Preface

This report is the result of a fruitful cooperation between Saab Automobile, Volvo Cars Corporation (VCC), Scania and the Swedish National Road and Transport Research Institute (VTI). The reported work has been undertaken in the project Secondary Task Workload Test Bench – 2TB within the competence centre Virtual Prototyping and Assessment by Simulation (ViP).

The scope of the project was to develop a simulator-based test method for assessing the load imposed on the driver from secondary tasks (in-vehicle equipment), i.e. to investigate commonly used performance indicators for their sensitivity to different levels of difficulty of secondary tasks loading on different modalities, and to evaluate their robustness across platforms.

The project was financed by the competence centre ViP, i.e. by ViP partners and the Swedish Governmental Agency for Innovation Systems (Vinnova).

We are grateful to all who contributed to the project, and special thanks go to ViP for financing the study.

Linköping, September 2012

Katja Kircher and the project team
Quality review

External peer review was performed in May 2012 by Håkan Alm, Luleå University and Paul Green, University of Michigan Transportation Research Institute (UMTRI). Katja Kircher has made the final alterations to the manuscript. The ViP Director Lena Nilsson examined and approved the report for publication on 26 March 2014.
Table of contents

Executive summary ........................................................................................................ 9
1  Preamble .................................................................................................................. 11
2  Introduction ............................................................................................................. 12
  2.1 Problem statement ............................................................................................ 12
  2.2 Driver performance metrics in interface design .................................................. 12
  2.3 Project objectives and goals ............................................................................. 13
3  Background ............................................................................................................... 15
  3.1 Reference task requirements ............................................................................. 15
  3.2 Reference tasks .................................................................................................. 16
4  Method ...................................................................................................................... 20
  4.1 Apparatus ........................................................................................................... 20
  4.2 Road ..................................................................................................................... 23
  4.3 Traffic .................................................................................................................. 24
  4.4 Participants ......................................................................................................... 24
  4.5 Design ................................................................................................................ 25
  4.6 Procedure ........................................................................................................... 26
  4.7 Research questions and analysis ....................................................................... 27
5  Results and interpretation ....................................................................................... 34
  5.1 Baseline comparisons ....................................................................................... 34
  5.2 Secondary task performance ............................................................................ 37
  5.3 Performance ratings ......................................................................................... 42
  5.4 Driving behaviour .............................................................................................. 45
  5.5 Driver behaviour ............................................................................................... 48
  5.6 Interpretation of driver and driving behaviour results .................................... 49
  5.7 Inter-individual differences .............................................................................. 50
6  Discussion ................................................................................................................ 53
7  Conclusions ............................................................................................................. 56

References .................................................................................................................... 57
### Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>DLP</td>
<td>Digital Light Processing projector</td>
</tr>
<tr>
<td>ESoP</td>
<td>European Statement of Principles</td>
</tr>
<tr>
<td>GLM</td>
<td>General Linear Model</td>
</tr>
<tr>
<td>HASTE</td>
<td>Human Machine Interface and the Safety of Traffic in Europe project</td>
</tr>
<tr>
<td>HMI</td>
<td>Human Machine Interface (interaction)</td>
</tr>
<tr>
<td>ISO</td>
<td>International Organization for Standardization</td>
</tr>
<tr>
<td>IVIS</td>
<td>In-Vehicle Information System</td>
</tr>
<tr>
<td>LCD</td>
<td>Liquid Crystal Display</td>
</tr>
<tr>
<td>LCT</td>
<td>Lane Change Test</td>
</tr>
<tr>
<td>PI</td>
<td>Performance Indicator</td>
</tr>
<tr>
<td>SAE</td>
<td>Society of Automotive Engineers</td>
</tr>
<tr>
<td>S-IVIS</td>
<td>Surrogate In-Vehicle Information System</td>
</tr>
<tr>
<td>TFT</td>
<td>Thin Film Transistor</td>
</tr>
<tr>
<td>US DOT</td>
<td>United States Department of Transport</td>
</tr>
<tr>
<td>VCC</td>
<td>Volvo Car Corporation</td>
</tr>
<tr>
<td>VTI</td>
<td>Swedish National Road and Transport Research Institute</td>
</tr>
</tbody>
</table>
List of figures

Figure 1: Visualisation of the numbers read from tape (t) and the expected answers (a) to be given. ............................................................... 18

Figure 2: Visualisation of the haptic task of inserting wooden bricks into a children’s play box without looking................................................................. 19

Figure 3: Graphical representation of the road curviness. VTI study to the left and VCC study to the right................................................................. 23

Figure 4: The roads used for the experiment in the VTI (left) and VCC (right) simulators. ................................................................. 24

Figure 5: Visualisation of one reference task block and one baseline block. The order of the sections within each block is balanced across and within participants, the levels of task difficulty always appeared in the same order................................................................. 26

