration was increased. There were no differences in mean speed, but the variance was greater for the rutted road. Mean lateral position differed very little, but the subjects tended to avoid driving in the ruts. It was concluded that the driving simulator studies was promise and that further development would be of value.

(de Waard et al. 2004) studied the influence of visual road information. In an advanced driving simulator they explored the questions, “How much visual information from the road is required for proper driving?”, and “How do people cope with a visually ambiguous road configuration?”. During the rides, performance (lateral position, speed) and heart rate were recorded continuously, and before transition to a new section drivers gave a rating for invested effort and for visibility in the (previous) road course. The experiment’s goal was to determine whether a shift in driving behaviour could be noticed given a certain amount of visual information. The main threshold was found to be between roads with ‘no delineation on the road surface at all' and 'a centre-line'. Elderly drivers appeared to need the visual aid of the centre line to a greater extent than young drivers, and in general drove slower and regulated their information input this way. Results from the study also supported the assumption that with increasing age people are more easily confused by ambiguous cues.

(Comte and Jamson 2000) used a driving simulator to study speed-reducing measures for curves. This study developed and tested four speed-reducing methods (Variable Message Sign, in-car advice, speed limiter and transverse bars) against a baseline condition. Driver performance, workload and acceptability were evaluated. As would be expected, speed limiters were the most effective measure; however, in terms of user acceptability this system was least preferred. All the other measures significantly reduced speeds when activated (although not as effectively as the speed limiter), in the order of approximately 6 km/h. It appears that the provision of information or support, in any format, could be effective in reducing speed on curves.

6.4 Experiment design

6.4.1 Data sources

Experiments in the NTNU/SINTEF driving simulator to study the effects of road markings on speed and lateral position were conducted using the Støren – Soknedal road model described in Section 2.5 . The real road is 8.5 metres wide. The driving simulator model makes it possible to alter the road, lane, and shoulder width of the test section of road.

Two data sources were used for comparable data of real world driver behaviour. Data have been collected on the real world Støren – Soknedal road. In addition data have been collected at E6 Lillehammer, where the effects of changed road markings have been evaluated in real life.

The different sources of data are shown in Fig. 56. The roads in the simulator are labelled RS (Road-Simulator) and the roads in the real world are labelled RR (Road-Real). Lines between the data sources show the comparisons that can be made.

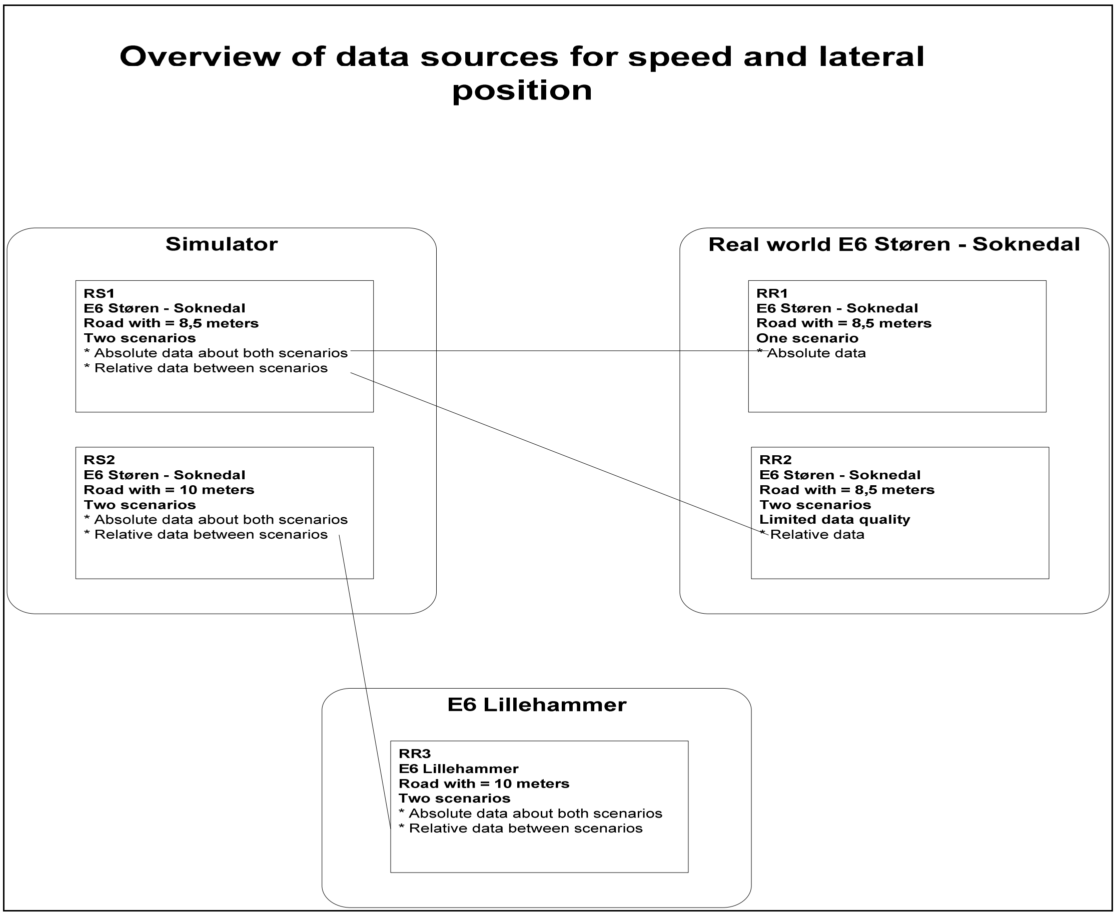


Fig. 56. Overview of data sources for speed and lateral position.

Driving simulator experiments were conducted on the standard road with 8.5 metre road width especially for this thesis. These measurements are labelled RS1 and an image of the situation is presented in Fig. 57.

The simulator study RS2 was a SINTEF project on the effects of visual lane division and the use of rumble stripes on 10 metre wide roads. The results were presented by (Giæver and Engen 2005a) and (Giæver and Engen 2005c). The 10 metre wide road was specially developed in the simulator for this project. The basic data files were available for the present work for further analysis. Fig. 58 shows the situation presented in the driving simulator.

Sakshaug et al. (2004) studied the effect of shoulder width and lane width on traffic safety on a real road that was 8.5 m wide. These measurements are labelled RR2. Of special interest in this case study are measurements of speed and lateral position when the road markings were changed. Measurements were made for the same road stretch that was visualized in the simulator. Unfortunately these measurements were of poor quality with many clearly incorrect measurements because of detector failure, and the data had to be extensively filtered to give reasonable results.

Supplementary measurements labelled RR1 were also made especially for this thesis on the real road, to extend the comparison between the real road and the simulator. RR1 and RR2 are shown in Fig. 61.

After the simulator studies were conducted, a real-world test of the 10 metre road with a visual division was implemented in Lillehammer. This study is labelled RR3. Images for the before and the after situation are presented in Fig. 59 and Fig. 60.

Fig. 57. RS1: E6 Støren – Soknedal, road width = 8.5 m.

Fig. 58. RS2: E6 Støren – Soknedal, road width = 10.0 m.

Fig. 59. RR3: E6 Lillehammer, before.

Fig. 60. RR3: E6 Lillehammer, after.

Fig. 61. RR1/RR2: E6 Støren – Soknedal.

6.4.2 Driving simulator scenarios

All the experiments were done on the Støren – Soknedal road, previously described in section 2.5.2 . There were four scenarios used with different road, lane and shoulder widths. These are described in Table 1. The scenarios are labelled SS1-SS4 (Simulator Scenarios).

Table 1 Driving simulator scenario description

Data source	Scenario Simulator	Description
RS1	SS1	Road width (RW) = 8.5 metres Lane width (LW) = 3.25 metres Shoulder width (SW) = 1.0 metre This is the standard scenario used in the simulator for ordinary experiments.
RS1	SS2	Road width (RW) = 8.5 metres Lane width (LW) = 3.0 metres Shoulder width (SW) = 1.25 metres This is a road with narrower lanes than ordinary.
RS2	SS3	Road width (RW) = 10 metres Lane width (LW) = 3.5 metres Shoulder width (SW) = 1.5 metre Central reserve width (CRW) = 0.0 metres This is the standard representation used in the simulator for 10 metre roads.
RS2	SS4	Road width (RW) = 10 metres Lane width (LW) = 3.5 metres Shoulder width (SW) = 1.0 metre Central reserve width (CRW) = 1.0 metre This is the new representation where a visual central median is employed.

In all 29 test subjects took part in the driving simulator experiment. Five test subjects were female and 24 subjects were male. Twelve test subjects were 40 years or older, but none were older than 55.

Both in RS1 and RS2, simulator studies were designed in such a way that the test subject drove along the road four times. Each subject drove either on the 10 metre road or the 8.5 metre road. They drove each scenario twice. The "priority" of the scenarios was randomly selected to eliminate any learning effects. All of the runs from the test subjects who drove the 10 metre wide road were used in the analysis. For the test subjects who drove the 8.5 metre wide road, two traffic situations were used. See section 7.4.1 for details of traffic situations. Only the two runs with free flow traffic are used in this chapter for Validation case 2 – Speed and lateral position, while two other runs were used in Chapter 7 Validation case 3 – Time gap.

Fifteen test subjects drove the 10 metre road and 14 others drove the 8.5 metre road. Each subject used approximately 8 minutes on the 10 to 11 km test road. The speed limit was 80 km/h.

6.4.3 Roadside measurements

As described in chapter 4, part of the road used in the driving simulator has already been built. The road is 8.5 metres wide. Before the road had its official opening, a temporary asphalt pavement was used. This pavement used the road markings from scenario SR2 in Table 2. Before the official opening, new pavement was laid and new road markings were made on the road. These road markings are the “normal” markings and are described as scenario SR 11. In Table 3 the different scenarios are presented with the scenarios labelled SR (Scenario Real road).

Table 2 Real road scenario description

	Scenario Real road	Description
RR1 and RR2	SR1	Road width (RW) = 8.5 metres Lane width (LW) = 3.25 metres Shoulder width (SW) = 1.0 metre This is the standard scenario used in the simulator for ordinary experiments.
RR2	SR2	Road width (RW) = 8.5 metres Lane width (LW) = 3.0 metres Shoulder width (SW) = 1.25 metres This is a road with narrower lanes than ordinary.
RR3	SR3	Road width (RW) = 10 metres Lane width (LW) = 3.5 metres Shoulder width (SW) = 1.5 metre Central reserve width (CRW) = 0.0 metre This is the standard representation used in the simulator for 10 metre roads.
RR3	SR4	Road width (RW) = 10 metres Lane width (LW) = 3.5 metres Shoulder width (SW) = 1.0 metre Central reserve width (CRW) = 1.0 metre This is the new representation where a visual central reserve or median is used.

All road measurements were done with a PTA (Portable Traffic Analyzer). This equipment can record both speed and lateral position by placing three wires over one lane. The images in Fig. 59 - Fig. 61 show the placement of the wires.

In all there were measurements at five points:

1. Left curve with radius 360 m.
2. Straight road.
3. Right curve with radius 1250 m.
4. Left curve with radius 500 m.
5. Right curve with radius 2500 m.

Measurements in RR1 were conducted at all five points, while measurements for RR2, described in Sakshaug et al. (2004), were conducted at points 4 and 5.

At the time of measurements for RR2 the detectors were of poor quality. Therefore the results had many errors and the results had to be filtered. These measurements were done on both scenarios SR1 and SR2 described in Table 2. The speed limit was 70 km/h at the time of measurements.

Since the time of the first measurements, SINTEF has acquired new detectors. New measurements were conducted. This time the measurements were only made in the same direction as the experiments in the driving simulator. In addition three new points were added. This time the speed limit was 80 km/h.

Lillehammer has a 10 metre wide road where a real-world test was conducted of visual lane division with road markings.

6.5 Results

6.5.1 Results of the simulator studies

On the 8.5 metre road, the mean velocity was higher in scenario SS 1 (LW = 3.25m, SW = 1.0m) than scenario SS2 (LW = 3.0m, SW = 1.25m). The difference in speed was between 1.2 km/h and 1.5 km/h at each point, but the difference was not statistically significant at the 95% confidence interval.

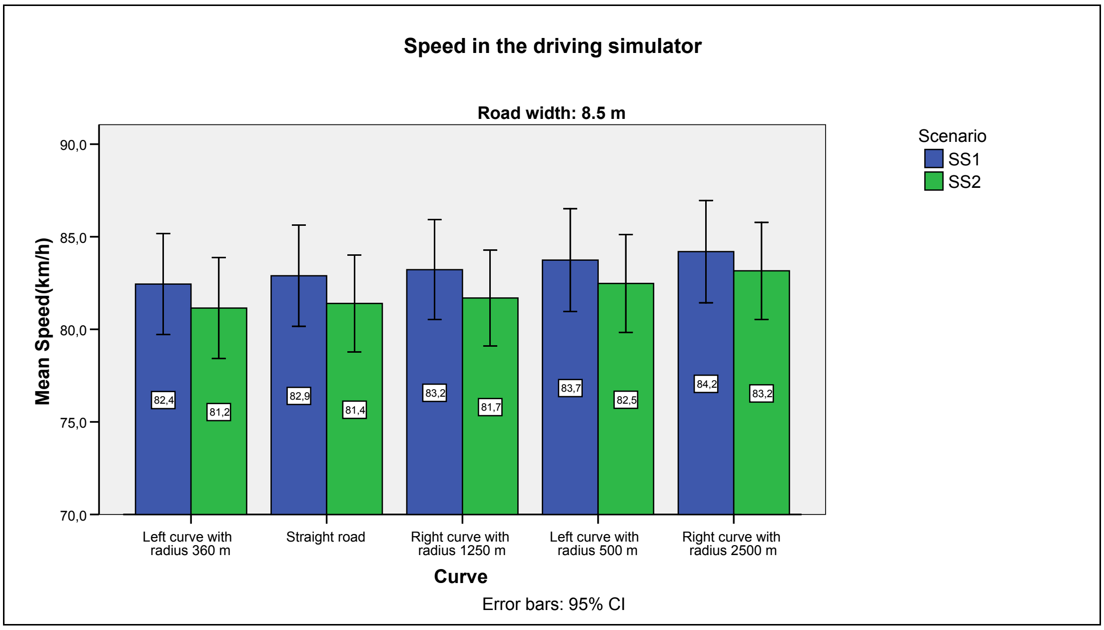


Fig. 62. Mean speed in the driving simulator on an 8.5 metre wide road

Fig. 63 shows the lateral position of the left wheel on an 8.5 metre road relative to the centre line. On roads with narrower lanes, the driver is closer to the centre line. The lanes width differs by 0.25 metres. The difference in lateral position is between 10 cm and 12 cm.

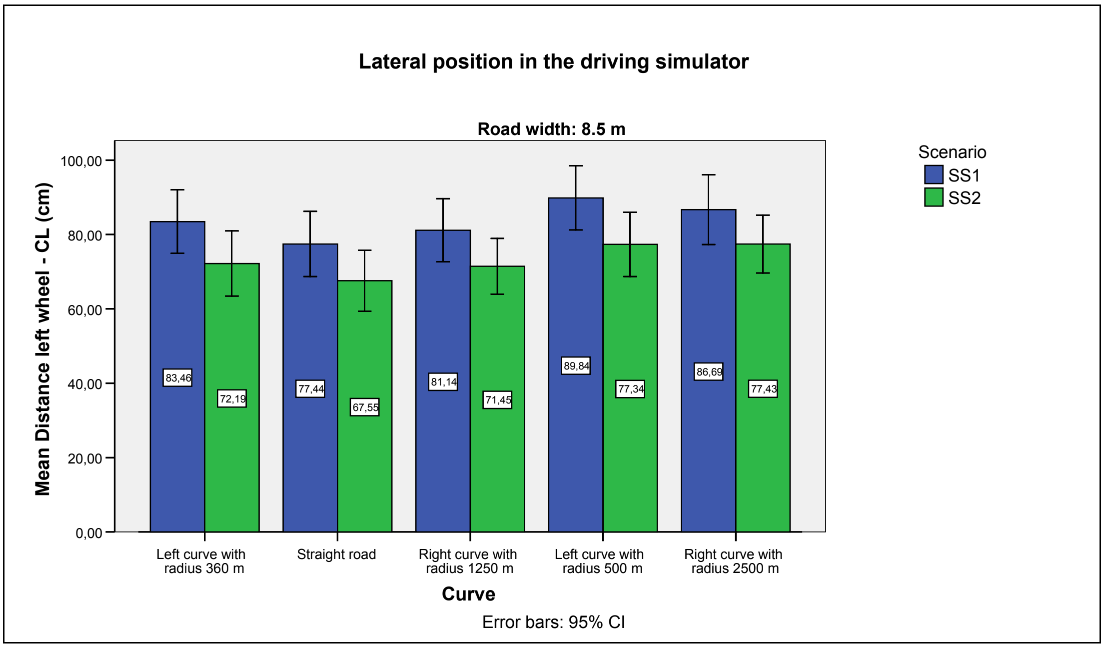


Fig. 63. Mean lateral position in the driving simulator on a 8.5 metre wide road.

On 10.0 metre wide roads, the speed on roads with a 1.0 metre visual central reserve (SS4) was lower than on the roads without such reserve (SS3), see Fig. 64. The mean speed was reduced between 0.2 km/h and 1.9 km/h.

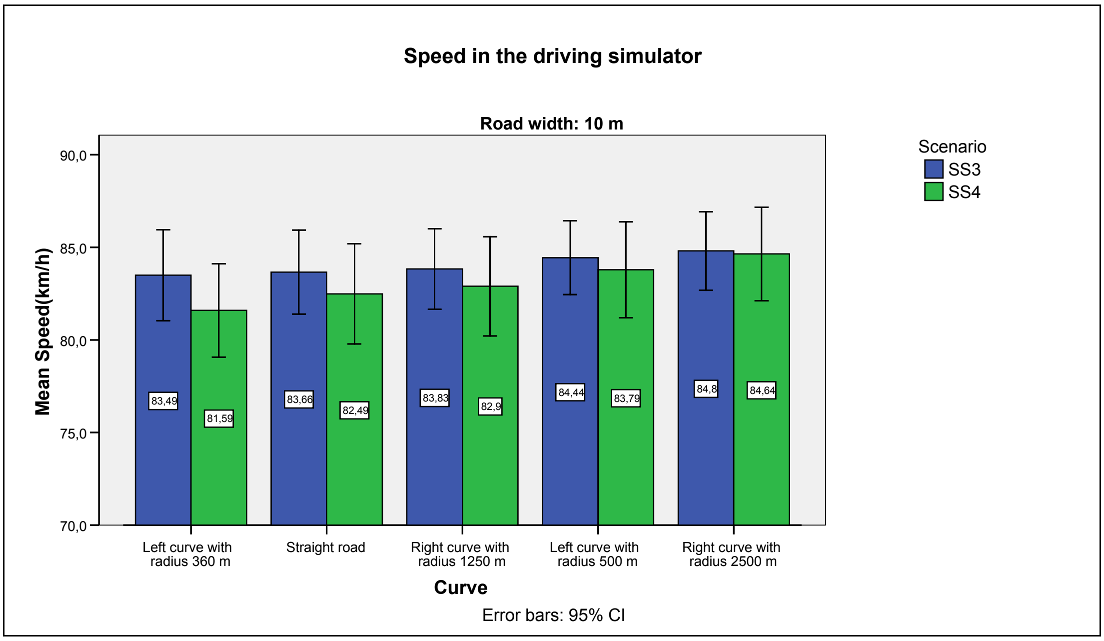


Fig. 64. Mean speed in the driving simulator on a 10 metre wide road.

The mean lateral positions on 10 metre wide roads differed significantly at the 95% confidence interval at all measurement points between the road with and without a visual central reserve. The road markings were “moved” 50 cm away from the centre of the road. The change in lateral position varied between 38 cm and 49 cm at the different observation sites. The results are shown in Fig. 65.