Figure 6: Graphical representation of the analyses. Light grey boxes represent baseline drives without lead car. Dark grey boxes represent baseline drives with lead car. Red, green and blue boxes represent the three different secondary tasks on three levels of difficulty (from unsaturated to saturated colours). The different shadings represent the driving condition under which the secondary task was executed (no black component: secondary task only; slight black component: driving without lead car; heavy black component: driving with lead car)................................................................. 29

Figure 7: Percentage answers to stimuli for the visual task across all participants, for both locations, all driving conditions and all task difficulty levels................................................................. 38

Figure 8: Percentage correct answers of all answered stimuli for the visual task across all participants, for both locations, all driving conditions and all task difficulty levels. 39

Figure 9: Mean task completion times (s) for the correct answers on the visual task across all participants, for both locations, all driving conditions and all task difficulty levels................................................................. 39

Figure 10: Number of correct answers on the cognitive task across all participants, for both locations, all driving conditions and all task difficulty levels................................................................. 40

Figure 11: Number of correctly placed bricks minus the number of incorrectly placed bricks across all participants, for both locations, all driving conditions and all task difficulty levels................................................................. 41

Figure 12: Self-rated performance on the driving task, depending on the concurrently executed secondary task (most difficult level) and lead car presence................................................................. 43

Figure 13: Self-rated secondary task performance on the most difficult level, depending on secondary task type and driving condition................................................................. 43

Figure 14: Percentage of mental effort invested in the secondary task as opposed to the driving task, depending on secondary task type and lead car presence................................................................. 44

Figure 15: Lane departure distance (m) for individual VTI (red) and VCC (blue) participants. Mean (fat solid line) and median (fat dashed line) values for all secondary task conditions, with and without lead car, and for all task difficulty levels (1, 2, and 3). ................................................................. 51
Figure 16: Percent road centre (PRC) with a radius of 8° for individual VTI (red) and VCC (blue) participants. Mean (fat solid line) and median (fat dashed line) values for all secondary task conditions, with and without lead car, and for all task difficulty levels (1, 2, and 3).
List of tables

Table 1: The arrow matrices for the three levels of difficulty in the visual reference task.......................................................... 17
Table 2: The numbers read from tape for the three levels of difficulty in the cognitive reference task.......................................................... 18
Table 3: Performance specifications of the VTI Simulator III.......................................................... 20
Table 4: Generalised research questions for the study at hand, with motivations........ 28
Table 5: Classification of the PIs into different categories. ...................................................... 30
Table 6: PI (performance indicator) definitions. .......................................................... 31
Table 7: F-values and levels of significance for the driving behaviour PIs for baseline comparisons. Within each row the upper values indicate the VTI results and the lower values indicate the VCC results. When a box is filled with grey, there is a significant difference (p < .05) between VTI and VCC for the factor combination in question. Rows marked with red indicate PIs that relate to traffic code violations. (*): p < .10; *: p < .05; **: p < .01. ...................................................... 35
Table 8: F-values and levels of significance for the driver behaviour PIs for baseline comparisons. Within each row the upper values indicate the VTI results and the lower values indicate the VCC results. When a box is filled with grey, there is a significant difference (p < .05) between VTI and VCC for the factor combination in question. (*): p < .10; *: p < .05; **: p < .01. ...................................................... 36
Table 9: F-values and significances for all tested driving behaviour PIs for main effects and 2-way interactions. Within each row the upper values indicate the VTI results and the lower values indicate the VCC results. When a box is filled with grey, there is a significant difference (p < .05) between VTI and VCC for the factor combination in question. Red bars indicate PIs related to a traffic rule violation. (*): p < .10; *: p < .05; **: p < .01. ...................................................... 46
Table 10: F-values and significances for all tested driver behaviour PIs for main effects and 2-way interactions. Within each row the upper values indicate the VTI results and the lower values indicate the VCC results. When a box is filled with grey, there is a significant difference (p < .05) between VTI and VCC for the factor combination in question. (*): p < .10; *: p < .05; **: p < .01. ...................................................... 48
Secondary Task Workload Test Bench - 2TB
Final Report
by Katja Kircher¹, Christer Ahlström¹, Annie Rydström², Mikael Ljung Aust², Daniel Ricknäs³, Susanne Almgren⁴, and Arne Nåbo⁴

¹ Swedish National Road and Transport Research Institute (VTI)
² Volvo Cars
³ Scania
⁴ Saab Automobile

Executive summary
The main aim of this study was to investigate a selection of commonly used performance indicators (PIs) that have been reported to be sensitive to distraction and workload. More specifically, the PIs were tested for their ability to differentiate between task modalities (visual, cognitive and haptic) and task difficulty (easy, medium and hard). It was investigated whether possible differences were constant across two traffic situations (with/without lead vehicle) and two driving simulators.