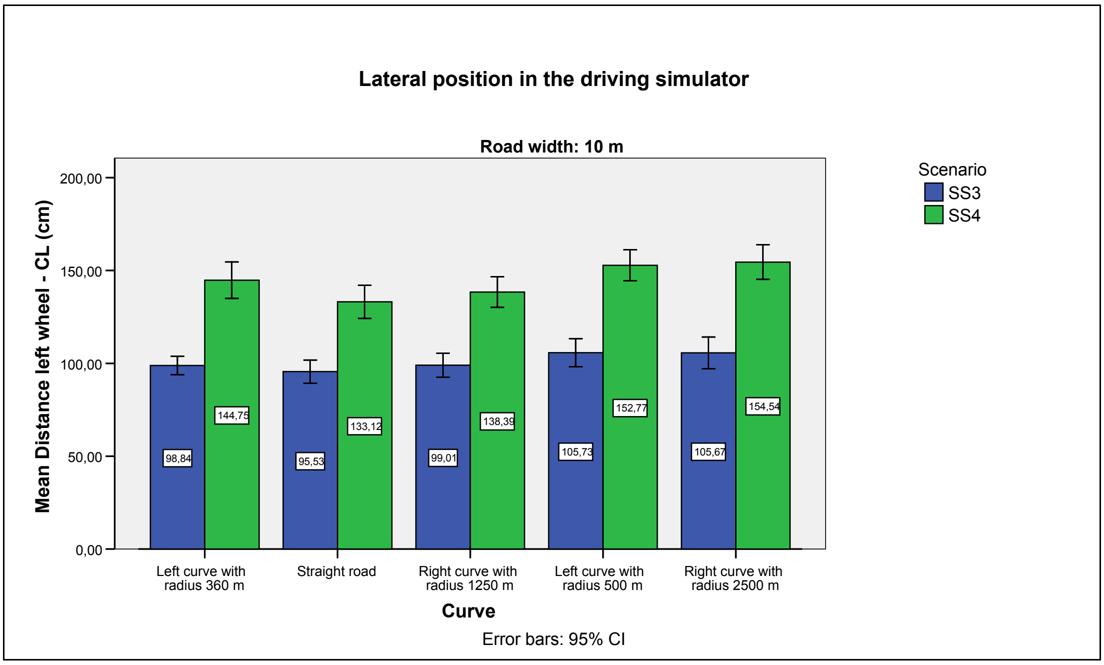


Fig. 65. Mean lateral position in the driving simulator on a 10 metre wide road.

6.5.2 Results of the real world studies at E6 Støren-Soknedal

Fig. 66 and Fig. 67 show the real world measurements (RR1) of speed and lateral position of cars and trucks. The vehicles used in the driving simulator were cars, and it is natural to compare the driving simulator data with measurements of cars in the real world. The difference in speed between cars and trucks did not show conclusively what type of vehicle was driven fastest. The difference of speed varied between cars that drove 1.1 km/h faster than trucks to trucks that drove 1.6 km/h faster than cars.

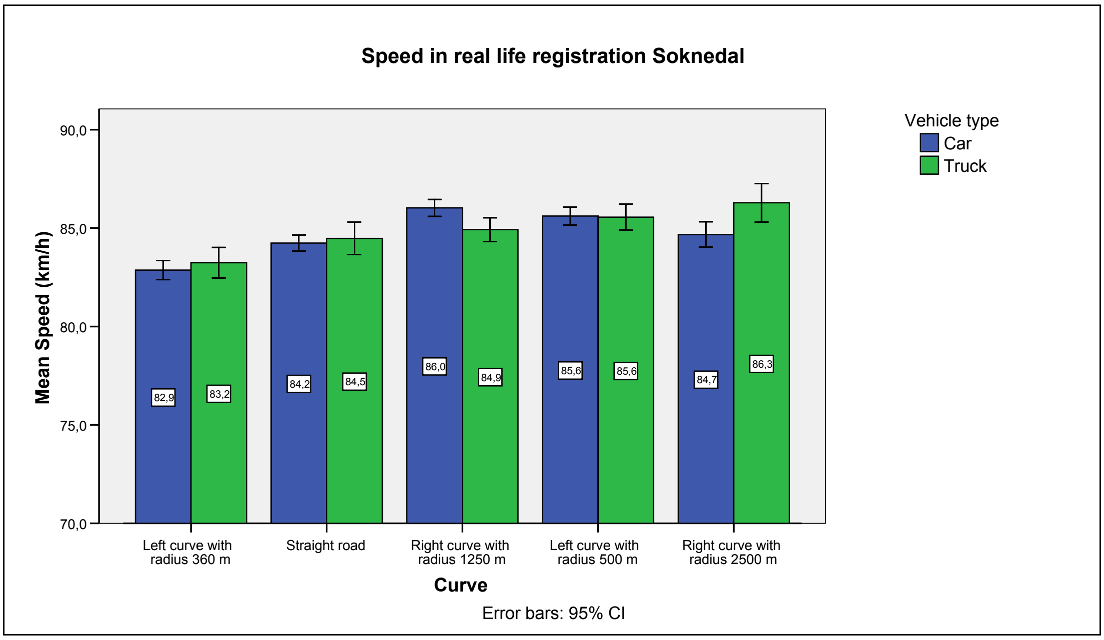


Fig. 66. Mean speed in real life on E6 Støren – Soknedal.

The slowest mean speed of cars measured at one single point was 82.9 km/h, while the fastest mean speed was 85.0 km/h. The speed measurements in the RS1 measurements in the simulator varied between 82.2 km/h and 84.2 km/h.

The difference in lateral position of the left wheel between cars and trucks was significant. The difference was between 19 cm and 34 cm. The lateral position of the left wheel related to the centre line of the car varied between 30 cm in a left turn and 97.40 in a right turn. This can be compared to the results in Fig. 63, where the lateral position measurements from RS1 in the simulator varied between 77 cm and 90 cm,

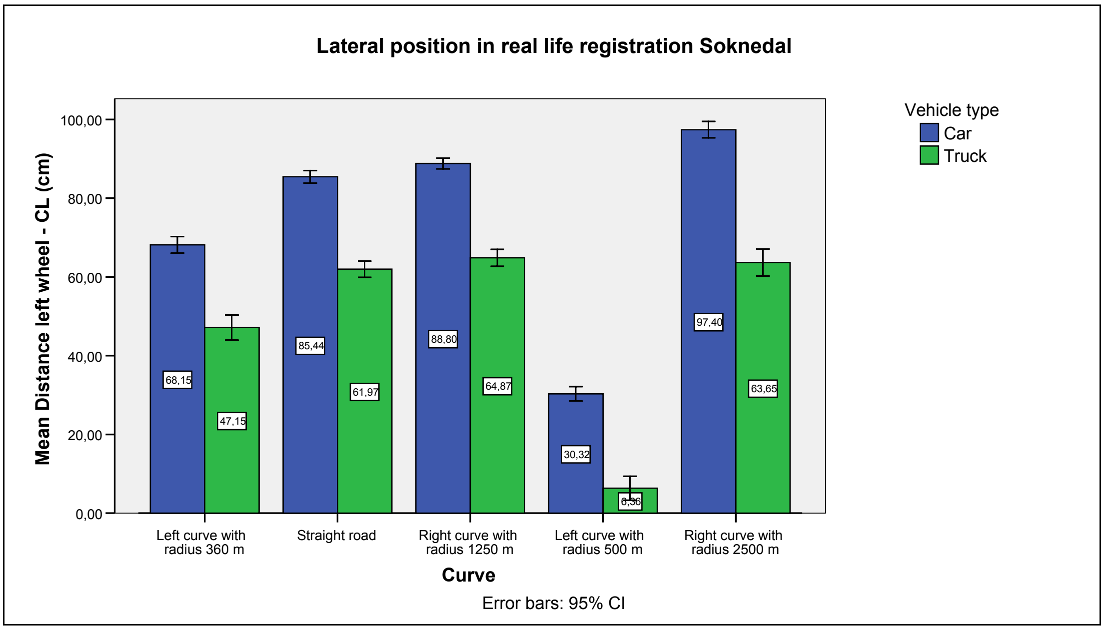


Fig. 67. Mean lateral position in real life on E6 Støren – Soknedal.

For the measurement of RR2 at the same places in E6 Støren – Soknedal, Sakshaug et al. (2004) found that the smaller lane width of 0.25 cm led to a change in lateral position of 14 cm measured from the right wheel and 23 cm measured from the left wheel. Sakshaug et al. (2004) had a large amount of data that were not used because it did not make sense. The end results must therefore be viewed with caution with respect to accuracy.

6.5.3 Results of the real world studies at Lillehammer

The introduction of a visual central reserve to the road in Lillehammer in 2005, allowed for before and after measurements of the real world effect of this measure. The exact properties of the road at the specified observation sites are described in Table 3.

Table 3: The Lillehammer road before and after installation of the visual central reserve.

Observation site	Road kilometre	Direction	Shoulder width		Lane width		Comment
			Before	After	Before	After	
1	E6 Hp06 km 1.1	North	160	105	350	350	
2n	E6 Hp07 km 0.4	North	150	100	350	350	Side guardrail
2s	E6 Hp07 km 0.4	South	140	100	360	350	Side guardrail
3	E6 Hp07 km 0.7	North	150	90	350	350	
4	E6 Hp08 km 1.7	North	147	100	350	350	
5	E6 Hp08 km 2.5	North	127	100	376	350	Side guardrail

At the time of the before measurements in 2005, new asphalt had been laid at observation sites 4 and 5, and the visual central reserve had been marked. The white side road marking was not carried through. During another project there had been similar measurements at sites 4 and 5 in 2004. These can be used as before measurements. The speed and lateral position measurements for ordinary cars are presented in Fig. 68 and Fig. 69.

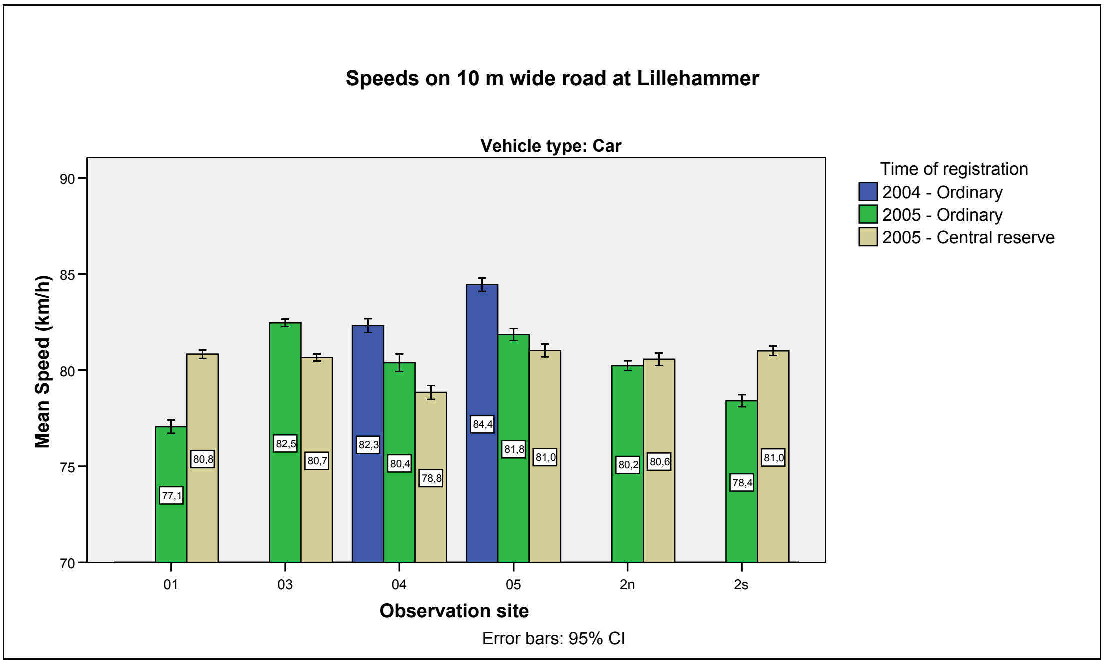


Fig. 68. Speed on the real world 10 metre wide road at Lillehammer.

The change in speed varied between an increase of 3.7 km/h and a decrease of 3.4 km/h from the before to the after situation.

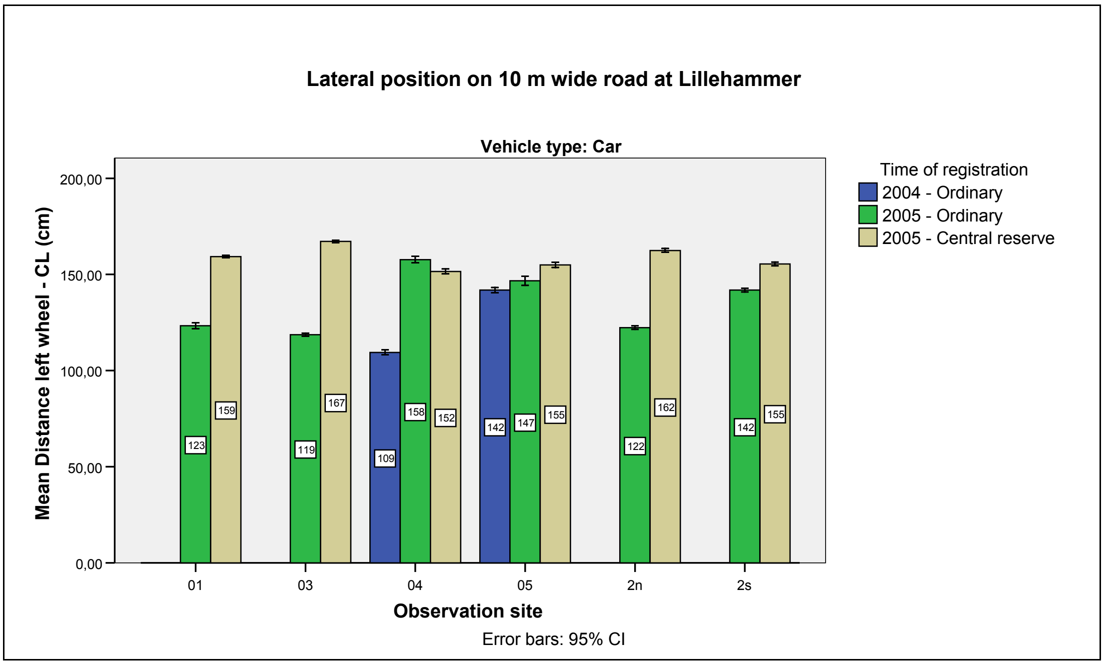


Fig. 69. Lateral positions on the real world 10 metre wide road at Lillehammer.

As shown in Fig. 56, several comparisons can be made between real world measurements and driving simulator experiments.
The lateral position differed between 12 and 48 centimetres in the before and after situations. The use of a central reserve resulted in the driver's distance from the road centre significantly increasing at a 95% confidence interval.

6.6 Discussion

6.6.1 Comparison of simulator and real world measurements.

In Fig. 56, three comparisons between results from the driver simulator and real world data are suggested:

- Absolute values of RS1 (Simulator studies) and RR1 (Real world measurements). Both have RW= 8.5m, LW = 3.5m.
- Relative change in values comparing RS1 (simulator studies) and RR2 (real world data) Both have RW =8.5m, while the lane width changes between LW=3.5m and 3.25m.
- Relative values between RS2 (Simulator studies) RR3 (Real world data). In both situations RW=10.0m, LW=3.5m, while the difference between the scenarios was based on a central reserve and no central reserve.

Absolute values of RS1 for speed showed that there was less fluctuation in the mean speed in the driving simulator compared to the real world measurements of RR1. The mean speed varied between 83.5 km/h and 84.8 km/h in the driving simulator, and 82.9 km/h and 86.0 km/h for cars in the real world. The differences in absolute values between the simulator studies and real world measurements was less than 0.5 km/h for three of the curves while the largest difference was 2.2 km/h. These differences were not very large. The difference of 0.25 cm lane width actually had effects that were comparable to a decrease in speed (reduction of up to 1.5 km/h). It is unlikely that the real world lane width was precisely 3.5 metres at all the curves. Additionally, the simulator measurements were not influenced by traffic volumes in the same way as in real traffic situations.

Comparison of absolute values for lateral position for simulator studies of RS1 and real world studies RR1 revealed some of the same characteristics as for speed. In the driving simulator, the distance only varied between 77 cm and 90 cm, while the distance varied between 30.32 cm and 97.40 cm for cars in the real world recordings. It seems that in the real world recordings, the largest distance to the road centre was found in right curves, while the shortest distance was found in left curves. These findings were not so evident in the driving simulator recordings. It seems that the driver in the real world uses more of the lane width in their driving than in the driving simulator, where the focus appears to be on driving at a constant speed and position in the lane. This focus on constant driving might as likely be due to the experiment situation as the properties of the driving simulator itself. Even with these uncertainties, the mean speed and mean lateral position seems to be of the same size in the driving simulator and in real world.

The simulator measurements of lateral positions in RS1 and in the real world measurements RR2 showed similar results. The relative change in lateral positions because of a change in lane width of 0.25 m was relatively small for both measurement methods. The best estimate of RR2 was a change in lateral position of between 14 and 23 cm, depending upon which wheel was measured. In the simulator the change in lateral position was between 10 cm and 12 cm. These results were quite similar.

When comparing the relative values between RS2 (Simulator studies) and RR3 (Real world data) from a 10 metre wide road, it might seem that the results concerning speed were not the same. In the driving simulator the mean speed was reduced between 0.2 km/h and 1.9 km/h at all the observation point when a visual central reserve was used. Based on the measurements alone it appeared that the real world measurement of speed sometimes increased and sometime decreased because of the introduction of a visual central reserve. At some measurement points there was an increase of 1.7 km/h in the after situation, while at other observation points there was a reduction of up to 3.4 km/h. In addition to the change in road markings, there was also a change to new asphalt from the before situation to the after situation. According to (Elvik and Vaa 2004), new asphalt can increase speeds between 2 and 5 km/h. If the effect of new asphalt is assumed to be 3.5 km/h, the real world measurements gave a speed reduction of between 0.1 km/h and 5.2 km/h.

The relative change in lateral position was between 12 cm and 48 cm in the real world measurements (RR3) when a visual centre reserve was introduced. In the driving simulator, the change in lateral position was between 38 cm and 49 cm at the different observation sites. As in all previous comparisons between real world and simulator measurements, the mean differences were of the same size, but there were greater variations in the real world measurements.

6.6.2 The effect of road, lane and shoulder width on speed and lateral position

Both the driving simulator studies and the real world measurements showed that the mean speed was reduced when the lanes were narrower and when a visual central reserve was in place. The lateral placement was also affected by the movement of the centre line or the edge line. A movement away from the centre of either line resulted in a greater distance from the centre of the road.

6.6.3 Validity of comparing real world data to driving simulator data

Statistical conclusion validity
Measurements error is often a problem when measuring data about speed and lateral position from real world. Instruments have to be placed according to strict rules. Quite frequently data are recorded with some kind of error. An example was found in (Sakshaug et al. 2004), where there were logical errors in the recorded data.

Internal validity
The real world measurements at Lillehammer were carried out using a before and after study. Even though there was little difference in time, there might have been changes in drivers or other factors. This road carries a lot of tourist traffic. The time of year of the measurements can influence the degree to which the drivers are tourists, which thereby also influences the speed.

In this study, the testing itself in the driving simulator can influence results. When driving the same situation four or five times, the subjects may change their speed because of the repetition.

In these studies it can be assumed that the experiment situation itself leads to specific driver behaviour. In this case the result was probably influenced by more concentration than usual and therefore less fluctuation in speed and lateral position in the driving simulator experiments.

Construct validity
The speed and lateral position measurements in real life are influenced by a number of factors such as weather, asphalt colour, asphalt quality and so on. A failure to adequately describe these constructs and not take them into the account when describing the effect of change in road markings will lead to errors. A special focus must be placed on the influence of new asphalt on the speed.

External validity
This chapter shows how the lack of knowledge about influences on speed and lateral position might lead to difficulties in making correct decisions about the effects. It might be just as difficult to generalize real world measurements as it is to generalize driving simulator measurements. A number of confounding variables have to be taken into account in real world measurements when calculating the effect of changes of road markings.

6.7 Conclusion

The most important findings were that reduction of lane width seems to reduce speed, and that marked central reserve increased the distance between oncoming vehicles significantly. All these findings point in the direction of improves safety. This conclusion is the same if you look at the results from the simulator studies separately or roadside registrations separately. When the results from theses different research tools are the same, the validity of the results is improved.

Lateral position and speed are two parameters for which it is seemingly easy to make comparison between real life measurements and driving simulator experiments. This chapter shows that this is not necessarily true.

All three comparisons showed that the mean speed and mean lateral position were of equal size both in terms of absolute values and in terms of relative values related to effects of change in road markings. The nature of driving in the driving simulator appears to result in a more constant speed and lateral position. This might be because of the experimental situation in the driving simulator, where the test subject concentrates more on driving in a more “correct” way. The lateral position is more stochastic in real life. This is probably because there are more confounding variables in real life. This does mean that when trying to generalize findings, both real life data and simulator data have their shortcomings, but when combined, the validity of both methods are greatly improved.