The experiment was conducted in the VTI Driving Simulator III, an advanced moving-base simulator, and in the Volvo Car Corporation (VCC) driving simulator, an advanced fixed-base simulator. Both simulators were equipped with Smart Eye Pro eye tracking systems. The simulated road was a rural road made up of alternating left and right curves. The posted speed limit was 90 km/h. In the moving-base simulator 24 drivers participated, and in the fixed-base simulator 12 drivers participated. The design was within-subjects for the factors task modality, task difficulty and traffic situation.

A visual, a cognitive and a haptic secondary task were chosen to test the ability of the PIs to distinguish between the tasks’ loading on different modalities. The visual task was a modification of the arrows task developed in the HASTE project, the cognitive task was a combined computation and memory task based on the paced serial addition task, and the haptic task was developed within the project. For the haptic task the participants had to insert wooden bricks of different shapes into matching holes without visual aid.

The longitudinal PIs tested were mean speed, standard deviation of speed, speeding index and minimum time headway (THW). Lateral PIs were mean lane position, modified standard deviation of lane position (SDLP), number of lane crossings, steering wheel reversal rate (SWRR), lane departure distance and lane departure area. Eye movement PIs were percent road centre (PRC), longest glance away from the road, total eyes-off-road time, and two PIs describing the actual attention level based on the AttenD algorithm.

The main results from the study were:

- There were only minor differences between the two simulator types for driving behaviour as described by longitudinal PIs. There was no overall offset, and the main difference was that the visual task led to stronger speed reductions in the moving-base simulator, which influenced both the mean speed and the speeding index.

- The minimum THW was similar for both simulators and all modality types, except
that it increased with increasing task difficulty level for the visual and the haptic tasks in the moving-base simulator. However, this increase was due to one outlier who reduced the speed massively and therefore increased the THW so much that the mean of all drivers got affected.

- Regarding lateral PIs, major differences between the two simulator types were found, both as a general offset and for those factor combinations that include modality and task difficulty level.

- With the visual or the haptic task active, the drivers positioned themselves further to the left and the SDLP was higher in the fixed-base simulator.

- The number of lane crossings did not differ considerably between the simulators, but the lane departure area was larger on average in the fixed-base simulator, again influenced by modality, with the largest lane departure areas for the visual task, and in the case of the fixed-base simulator for the haptic task as well.

- For all lateral PIs except the SWRR, for both simulators, and especially for the visual and the haptic tasks, the differences were only evident for the medium task difficulty level.

- Most of the eye movement related PIs had a general offset between the simulators. The drivers in the fixed-base simulator accumulated more time with their eyes off the road, especially during the visual and the cognitive tasks, while the drivers in the moving-base simulator cast longer single glances at the display.

Generally speaking, the longitudinal PIs behaved more similarly across simulator platforms, while the lateral control was much more impaired in the fixed-base simulator. Caution should be exerted in the interpretation of outcomes based on lateral PIs in fixed-base simulators.

It is unexpected that the lateral PIs should show the most extreme values for the medium level of task difficulty, especially as this could also be observed for the baseline condition, in which no task at all was conducted.

There were large intra-individual differences found across most of the investigated PIs. In some cases the behaviour is quantitatively different, with some drivers having much higher values than others but with the general trend being equal, while for other PIs the differences appear to be of a qualitative nature.

To conclude, assessing the effects of secondary tasks on traffic safety via PIs that are based on the control level of driving behaviour is difficult. Results vary widely between drivers, and the implications of changes in some PIs on safety are not always clear. It is therefore suggested to investigate the possibility of using PIs on the tactical level of driving behaviour, and instead of reporting mean values, that potentially are not very expressive, it may be worth exploring the notion that a certain safety envelope may not be breached by any or x % of the drivers. This may make the interpretation of results more straightforward.
Preamble

In the original proposal it was planned that simulator studies would be conducted at three facilities; at VTI, Saab Automobile and Volvo Cars (VCC). Due to economic circumstances, Saab was not able to fulfil this mission during the course of the project.
2 Introduction

Windscreen wipers, climate controls and radios are omnipresent in our vehicles, and there are not many people out there who question this. As vehicles develop, there will be more and more new features and applications just waiting to find their way into the cockpits of modern cars. These features and applications can come either as embedded devices, integrated into the vehicle system, or as nomadic devices, portable and not necessarily built for in-vehicle use. Depending on the system, its usage can have more or less impact on the driver’s workload. Here, all load on the driver coming from tasks not directly related to driving is called “secondary task workload”.

Within the eSafety initiative a group of experts developed the “European Statement of Principles” (ESoP) based on Recommendation 2000/53/CE from 1999, giving guidance for the development of safe human machine interfaces (HMI) for in-vehicle information and communication systems (ESoP Expert Group, 2005). An update was published in 2007 (ESoP Expert Group, 2007), taking into account the development in technology. ESoP recommendations include highly relevant recommendations, such as “The system does not distract or visually entertain the driver”. While this is immediately comprehensible, it is not easily operationalized in a consistent manner.

Just after the study presented here had been carried out the US DOT suggested a testing procedure for embedded devices and their potential to distract drivers visually and manually (US DOT, 2012). At the time of writing this report, it was still possible to leave comments on these suggestions.

2.1 Problem statement

The testing of secondary task workload for a vehicle meant for production can be quite substantial due to the amount of systems that need to be tested. It is therefore necessary to have a test method that is efficient but still relevant to a real driving situation. Also, the method would benefit from being standardized, meaning that results can be compared and communicated in a consistent manner. The outcome of such a test would preferably be whether driving and system operation can be done safely together in a given situation or not.

2.2 Driver performance metrics in interface design

In an attempt to operationalize distraction and standardize a test method for the suitability of navigation system usage while driving the Society of Automotive Engineers suggested the so-called 15-second rule (SAE, 2000), based on the original suggestion by Green (1999). The idea was that a task, which can be completed within 15 seconds while the vehicle is standing still, should be acceptable to execute while driving. This rule has been debated vigorously, and an expansion to accommodate adaptation for the type of load – visual, cognitive or haptic – has been suggested (Green, 2000; Reed-Jones, Trick, & Matthews, 2008; Tijerina, 2000; Tijerina, Johnston, Parmer, Pham, & Winterbottom, 2000; Tijerina, Johnston, Parmer, Winterbottom, & Goodman, 2000). Related to that, the SAE Standard 2365-2002 (SAE, 2002) provides guidelines on how to calculate task duration for navigation system usage, based on a task analysis.

ISO 16673:2007 (ISO, 2007) provides a procedure for measuring visual demand due to the use of visual or visual-manual interfaces. With this technique, the driver’s vision is occluded with a special pair of glasses that opens and shuts at various time intervals (Weir, Chiang &
The phase with the driver’s vision occluded simulates the time they are looking at the road, while the open phase represents the time that they are looking at the in-vehicle device. By using this method it is possible to evaluate whether a secondary task can be carried out using only short, periodic glances. Otherwise, the task is considered unsafe. Advantages with the visual occlusion technique is that it can be carried out while the car is standing still and also that it applies to both embedded and nomadic devices.

Another standardised task to assess the influence of secondary tasks on driver performance is the lane change test (LCT), which is subject of a proposed ISO standard. The LCT tries to reach the reliability of a laboratory experiment while increasing the validity by mimicking a real driving scenario (Mattes, 2003). Compared to a laboratory experiment, the artificial stimulus is replaced by road signs and artificial key presses are replaced by actual lane changes. Lane changing performance (primary task) is then used as a measure of how distracting the secondary task is.

Detection response tasks, such as the peripheral response task (van Winsum, Martens, & Herland, 1999) or the tactile response task (Engström, Aberg, Johansson, & Hammmarbäck, 2005), have also been introduced as measures of driver distraction. Detection response tasks measure the drivers’ ability to detect different stimuli. For example, in the peripheral response task, a visual stimulus is presented in the peripheral field of view. Depending on detection performance, it is possible to assess the impact of different secondary tasks.

Distraction assessment based on visual behaviour has also been suggested. ISO 15007:2002 describes definitions, techniques and procedures for assessing visual behaviour in terms of glances towards some display or other target in the vehicle (ISO, 2002; ISO under development). The underlying idea is similar to the occlusion technique, namely that during the execution of some rather complex visual task, several glances to the peripheral object take place. Glance durations are typically between 0.6 and 1.6 seconds, and drivers are generally unwilling to look away from the road for more than 2 seconds (Green, 2002). If thresholds for total glance duration, glance frequency, single glance duration and total task time are exceeded, the secondary task is considered unsafe. An evaluation of the occlusion technique and its usefulness for the assessment of distraction imposed by driver support systems was provided by Pettitt et al. (2006).

A drawback with visual behaviour analysis according to the ISO standard is that it is based on manual video transcription. Different solutions based on eye tracking have therefore been suggested. These include the percent road centre measure (Victor, Harbluk, & Engström, 2005), i.e. the percentage of time that the driver looks at the forward roadway, and the visual demand metric (Engström & Mårdh, 2007), which also accounts for eccentricity and duration of individual glances.


### 2.3 Project objectives and goals

There are a number of driver performance metrics available for guidance in interface design. Most of them come with both advantages and disadvantages. The validity of approaches that are performed in a laboratory setting, such as the 15-second rule, may be questioned. Assessing secondary task load by adding another (third) task, as in the detection response tasks, may also be questioned in many cases. More general concerns that are not addressed by any of the methods are factors such as familiarity, expectancy and the complexity of the
surrounding traffic environment. In this project, high validity and safety will be achieved by performing the test in high-fidelity driving simulators (with and without moving-base). Further, distraction assessment will not rely on any imposed third task.

The reported project should be seen as a first phase, where the main goal is to identify performance indicators for driving behaviour, driver behaviour and subjective ratings that reliably indicate changes in workload induced by a secondary task of visual, cognitive or haptic/tactile character. It should preferably be possible to differentiate between different levels of secondary task difficulty based on the output of the identified performance indicators.