The measurements show that the differences between the results from the driving simulator and the real world measurements have plausible explanations.

7. Validation case 3 – Time gap

7.1 Background

The purpose of this case study was to look at the ability to study and analyse the interaction between adjacent vehicles. Models of this interaction are usually called car following models. One key concept in car following models is called the time gap or headway. Time gap is the time between the passing of the rear of a vehicle to the passing of the front of the following vehicle. Headway denotes the time between the passing of the front part of two successive vehicles.

In this chapter the intention is to make an introductory study of methods for measuring time gap and headway. There have been no such studies at NTNU/SINTEF in either the driving simulator or the instrumented vehicle; the only related information comes from roadside measurements. Results from one such study by (Giæver 1993) is presented as a reference for data collected using the driving simulator and the instrumented vehicle. Data collection in the driving simulator and in the instrumented vehicle was done in conjunction with other projects. The purpose of this exercise was to get experience conducting this kind of research. The presentation of the research is therefore also at a simpler level than the results from validation cases 1 and 2.

7.2 Concept of time gap and headway

In addition to being used in car following models, time gap and headway are used to calculation of maximum traffic volume, junction capacity studies and studies of traffic safety.

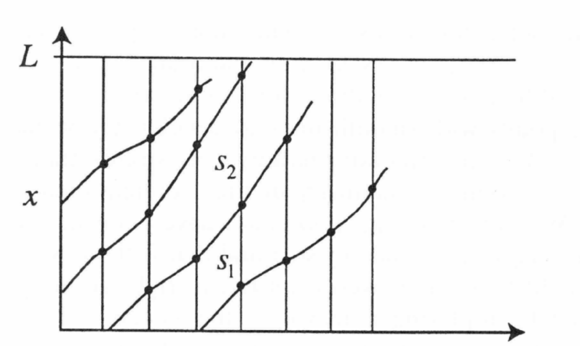


Fig. 70. Time – space diagram where traffic density at a road with length L at certain times (t1-t6) is measured (from (Daganzo 1997).

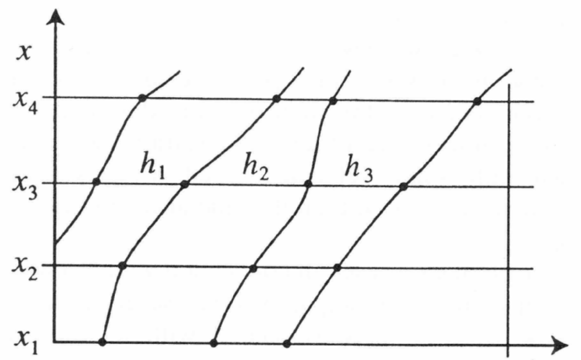


Fig. 71. Time-space diagrams where traffic volume at time interval T at certain points (x1-x4) is measured (from (Daganzo 1997).

Time gap and headway can be measured in several different ways. Fig. 70 shows a time – space diagram where several individual vehicles are plotted. The distance separating between consecutive vehicles at a given time is usually called spacing. An example of the kind of measurements used to plot these kinds of diagrams are aerial photos taken at specific time intervals. Time gap and headway are calculated based on measured speed and distances between the vehicles. This kind of measurement will also result in information about traffic density.

Fig. 71 shows a time – space diagram where measurements were conducted at certain fixed locations. Time gap and headway were registered directly at the fixed locations. The data collection can be done with ordinary roadside equipment. This will result in information about the traffic volume.

To be able to draw the trajectory between the measurements in both Fig. 70 and Fig. 71, we have to be able to identify each vehicle.

One option for obtaining more detailed information about the time gap is to measure it using a vehicle that is in the traffic flow. (May 1990) described how these kinds of measurements were done in the 1960s, where time gap was calculated based on continuous measurements of speed distance between cars using a wire between the cars. Today such measurements are much easier to undertake using either radars or lasers.

Car following models are becoming more and more important. These types of models have traditionally been used in used in micro simulations, but are now also being used in advanced vehicle control and safety systems. As more and better data have become available, more sophisticated models have been created. (Brackstone and McDonald 1999) have described the history of car following models.

7.3 A literature review of time gap and headway measurements in driving simulators

(Klee and Radwan 2004) assessed the use of a driving simulator for traffic engineering and human factors studies. Two studies involving traffic engineering and human factors were undertaken to see if they could be conducted in a driving simulator instead of relying on actual field test data. Two separate projects were chosen. The first involved the subject of gap acceptance by drivers. Published data for gap acceptance for a specific type of merging manoeuvre was compared to the experimental results obtained in the driving simulator under similar merging conditions. The second study was a human factors investigation of a radar-based safety warning system using in-vehicle voice and text warnings to alert drivers about the presence of impending conditions requiring their attention. Findings from the gap acceptance study contradicted published AASHTO (2001) data, which suggested that the minimum acceptable gap for a left turn from a minor road on to a major road was independent of the traffic speed on the major road. The simulator results indicated that the acceptable gap was reduced as the major road speed increased. In the human factors study, responses from the subjects indicated that the safety warning system accomplished the intended purpose of informing drivers and raised their perceived awareness levels to potential road hazards without confusing or distracting them. However, the subjects were undecided as to any perceived safety benefits.

(Boer et al. 2005) studied driver performance assessment with a car following model.

Test track car following data was used to show how drivers differ in their car following control strategies by introducing and identifying a driver car following model for each driver. It was demonstrated that the adopted target time strongly influenced the associated control strategy as well as the safety margin.

(Abe and Richardson 2006) studied how drivers might respond to Forward Collision Warning Systems (FCWS). The driving simulator study focused on alarm timing and its impact on driver response to alarm. The experimental investigation considered driver perception of alarm timings and their influence on trust at three driving speeds and two time headways. The results showed that alarm effectiveness varied with driving conditions. Alarm promptness had a greater influence on ratings of trust than improvements in braking performance enabled by the alarm system. Moreover, alarms that were presented after braking actions had been initiated were viewed as late alarms. It was concluded that drivers typically expected alarms to be presented before they initiated braking actions and when this does not happen, driver trust in the system is substantially decreased.

(Sauer et al. 2003) studied a model for car following based solely on optical parameters. The model was developed and compared with the performance of human drivers in a simulator. The model used the optical size of the back of the car being followed and the first derivative of its optical size as inputs. The model consisted of two components: one that accelerated or decelerated to maintain the visual size of the leading car, and another that accelerated or decelerated to minimize changes in the rate of change of the visual size of the leading car. The simulator presented drivers with a leading car that was changing its speed according to a sum of non-harmonic sines. Comparisons of human drivers' performance with the models' showed a high degree of similarity.

7.4 Experiment design

7.4.1 NTNU/SINTEF driving simulator

The driving simulator scenarios used in this chapter have been described in section 6.4.2 as scenario SS1 and SS2 with a road width of 8.5 metres. At certain parts of the track a vehicle was placed in front of the test subject's vehicle. The vehicle drove at a constant speed. The scenario was designed in a manner that made it difficult to overtake the vehicle, forcing the test subject to drive behind the slow moving vehicle for a longer period. In total 14 test subjects drove the road 4 times. Fig. 72 shows how the scenarios were used.

	Lane width = 3.5 metres	Lane width = 3.25 metres
Leading vehicle/ queue situation	validation case 3 (Chapter 7)	validation case 3 (Chapter 7)
No leading vehicle	validation case 2 Chapter 6)	validation case 2 (Chapter 6)

Fig. 72. Use of the different scenarios driven by the test subjects.

Car following was defined as a time gap less than 5.0 seconds. Only situations where the time gap was less than 5.0 seconds were used in statistical calculations.

7.4.2 Instrumented vehicle

The data for time gap in the instrumented vehicle were collected in conjunction with the “Instrumented road” research project. At the time of measurement, two driving support systems were in use in the instrumented vehicle: ISA and Mobil eye. The ISA system can influence speeding situations. The Mobil eye system has both a lane departing warning system and a warning for short time gaps. Detailed data were collected from 4 drivers, 2 females and 2 males. The ERS-logger described in section 3.4.4 was used to collect the data and the radar in front was used to measure distance to the vehicle in front. Data were collected at a rate of 10 Hz. The time gap was calculated on the basis of speed of the instrumented car and the measured distance between the vehicles.

Based on video images, parts of the measured data were selected and used for the analysis. The parts were selected based on two parameters:

- The road is an ordinary two-lane road.
- There is a vehicle visible in front.

Only situations where time gap was less than 5.0 seconds were used for statistical calculations.

It was also planned to use data from the rear facing radar to calculate the time gap for the following vehicle, but there were no data most of the time. Error checking, calibration and further tests have to be conducted to see if the rear facing radar can be used for its intended purpose.

7.5 Results

7.5.1 Driving simulator

In those situations where there was a car following situation in the driving simulator, the mean time gap was 1.13 seconds, while the standard deviation was 0.64.

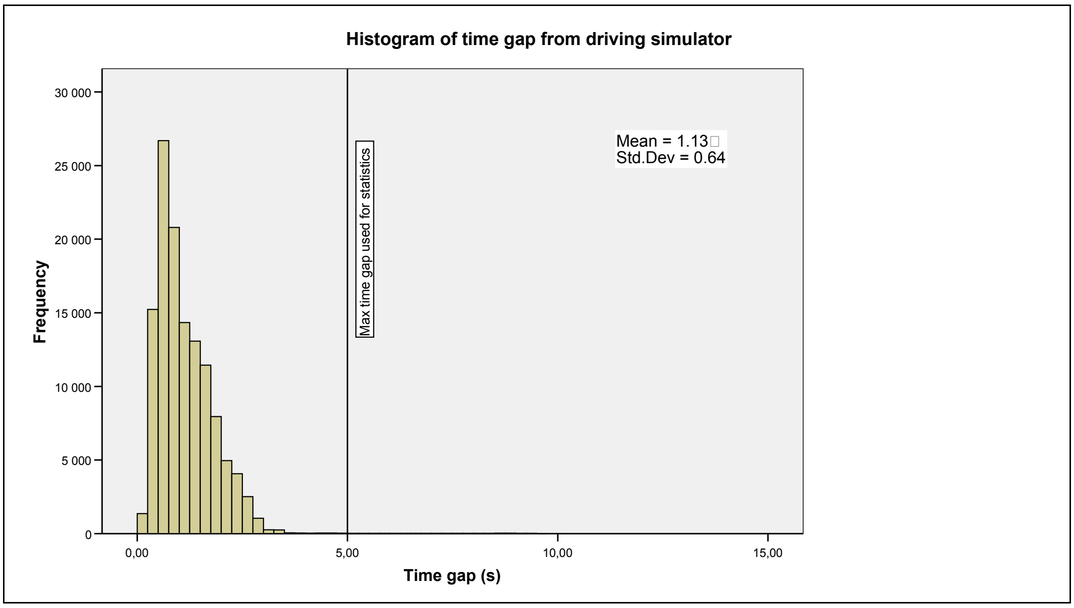


Fig. 73. Histogram of the time gap measured in the driving simulator.