In a future second phase, the surrounding traffic environment will be of increasing complexity. The idea is to evaluate in which situations a certain task will be safe to perform. A secondary goal of the second phase will be to fully automatize the result generation process.

The concrete objectives of the project are to:

1. Develop a testing procedure for secondary task workload measurement.
2. Perform a pilot study.
3. Develop a proposal for a full-scale project starting 2011 and apply for FFI funding.

In this first phase of the project, the method will:

- Use driving simulators of varying fidelity. They could range from a desktop solution to a high-fidelity solution.
- Use state-of-the-art measurement equipment and data treatment methods.
- Have a reference secondary task to provide a solid anchor task.

In the second phase of the project, the method will:

- Have a range of driving scenarios with different levels of “driving workload”.
- Describe the selection of participants to get a representative group.
- Describe the number of participants needed to get reliable statistics.
- Use automated result generation in order to be efficient.

As mentioned, the approach is to tackle the issue in two steps. During the first, the participants’ driving experience as well as the complexity of the secondary task are varied, while the complexity of the driving situation will be kept constant. During the second phase, variations in situational complexity will be incorporated and the analysis will be automatized as much as possible.
3 Background

Theories on attention generally accept the notion that there are limits to how much humans can process at one time. Whether there is a central bottleneck or a modality dependent bottleneck as suggested in the multiple resource theory (Wickens, 1984), and in which way automation, experience and individual capacity differences play a role has been discussed widely in the literature (for a relatively recent overview see Wickens & McCarley, 2008).

Driving is a dynamic task in a complex system (Hollnagel & Woods, 1983), which requires sustained attention from the driver. Exactly how much attention is required, and whether secondary tasks that include visual, manual or solely cognitive components have detrimental effects on driving performance or not, and how they differ depending on the load and the current situation has been debated at length (Regan, Lee, & Young, 2008).

The EU project HASTE has made a substantial contribution in investigating the suitability of different performance indicators for driver state assessment (Östlund et al., 2004). Tests were conducted both in the field and in different simulators. In HASTE focus was put on the visual and the cognitive channels, and so-called surrogate in-vehicle information systems (S-IVIS) were developed, which should represent the load put on drivers by real IVIS, but which could be manipulated for varying difficulty.

No S-IVIS for purely haptic load were developed, and in many studies visual and haptic activities are lumped together. It is, however, totally feasible to assume completely haptic interfaces that do not require any visual control. However, some cognitive resources are probably always needed for secondary task execution, regardless of whether the task is classified as visual, haptic, auditory or cognitive.

Load on different modalities have previously been shown to affect driver behaviour differently (Recarte & Nunes, 2000; Victor, Harbluk, & Engström, 2005; Östlund et al., 2004). Therefore, the same function can possibly affect the driver differently, depending on the interaction modality used. These differences in behavioural changes might have different effects on traffic safety in different traffic situations.

The current project was conceived as a preparation project for a larger scale investigation on what secondary tasks are acceptable to execute in different types of traffic situations. The present project focuses mainly on the identification of performance indicators (PIs) that can reliably assess a cognitive vs. visual vs. haptic load. For this purpose, secondary tasks with a known type of load should be used. As soon as those PIs are established, they can be used to assess functions whose type of load is not as clearly known.

In the present project the difficulty of the driving task is kept constant to enable assessing the effect of the different levels of load on different modalities. However, for the final assessment method it is planned to gradually increase the complexity of the traffic situation, in order to find possible cut-off points, and to watch for non-linear behavioural decrements.

3.1 Reference task requirements

For the reference tasks to work as such, they need to fulfil a number of requirements, which are described here.

1. Load on correct channel: The tasks must load highly on the modality they are intended to load on, which is either visual, cognitive (including speech) or haptic (including motoric load), and as little as possible on the other modalities. This is necessary to ensure that changes in the performance indicators really are due to a
capacity reduction in the modality in question, or, for that matter, in some central region, and not because another modality was drawn upon as well. Here, no physiological measurements of brain activity in different areas were conducted, but a simpler approach was chosen, which was based on expert evaluations of the load profile of the different tasks.

2. **Simplicity:** The task must be simple enough so that all participants have a very good chance of understanding it easily, and that the learning curve is very steep. It is important that the participants will reach their performance asymptote, so they do not improve their performance significantly during the course of the experiment.

3. **Levels of difficulty:** It must be possible to vary the difficulty of the task reliably, as it should be investigated whether an increased level of difficulty in the reference task will lead to a decreased performance in driving and driver behaviour. The task difficulty was determined via a set of pre-tests with the reference tasks in question.

4. **Readily resumed:** The participants must be able to resume the task quickly in case they lost track of the task at some point in time. Therefore, the consecutive actions should be stand-alone and not be dependent on the previous one.

5. **System-paced:** The task should be system-paced and not user-paced, so that all participants can be exposed to the same conditions.

6. **Task performance:** The task performance should be possible to quantify, if necessary on more than one dimension. This could include the time to answer, the correctness of the answer, precision, etc. This is necessary to be able to observe performance degradations in the reference task, which can then be related to performance in the driving task, and in baseline trials.

### 3.2 Reference tasks

The task prototypes were chosen based on the requirements listed above. The visual task was based on a widely used and published task, the cognitive task had been used successfully at VTI before, and the haptic task was developed specifically for the current project. It is acknowledged that all tasks incorporate a cognitive component, and the visual task includes a haptic component, but it is assumed that the channel after which the task has been named takes the biggest load for each respective task.

#### 3.2.1 Visual task

The visual task is a modification of the arrows task developed in the EU project HASTE (Östlund et al., 2004). It is a visual search task with the goal to determine whether an arrow pointing upward is present or not in a matrix of arrows pointing to the left, to the right or downward. In order to obtain suitable levels of difficulty desktop pre-tests were made with matrix sizes of 4 x 4, 5 x 5, and 6 x 6 arrows. A target arrow was present in half of the cases in a random position in the matrix. The distracting arrows could per matrix either point only to the left, only to the right, only down, or in all three directions, which yielded eight different test cases.

When executing this task as primary task in a desktop situation the error rate was minimal, therefore reaction time for correct answers was used as the indicator of task difficulty. Visual inspection of the reaction times for the different test cases and matrix sizes led to a selection of a 4 x 4 matrix with arrows either pointing to the left or to the right, with and without a target arrow, as the easiest level, Level I. The mean reaction time to produce a correct answer
for these cases lay at 1.10 s. Level 2, the medium difficulty level, was a 6 x 6 matrix with arrows pointing downward, with and without a target arrow. The mean reaction time for these cases was 1.45 s. Level 3, the most difficult level, was represented by a 6 x 6 matrix with arrows pointing in all three directions, with and without a target arrow. The mean reaction time for this level was 2.6 s. The arrow matrices for the different difficulty levels are listed in Table 1.

Table 1: The arrow matrices for the three levels of difficulty in the visual reference task.

<table>
<thead>
<tr>
<th>Levels</th>
<th>Cases selected per level</th>
<th>no target arrow present</th>
<th>target arrow present</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level 1</td>
<td></td>
<td>← ← ← ← ← ←</td>
<td>← ← ← ← ← ←</td>
</tr>
<tr>
<td>Level 2</td>
<td></td>
<td>↓ ↓ ↓ ↓ ↓ ↓</td>
<td>↓ ↓ ↑ ↓ ↓ ↓</td>
</tr>
<tr>
<td>Level 3</td>
<td></td>
<td>↑ ← ↓ ← ↑ ←</td>
<td>← ↓ ← ↓ ← ↓</td>
</tr>
</tbody>
</table>

3.2.2 Cognitive task
The cognitive task is a slightly modified version of the Paced Serial Addition Task (Gronwall, 1977; Sampson, 1956). The cognitive task loaded both on memory and the capacity to handle additions. Single digit numbers were read from tape, with a new number being read every third second. The participant had the task to respond with the sum of the last two numbers read, that is, he or she had to remember the last number from tape while responding to the current addition, and then wait for the next number to add to the one remembered (Figure 1).
Pilot trials with varying levels of difficulty of the additions as well as varying speed of the presented numbers were run to determine three distinct levels of difficulty. The correctness of the answers given, in combination with subjective ratings of the nine pilot participants, led to the selection of the three levels of difficulty (presented in Table 2), and the speed of a new number being read every third second. The number sequences could be read both from the first to the last number and from the last to the first number, which resulted in the same answers.

Table 2: The numbers read from tape for the three levels of difficulty in the cognitive reference task.

<table>
<thead>
<tr>
<th>Level 1</th>
<th>Level 2</th>
<th>Level 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>7</td>
</tr>
<tr>
<td>2</td>
<td>7</td>
<td>5</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td>1</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>1</td>
<td>5</td>
<td>9</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>8</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>5</td>
</tr>
</tbody>
</table>

3.2.3 Haptic task

To the knowledge of the authors, no readily available haptic task existed that fulfilled the requirement of not having a visual component. Several alternatives, like the usage of a knob or dial, were considered before the final arrangement was selected. Its advantages are that no vision is involved, that performance could be logged automatically via micro switches, and that the setup was easily transportable between the test sites.

The goal of the haptic task was to insert wooden bricks of three different shapes into matching holes in a wooden box. The box was placed between the front seats, within easy reach of the driver. To prevent the driver from looking down at the bricks the box was covered with a black cloth (Figure 2).

There were three cylindrical, three triangular and three cubic wooden bricks to choose from. Whenever a brick was inserted into the fitting hole it pressed a micro switch which registered the inserted brick in the log data.
Three levels of difficulty for the haptic task were created in pilot trials by using bricks of different shapes, and by varying the order in which the bricks should be inserted. On the easiest level the drivers were asked to only insert the cylindrical bricks, on the medium level they were asked to insert cubic and/or triangular bricks in any order, and on the hardest level the drivers should insert alternate between cubic and triangular bricks.