Fig. 73 shows a histogram of the time gap measured in the driving simulator. No time gap over 10 seconds was measured. Using the maximum time gap of 5 seconds when calculating statistics, we get a time gap of 0.38 seconds for the 5th percentile, 0.62 seconds for the 25th percentile, and a median of 0.97 seconds.

7.5.2 Instrumented vehicle

Within those time periods where it can be subjectively determined from the video image that there is a vehicle in front, there were measurements of distance in 74.5% of the cases. There are 25.5% missing data about distance.

Fig. 74 shows a histogram of the time gaps measured in the instrumented vehicle. The mean time gap was 2.57 seconds while the standard deviation was 1.12.

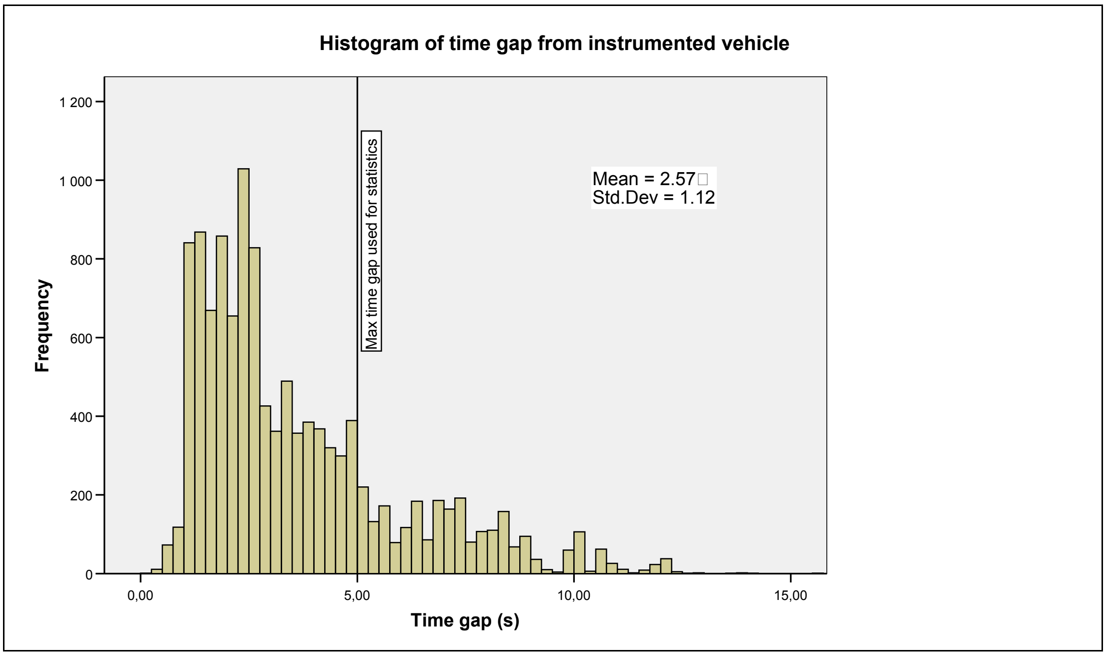


Fig. 74. Histogram of time gap measured in the instrumented vehicle.

Using the maximum time gap of 5 seconds when calculating statistics we get a time gap of 1.1 seconds for the 5th percentile, 1.6 seconds for the 25th percentile, and a median of 2.41 seconds.

The distribution of the time gap measurements of the instrumented vehicle and the driving simulator showed a skewed distribution, but had different mean and Standard deviation.

7.5.3 Roadside measurements

The results here were found in (Giæver 1993), where headway were measured in situations where the headway was less than 5 seconds. Measurements were done at 6 observation sites at two different time periods (morning and afternoon).

If a car length was set to 5 metres and the mean speed was set to 80 km/h, the comparable time gap was 0.23 seconds lower than the headway. In other words, one must reduce the headway times presented here by 0.23 seconds to directly compare it with time gap measurements in the simulator and instrumented vehicle.

In the roadside measurements the mean headway was between 2.07 seconds and 2.73 seconds at the different observation sites when there was no rain and the asphalt was dry. The comparable time gaps were between 1.84 seconds and 2.5 seconds.

7.6 Discussion

7.6.1 Time gap measurements in the driving simulator

The time gaps that were measured in the driving simulator were very low. A mean time gap of 1.13 (about 1.35 seconds headway) would result in a capacity of one lane being between 2600 and 2700 vehicles an hour. I am not aware of any measurements of lane capacity of this magnitude in Norway, but in good highway conditions in the United States there might be capacities of this magnitude.

There might be several reasons for this low time gap:

- There was no dangerous situation that would normally make the test subject maintain a larger time gap.
- The vehicle in front was driving at a constant speed. This made the situation more predictable for the test subject.
- There were just one or two cars in front of the driver, making it easy to control the traffic situations facing them.
- Even if special efforts were made to ensure that no obvious situation occurred so the test subject would overtake the preceding vehicle, the test subject would not know this and would be constantly looking for an overtaking situation.
- The driving conditions were the best imaginable, e.g. there was no rain, the asphalt was dry and there were no events that would disturb the driver during the session.

All of these reasons can to some extent be corrected in the driving simulator. More complex driving situations can be introduced, more traffic can be present, the driving conditions can be altered, and the driver can be asked to do a specific task while driving.

An example of the use of a more complex model for speed of the autonomous vehicles is given by (Sauer et al. 2003), where the speed was modelled according to a sum of non-harmonic sines.

7.6.2 Instrumented vehicle

The mean time gap measured in the instrumented vehicle was much higher than those in the driving simulator. A mean time gap of 2.57 seconds seems to be high. The corresponding traffic volume is between 1200 and 1300 vehicles/hour. The mean time gap was very dependant on the selected maximum time gap. Reducing the maximum time gap by 0.23 seconds (corresponding to a maximum headway of 5 seconds) reduced the mean time gap to 2.47 seconds.

There might be several reasons for this large time gap:

- The driver support systems may probably have influenced the time gap. The Mobil eye system most likely has an effect since a warning signal was given when the time gap is too short.
- The mean time gap was very dependent on the chosen maximum time gap for the situations included in the statistical analyses.
- There were few situations where overtaking was an option, resulting in a more relaxed driving style.
- The situations that did occur where time gaps were below 5 seconds might not be real car following situations, because no queue situation was present.

7.6.3 Roadside measurements

The mean headway varied quite a bit between the different observation sites. The difference between 2.07 seconds and 2.73 seconds is quite substantial, but nevertheless, all results fell between the measurements from the driving simulator and those from the instrumented car. It can be noted that the difference was there even if the measurement method and driving conditions were controlled.

7.6.4 Discussion related to validity

Statistical conclusion validity
In the real world study using the instrumented vehicle, only four test subjects were evaluated. One must also be aware that even if the numbers of measurements were very high because of the continuous measurements, these were not completely independent of each other. The number of independent measurements can be increased by increasing the number of test subjects or by increasing the number of distinct car following situations.

Internal validity
The instrumentation of the instrumented vehicle might play a role in the measurements. The radar was made for the purpose of Automatic Cruise Control and was therefore optimized for such use.

Construct validity
In a comparison of the situation of the driving simulator and the instrumented vehicle, one can see that several factors were not equal. In the driving simulator there was a situation where the driver might be looking for possibilities to overtake. The driving simulator also featured a situation where the driver in front drove at a constant speed. This is not the case in most real world situations.

External validity
The results from this chapter show how important it is to control the confounding factors both in the driving simulator and in the real world. The possibility of using the results in different situations demands detailed knowledge about how confounding variables are controlled for.

7.7 Conclusions

7.7.1 Time gap

When conducting research related to time gap in the driving simulator, close attention must be paid to the confounding variables. Several reasons for the short time gaps found with the driving simulator and the long time gap using the instrumented vehicle are given. Other roadside measurements also fluctuate to some extent.

Taken into account the difference in situations between the driving simulator situation, the instrumented car situation and the roadside measurements, all measurements seems plausible.

7.7.2 Validity of using driving simulator in time gap studies

The stochastic variables presented by real world driving situations demand special attention. A number of variables can be altered in the driving simulator, creating another car following situation. In the real world, several confounding variables will influence the time gap, and if these are not considered when analysing results from the driving simulator, the wrong conclusions may be reached.

8. Conclusions

8.1 Use of driving simulators

Driving simulators are steadily becoming a more and more important research tool. A great deal of research has been conducted using both real world data and data from driving simulators. Advancements in computer science have steadily improved the performance of driving simulators, making them cheaper and more widely available.

The use of driving simulators gives us new measuring methods and the ability to control situations. The control of variables is one advantage of the driving simulator, but it is also a challenge. In the time gap validation (case 3), the results were lower than initially expected partly because the situation designed for the test was too simple relative to normal real world situations.

All traffic research improves our knowledge about traffic and behaviour. At the same time it is important to use this knowledge when conducting and designing new research projects. This makes it possible to come to additional conclusions about the validity of results from driving simulator experiments, which in turn helps make driving simulators more acceptable as a possible research tool.

The advantages of using driving simulators are often said to be lower costs, and the ability to conduct research that would otherwise not be possible in the real world. This thesis demonstrates that an additional advantage of using driving simulators for research is being able to control confounding variables. This ability in turn leads to more accurate measurement and estimates of effects.

The use of a driving simulator cannot replace other research and measurements, but it is a valuable addition, in that it makes it possible to conduct new and more specific research.

Some methods related to observational studies describe change in behaviour, but it is often difficult to accurately describe the reason for the change. The experimental nature of using a driving simulator eliminates this problem. Driving simulators enable researchers to have better control over experimental situations and confounding variables, which makes it possible to repeat experiments.

Variables such as workload are usually difficult to measure directly. Instead, indicators that can be measured directly, such as speed, lateral position and eye movement, can be used as indirect measures of workload or attention to primary tasks.

8.2 Improving driving simulator validity as a research tool

The main focus of my thesis work was to compare the research and the results from the driving simulator with other research sources, and to provide more knowledge about the use and validity of driving simulators.

Based on the literature review, it appears that studies dealing with the validation of driving simulators show a fairly good match between speed in real life and speed in the driving simulator. Most drivers adapt to the driving situation in the driving simulator. There might be an unpredictable effect of using the driving simulator for those who have a high cognitive load while driving, such as in individuals with Alzheimer's disease.

The results from validity cases 1 – 3 show that the validity of using the driving simulator is strongly related to the research purpose. The design of the scenarios might have just as much influence on the results as the design of the simulator itself.

The use of general models of validity makes it evident that the whole construct and operationalization of the driving simulator must be validated, not just the driving simulator as a research tool.

The use of a driving simulator as a research tool demands a greater understanding of traffic situations than real world measurements, because the results are so dependent on the design and on traffic situations. The extended use and improved knowledge of driving simulators continually makes it possible to improve the validity of their use. At the same time the use of driving simulators cannot replace real world measurements because it is important to use alternative measurement methods.

In this thesis I have discussed four main aspects of validity, statistical conclusion validity, internal validity, construct validity, and external validity. This discussion is made both on a general basis and related to the three cases. Some conclusions can be drawn from this. To reduce the threats to validity of using a driving simulator it is important to pay attention to the whole research process. Test subjects have to be given a thorough introduction to driving in the driving simulator. Using a pool of drivers experienced with the driving simulator has benefits, but must be taken into consideration in the design of the experiment. The ability to control settings makes it possible for the test subjects to deduce the purpose of the study, which introduces new threats to the validity of the test. The lack of stochastic variations in the driving simulator also means that test subjects can quickly learn to expect the lack of variation, and adapt their driving behaviour accordingly.

All of my validation cases provided logical results in a comparison of driving simulator studies to real world studies. The differences between real world results and driving simulator results were generally not larger than the variances in real world results. The most important difference between real world measurements and driving simulator studies was the smaller standard deviation in the driving simulator results.

References

Abe, G., and Richardson, J. (2006). "Alarm timing, trust and driver expectation for forward collision warning systems." *Applied Ergonomics*, 37(5), 577-586.

Allen, R. W., Fancher, P. S., Levinson, W. H., Machey, J., Mourant, R. R., Schnell, T., and Srinivasan, R. (2000). "Simulation and measurement of driver and vehicle performance." Transportation Research Board.

Allen, R. W., Park, G. D., Fiorentino, D., Rosenthal, T. J., and Cook, M. L. "Analysis of Simulator Sickness as a Function of Age and Gender." *DSC Europe 2006*.

American Association of State Highway and Transportation Officials. (2001). *A Policy on geometric design of highways and streets*, American Association of State Highway and Transportation Officials, Washington, D.C.

Anderson, S. W., Rizzo, M., Shi, Q., Uc, E. Y., and Dawson, J. D. "Cognitive abilities related to driving performance in a simulator and crashing on the road." *3rd International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design*

Anund, A., Hjälmdahl, M., Palmkvist, G., Sehammar, H., and Thorslund, B. (2005). "Placering och utformning av frästa räfflor på landsväg – en simulatorstudie." *VTI rapport 523A*.

Arnedt, J. T., Geddes, M. A. C., and MacLean, A. W. (2005). "Comparative sensitivity of a simulated driving task to self-report, physiological, and other performance measures during prolonged wakefulness." *Journal of Psychosomatic Research*, 58(1), 61-71.

Bella, F. (2005). "Validation of a Driving Simulator for Work Zone Design." *Transportation Research Record*(1937), 136-144.

Boer, E. R. (2001). "Behavioral entropy as a measure of driving performance." *Proceedings of the First International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design*, 225-229.

Boer, E. R., Ward, N. J., Manser, M. P., Yamamura, T., and Kuge, N. "Driver performance assessment with a car following model." *Third International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design*.

Boff, K. R., Lincoln, J. E., and Harry G. Armstrong Aerospace Medical Research Laboratory. (1988). *Engineering data compendium : human perception and performance*, Harry G. Armstrong Aerospace Medical Research Laboratory, Wright-Patterson Air Force Base, Ohio.

Brackstone, M., and McDonald, M. (1999). "Car-following: a historical review." *Transportation Research Part F: Traffic Psychology and Behaviour*, 2(4), 181-196.

Brünger-Koch, M., Briest, S., and Vollrath, M. "Virtual driving with different motion characteristics - braking manoeuvre analysis and validation." *DSC 2006 Europe*.

Campagne, A., Pebayle, T., and Muzet, A. (2004). "Correlation between driving errors and vigilance level: influence of the driver's age." *Physiology & Behavior*, 80(4), 515-524.

Carsten, O. M. J., Groeger, J. A., Blana, E., and Jamson, A. H. (1997). "Driver performance in the EPSRC Driving Simulator: A validation study (Report no. GR/K56162)." *Leeds: University of Leeds*.

Chrstos, J. P., and Grygier, P. A. (1997). "Experimental testing of a 1994 ford taurus for nadsdyna validation." *SAE transactions*, 106(6), 895-908.

Clarion, A., Leclerc, F., Petit, C., Holler, S., Collet, C., and Mollard, R. "Revelant indicators to assess driver familiarization period in a dynamic simulator." *DSC 2006 Europe*.

Comte, S. L., and Jamson, A. H. (2000). "Traditional and innovative speed-reducing measures for curves: an investigation of driver behaviour using a driving simulator." *Safety Science*, 36(3), 137-150.

Cozby, P. C. (2003). *Methods in behavioral research*, McGraw-Hill, Boston.

Daganzo, C. F. (1997). *Fundamentals of transportation and traffic operations*, Pergamon, Oxford.

Dagdelen, M., Berlioux, J.-C., Panerai, F., Reymond, G., and Kemeny, A. "Validation process of the ULTIMATE high-performance driving simulator." *DSC 2006 Europe*.

de Waard, D., Steyvers, F. J. J. M., and Brookhuis, K. A. (2004). "How much visual road information is needed to drive safely and comfortably?" *Safety Science*, 42(7), 639-655.

Department Of The Army. (2000). "Army Model and Simulation Standards Report FY00." Department Of The Army.

Drews, F. A., Burns, S., and Strayer, D. L. "The development and evaluation of a high-fidelity simulator training program for snowplow operators " *Third International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design*

Edwards, C. J., Creaser, J. I., Caird, J. K., Lamsdale, A. M., and Chisholm, S. L. "Older and younger driver performance at complex intersections: implications for using perception-response time and driving simulation." *Second International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design*.

Elvik, R. (1999). *Assessing the validity of evaluation research by means of meta-analysis : case illustrations from road safety research*, Transportøkonomisk institutt, Oslo ,.

Elvik, R., Christensen, P., and Amundsen, A. (2004). "Speed and Road accidents–an evaluation of power model." Oslo TÖI, 2004.

Elvik, R., and Vaa, T. (2004). *The Handbook of road safety measures*, Elsevier, Amsterdam.

Engen, T., and Giæver, T. (2004). *Reaksjonstid i vegtrafikken*, SINTEF, Teknologi og samfunn, Veg og samferdsel, Trondheim.

Fambro, D. B., Fitzpatrick, K., and Koppa, R. J. (1997). *Determination of stopping sight distances*, National Academy Press, Washington, D.C.

Freund, B., Gravenstein, S., Ferris, R., and Shaheen, E. (2002). "Evaluating driving performance of cognitively impaired and healthy older adults: A pilot study comparing on-road testing and driving simulation." *Journal of the American Geriatrics Society*, 50(7), 1309-1310.

Garay-Vega, L., and Fisher, D. L. (2005). "Can novice drivers recognize foreshadowing risks as easily as experienced drivers?" 3rd International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design, Rockport, Maine USA

Giacomin, J., and Fustes, F. (2005). "Subjective equivalence of steering wheel vibration and sound." *International Journal of Industrial Ergonomics*, 35(6), 517-526.

Giæver, T. (1993). "Trafikkavvikling under vinterforhold." *STF 63 A93014*, SINTEF Samferdslsteknikk, Trondheim.

Giæver, T., and Engen, T. (2005a). "Evaluering av midtfelt - Lillehammer." *N-06/05*.

Giæver, T., and Engen, T. (2005b). *Testing av rumlefelt på 10 meters veg*, SINTEF, Teknologi og samfunn, Veg og samferdsel, Trondheim.

Giæver, T., and Engen, T. (2005c). *Testing av visuell midtdeler*, SINTEF, Teknologi og samfunn, Veg og samferdsel, Trondheim.

Godley, S. T. (1999). "A driving simulator investigation of perceptual countermeasures to speeding." Monash University. Dept. of Psychology.

Godley, S. T., Triggs, T. J., and Fildes, B. N. (2002). "Driving simulator validation for speed research." *Accident Analysis and Prevention*, 34(5), 589-600.

Gray, R. R., Rogers, D. M., Hewitt, G., Eberley, P., and Heckart, R. (2007). "Federal Aviation Administration (FAA) Human Factors Awareness Web Course." F. A. Administration, ed.

Green, M. (2000). "How long does it take to stop? Methodological analysis of driver perception-brake times." *Transportation Human Factors*, 2(3), 195-216.

Greenberg, J., Curry, R., Blommer, M., Kozak, K., Artz, B., Cathey, L., and Kao, B. "The Validity of Last-Second Braking and Steering Judgments in Advanced Driving Simulators." *DSC 2006 Europe*.

Guzek, M., and Jurecki, R. "Comparative analyses of driver behaviour on the track and in virtual environment." *DSC 2006 Europe*, Paris.

Hakamies-Blomqvist, L., Östlund, J., Henriksson, P., and Heikkinen, S. (2000). *Äldre bilförare i simulator–en valideringsstudie (VTI Rapport 464).*

Hauer, E. (1997). *Observational before-after studies in road safety : estimating the effect of highway and traffic engineering measures on road safety*, Pergamon, Oxford, Engl. ; Tarrytown, N.Y.

Helmers, G., and Törnros, J. (2006). *Effekt av övergångskurvor på förares säkerhetsmarginal samt inverkan av träning : ett försök i körsimulator*, VTI, Linköping.

Hjälmdahl, M., and Várhelyi, A. "Workload-study of ISA-drivers a method description." *13th ICTCT workshop*, Corfu, Geece.