<table>
<thead>
<tr>
<th>Picture</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Image" /></td>
<td>A children’s play box was modified such that the bricks inserted in the box exit the box automatically and reach the open part, where they can be picked up again. Three bricks of each shape were available.</td>
</tr>
<tr>
<td><img src="image2.png" alt="Image" /></td>
<td>The play box is installed in the vehicle between the front seats. Here it can be seen how a round brick is inserted, but during the study a black cloth covered the box and the hand, such that no visual information could be obtained.</td>
</tr>
<tr>
<td><img src="image3.png" alt="Image" /></td>
<td>Here the black cloth is in place, but in the combined picture the hand holding the brick underneath the cloth shows through.</td>
</tr>
</tbody>
</table>

*Figure 2: Visualisation of the haptic task of inserting wooden bricks into a children’s play box without looking.*
4 Method

It was a goal of the project to keep the experimental design as similar as possible across the different test sites, to be able to compare the results for the different simulators.

4.1 Apparatus

The study was run in two different locations, in two different simulators, but with the same secondary tasks. It had been planned to conduct the experiment in three simulators, but due to economic issues this could unfortunately not be realised.

4.1.1 Simulators

Two passenger car simulators were included in the study, one with moving-base and one with fixed-base.

VTI

The experiment was conducted in the VTI Driving Simulator III, an advanced moving-base simulator. A Saab 9-3 cabin was mounted on the motion platform. The moving-base is used to generate forces felt by the driver while driving. It can be divided into three separate parts: a large linear motion, tilt motion and a vibration table. The linear drive has world leading performance, it can achieve the highest acceleration of all simulators in the world. Linear motion was used in the lateral direction in the present study. The tilt motion is used in the roll and pitch directions to simulate long term accelerations such as driving in a curve or longitudinal acceleration and deceleration. The vibration table provides additional capabilities to generate road roughness for higher frequencies. While the tilt motion tilts both the cabin and the graphics projection screen, the vibration table moves the cabin relative to the projection screen. Detailed performance specifications for the motion platform of the VTI simulator III are provided in Table 3.

Table 3: Performance specifications of the VTI Simulator III.

<table>
<thead>
<tr>
<th>System</th>
<th>Type of movement</th>
<th>Max values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cabin</td>
<td>Pitch angle</td>
<td>- 9 degrees to + 14 degrees</td>
</tr>
<tr>
<td>Roll angle</td>
<td></td>
<td>± 24 degrees</td>
</tr>
<tr>
<td>External linear motion</td>
<td>Maximum amplitude</td>
<td>± 3.75 m</td>
</tr>
<tr>
<td></td>
<td>Maximum speed</td>
<td>± 4.0 m/s</td>
</tr>
<tr>
<td></td>
<td>Maximum acceleration</td>
<td>± 0.8 g</td>
</tr>
<tr>
<td>Vibration table</td>
<td>Vertical movement</td>
<td>± 6.0 cm</td>
</tr>
<tr>
<td></td>
<td>Longitudinal movement</td>
<td>±6.0 cm</td>
</tr>
<tr>
<td>Roll angle</td>
<td></td>
<td>± 6 degrees</td>
</tr>
<tr>
<td>Pitch angle</td>
<td></td>
<td>± 3 degrees</td>
</tr>
</tbody>
</table>
The visual system in the VTI simulator III consists of 3 DLP projectors providing a 120 degrees forward field of view and 3 LCD displays for the rear view mirrors. The projectors have a resolution of 1280 x 1024 pixels.

The sound system consists of five speakers, two speakers in the dashboard close to the windshield, one speaker in each of the front doors, and one rear speaker. These speakers are controlled by a separate computer which can play sound in any of the five speakers. This computer dedicated to the control of the sound interacts with its environment through the network.

VTI

At VTI Smart Eye Pro 5.7 was used with four installed cameras. Data were logged with 120 Hz. In the preparation phase of the project it turned out that it was not possible to calibrate the Smart Eye cameras when the simulator was parked, and then track the eye movements successfully with the simulator in driving mode with the motion system switched on. The reason was that the cameras shifted position slightly in relation to each other when the motion system was switched on, which transports the cabin a few metres into the middle of the linear motion platform and adjusts the pitch angle.

This necessitated a camera calibration in place, when the motion system was on and the cabin position settled. Due to mechanical and safety restrictions it is not possible, however, to enter or exit the cabin with the system in motion without provoking an emergency stop. Therefore the participants had to be taught how to calibrate the cameras, which they then did when the cabin had reached the driving position.

Auto-calibration, which checks the calibration over time and corrects deviations based on the head model, was switched off in the present study.