Hjälmdahl, M., and Várhelyi, A. (2004). "Validation of in-car observations, a method for driver assessment." *Transportation Research Part a-Policy and Practice*, 38(2), 127-142.

Horberry, T., Anderson, J., and Regan, M. A. (2006a). "The possible safety benefits of enhanced road markings: A driving simulator evaluation." *Transportation Research Part F-Traffic Psychology and Behaviour*, 9(1), 77-87.

Horberry, T., Anderson, J., Regan, M. A., Triggs, T. J., and Brown, J. (2006b). "Driver distraction: The effects of concurrent in-vehicle tasks, road environment complexity and age on driving performance." *Accident Analysis and Prevention*, 38(1), 185-191.

Hoskins, A. H., and El-Gindy, M. (2006). "Technical report: Literature survey on driving simulator validation studies." *International Journal of Heavy Vehicle Systems*, 13(3), 241-252.

Hugemann, W. (2002). "Driver Reaction Times in Road Traffic." *EVU annual conference, Portoroz, Slovenia, September*.

Ihs, A. (2006). *Vägutrustningars effekt på trafiksäkerhet, tillgänglighet och komfort : en körsimulatorstudie rörande effekten av mittlinje respektive kantstolpar*, VTI, Linköping.

Jamson, H. (1999). "Curve negotiation in the Leeds Driving Simulator: The Role of Driver Experiences." *Engineering in Psychology and Cognitive Ergonomics, Ashgate, Aldershot*, 3, 351–358.

Johansson, G., and Rumar, K. (1971). "Drivers'brake reaction times." *Human Factors*, 13(1), 23-7.

Kaptein, N. A., Theeuwes, J., and Van Der Horst, R. (1996). "Driving simulator validity: Some considerations." *Transportation research record*(1550), 30-36.

Karlsson, R. (2006). *Evaluating driver distraction countermeasures*, VTI.

Kemeny, A., and Panerai, F. (2003). "Evaluating perception in driving simulation experiments." *Trends in Cognitive Sciences*, 7(1), 31-37.

Kircher, A., Sandin, J., and Uddman, M. (2002). *Vehicle control and drowsiness*.

Klee, D. H., and Radwan, D. E. (2004). *Assessment of the Use of a Driving Simulator for Traffic Engineering and Human Factors Studies*

Ku, J. H., Jang, D. P., Lee, B. S., Lee, J. H., Kim, I. Y., and Kim, S. I. (2002). "Development and validation of virtual driving simulator for the spinal injury patient." *Cyberpsychology & Behavior*, 5(2), 151-156.

Lee, H. C. (2002). "The Validity of Driving Simulator to Measure On-Road Driving Performance of Older Drivers." Conference of Australian Institutes of Transport Resaerch, CAITRR.

Lee, H. C., Lee, A. H., and Cameron, D. (2003). "Validation of a driving simulator by measuring the visual attention skill of older adult drivers." *American Journal of Occupational Therapy*, 57(3), 324-328.

Lenneman, J. K., Shelley, J. R., and Backs, R. W. "Deciphering psychological-physiological mappings while driving and performing a secondary memory task." *3rd International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design*

Lerner, N. "Disconnect between driver behavior/performance studies and crash experience: lessons from the study of young/inexperienced drivers." *International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design*.

Lew, H. L., Poole, J. H., Lee, E. H., Jaffe, D. L., Huang, H. C., and Brodd, E. (2005). "Predictive validity of driving-simulator assessments following traumatic brain injury: a preliminary study." *Brain Injury*, 19(3), 177-188.

Lindsey, J. T., and Barron, A. E. "The perceptions of emergency vehicle drivers using simulation in driver training " *Third International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design*

Lotan, T., and Toledo, T. "Evaluating the safety implications and benefits of an in-vehicle data recorder to young drivers." *Third International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design*

Marcotte, T. D., Roberts, E., Rosenthal, T. J., Heaton, R. K., Bentley, H., and Grant, I. "Test-retest reliability of standard deviation of lane position as assessed on a pc-based driving simulator." *Second International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design*.

May, A. D. (1990). *Traffic flow fundamentals*, Prentice Hall, Englewood Cliffs, N.J.

McGehee, D., Mazzae, E., and Baldwin, G. H. S. (2000). "Driver reaction time in crash avoidance research: Validation of a driving simulator study on a test track." *Proceedings of the International Ergonomics Association 2000 Conference*.

McGehee, D. V., Lee, J. D., Rizzo, M., and Bateman, K. "Examination of older driver steering adaptation on a high performance driving simulator." *International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design*.

Meek, S. (1990). "En studie av reaksjonstider på "Head Up Display"-symboler under ulike trafikkforhold," Hovedoppgave, Norges tekniske høgskole, Trondheim.

Moe, D. (1995). "Evaluering av bilsimulator Autosim A/S – del1: Alphatest." Sintef samferdselsteknikk, Trondheim.

Muttart, J. W. "Quantifying driver response times based upon research and real life data " *Third International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design*

NADS. (2006). "The National Advanced Driving Simulator portfolio." 12.

Nilsson, G. (2000). "Hastighetsförändringar och trafiksäkerhetseffekter." VTI notat 76-2000

Nilsson, G. (2004). "Traffic Safety Dimensions and the Power Model to Describe the Effect of Speed on Safety."

Nordmark, S., and Åström, H. "Improvement and validation of the VTI driving simulator." *2nd International Colloquium On Vehicle Tyre Road Interaction*, Florence, 30.

OKTAL. (2006). "Scanner II Solution for driving Simulation, User manual, Release 2.14."

Olson, P. L., and Farber, E. (2003). *Forensic aspects of driver perception and response*, Lawyers & Judges Publ., Tucson, Ariz.

Owens, J. M., and Lehman, R. "The effects of age and distraction on reaction time in a driving simulator." *International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design*.

Paul, A., Boyle, L. N., Boer, E. R., Tippin, J., and Rizzo, M. "Steering entropy changes as a function of microsleeps " *Third International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design*

Pretto, P., and Chatziastros, A. "Changes in optic flow and scene contrast affect the driving speed." *DSC 2006 Europe*, Paris.

Reimer, B., D'Ambrosio, L. A., Coughun, J. F., Kafrissen, M. E., and Biederman, J. (2006). "Using self-reported data to assess the validity of driving simulation data." *Behavior Research Methods*, 38(2), 314-324.

Reymond, G., Heidet, A., Canry, M., and Kemeny, A. (2000). "Validation of Renault's dynamic simulator for Adaptive Cruise Control experiments." *Proceedings of the Driving Simulation Conference DSC*, 6–8.

Rødseth, J., Moen, T., Jensen, G. D., Asmusvang, B., Danielsen, H., Høseggen, S., and G.Ihler. (2002). *IKT i vegtrafikken : effektive, sjåførvennlige og trafikksikre IKT-løsninger for transport på veg*, SINTEF, Bygg og miljø, Veg og samferdsel, Trondheim.

Sakshaug, K. (2002). *Skilting av fartsgrense forbi arbeidssted*, SINTEF, Bygg og miljø, Veg og samferdsel, Trondheim.

Sakshaug, K., Giæver, T., and Lervåg, L.-E. (2004). *Grunnlag for revisjon av håndbok 017 Veg- og gateutforming: Skulder- og kjørebanebreddens betydning for trafikksikkerheten*, SINTEF, Bygg og miljø, Veg og samferdsel, Trondheim.

Salaani, M. K., and Heydinger, G. J. (2000). "Model Validation of The 1997 Jeep Cherokee For the National Advanced Driving Simulator." *2000-01-0700.*

Sauer, C. W., Andersen, G. J., and Saidpour, A. (2003). "Car following by optical parameters." 2nd International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design, Park City, Utah USA

Shadish, W. R., Cook, T. D., and Campbell, D. T. (2001). *Experimental and quasi-experimental designs for generalized causal inference*, Houghton Mifflin, Boston.

Slick, R. F., Cady, E. T., and Tran, T. Q. "Workload changes in teenaged drivers driving with distractions." *Third International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design.*

Soanes, C., and Stevenson, A. (2003). *Oxford dictionary of English*, Oxford Univ. Press, Oxford.

Stolwyk, R. J., Triggs, T. J., Charlton, J. L., Iansek, R., and Bradshaw, J. L. (2005). "Impact of internal versus external cueing on driving performance in people with Parkinson's disease." *Movement Disorders*, 20(7), 846-857.

Strayer, D. L., Drews, F. A., and Crouch, D. J. "Fatal distraction? A comparison of the cell-phone driver and the drunk driver." *Second International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design*

Summala, H., Lamble, D., and Laakso, M. (1998). "Driving experience and perception of the lead car's braking when looking at in-car targets." *Accident Analysis and Prevention*, 30(4), 401-407.

Tornros, J. (1998). "Driving behaviour in a real and a simulated road tunnel - A validation study." *Accident Analysis and Prevention*, 30(4), 497-503.

Törnros, J., and Wallman, C.-G. (2003). "Inverkan av spår i beläggningen på förarbeteendet – en förstudie i VTI:s körsimulator." *VTI meddelande 940.*

Vegdirektoratet. (1992). *Veg- og gateutforming*, Vegdirektoratet, Oslo.

Victor, T. W., Harbluk, J. L., and Engstrom, J. A. (2005). "Sensitivity of eye-movement measures to in-vehicle task difficulty." *Transportation Research Part F-Traffic Psychology and Behaviour*, 8(2), 167-190.

Wahl, R., Flø, M., Moen, T., Johannessen, B., Engen, T., and Haugen, T. (2002). *Elektronisk datafangst for godstransport i byområder*, SINTEF, Bygg og miljø, Veg og samferdsel, Trondheim.

Wentworth, J. A., Knaus, R., and Aougab, H. (1995). *Verification, Validation and Evaluation of Expert Systems.*

Wood, C., and Hurwitz, J. "Driver workload management during cell phone conversations." *3rd International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design*

Wortman, R. H., and Matthias, J. S. (1983). *An evaluation of driver behavior at signalized intersections*, Arizona Dept. of Transportation; National Technical Information Service.

Young, R., Aryal, B., Muresan, M., Ding, X., Oja, S., and Simpson, S. N. "Road-to-lab: validation of the static load test for predicting on-road driving performance while using advanced in-vehicle information and communication devices " *Third International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design*

Östlund, J., Nilsson, L., Törnros, J., and Forsman, Å. (2006). *Effects of cognitive and visual load in real and simulated driving*, VTI.