VCC

The procedure at VCC was similar, except that Smart Eye Pro 5.6 with three cameras was used, data were logged with a frequency of 60 Hz, and calibration was made by the experimenter.
4.1.3 Reference tasks

Each reference task was presented on three levels of difficulty, with each level lasting for 30 s, and with a 30 s break between levels as well as before the first and after the last level. Level 1 was always presented first, followed by Level 2 and finally Level 3.

**Visual task**

In the study each participant was presented with matrices of arrows from Level 1, Level 2 and Level 3 (see Chapter 3.2.1). Every 5th second a new matrix appeared. A sound announced the appearance of a new arrow matrix. The arrows were displayed on a touch screen, which was mounted to the right of the middle console in the simulator.

The participant had to determine whether one of the arrows pointed upwards or not. If one such target arrow was present the participant was required to press a “yes” button on the touch screen. If no upwards pointing arrow was present a “no” button should be pressed. As soon as a button was pressed the arrows disappeared. No feedback for the correctness of the answer was given.

The task was presented and logged via a dedicated computer, and the log was time synched to the log from the simulator. The current target case, the correctness of the answer and the reaction time were logged.

**Cognitive task**

The cognitive task was presented such that a new number was played every third second (see Chapter 3.2.2). As each sequence was made up of 9 numbers, and the participant needed a few seconds to answer the last number, each block took about 30 s. The experimental design required each participant to do each level three times. For half of the participants the sequences were read in the order first-to-last, last-to-first, first-to-last, and for the other half of the participants the order was read the other way round. The start of each cognitive task sequence was announced auditorily by the first number being read.

**Haptic task**

The beginning of each 30-second episode of the haptic task was announced by a sound. Then the participant started inserting the wooden bricks into the holes with the goal to manage with as many bricks as possible before another sound announced the end of the 30-second segment (see Chapter 3.2.3).

4.1.4 Performance rating scale

Always when Level 3 of a secondary task was completed, and in the end of the baseline drives, the participants were asked to rate their driving performance and their performance on the secondary task for Level 3, or for the last portion of the baseline drive. Ratings were made by giving a number between 1, which indicated the worst performance possible, and 10, which indicated the best performance possible. The participants gave a verbal rating upon request by the experimenter.
4.1.5 Effort distribution rating

When about half of the participants had been run at VTI it was decided that an additional rating estimating the participants’ experienced distribution of effort should be included. Therefore, no such data are available from the first 14 participants in the study. Participants were asked to indicate how much attention they had allocated to the driving task and to the secondary task, respectively, provided that they had 100% attention to distribute. The participants answered verbally upon a question by the experimenter, after having given the performance rating.

4.2 Road

The road was a rural road with a width of 3.2 m per lane. It was made up of curves with a radius of 1000 m. At VTI the curve direction changed when a circle segment of 90° was completed, that is when ca. 1571 m were driven (Figure 3 left). At VCC, a similar curve design was used but here the road was on a closed curve (Figure 3 right).

![VTI VCC](image)

Figure 3: Graphical representation of the road curviness. VTI study to the left and VCC study to the right.

The road led through a forest with trees on either side of the road. The road markings differed at the two test sites: At VTI the central road marking was made up of a double solid line, indicating that passing was banned, and the road markings on the sides were interrupted lines (Figure 4 left). At VCC the central marking was an interrupted line with solid lines as outer borders (Figure 4 right).

The posted speed limit was 90 km/h. One speed limit sign was placed in the beginning of each driving section.
4.3 Traffic

All surrounding traffic consisted of passenger cars of varying make and colour, which were drawn randomly from the available cars.

In half of the driving sections the participants had two lead cars in front of them. These cars drove 50 m apart from each other and held a mean speed of 87 km/h, around which they varied their speed according to Equation 1. Thus, the speed of the lead car was not coupled to the participants’ speed. This allowed the participant to vary the distance according to his or her wishes, thereby influencing the own workload.

Equation 1. Computation of the lead vehicle speed. The iteration is based on the iteration speed of the simulator programme, which is 200 Hz.

\[ v = 87 + 2 \times \sin\left(\frac{\text{iteration}}{3000}\right) + \max\left(3 \times \sin\left(\frac{\text{iteration}}{10000} - \frac{\pi}{4}\right), 0\right) [\text{km/h}] \]

In all driving sections there was oncoming traffic. The traffic appeared randomly, with the restrictions that no cars could follow each other closer than with a distance of 30 m, and the time interval between oncoming cars should not be larger than 35 s. The oncoming traffic kept to its own lane and did not overtake.

4.4 Participants

In the VTI simulator 24 participants took part, and in the VCC study 12 participants took part.

VTI

The participants were recruited via an external company. The company was provided with the following inclusion criteria: Participants were required to be between 30 and 45 years of age, which was meant to narrow the variance in driving experience and age related effects. Further, participants were required not to use glasses while driving, as glasses can interfere with the eye tracker and degrade gaze tracking performance. Additionally, the participants
should have some experience with modern technology like mobile phones/smartphones, navigation systems, and the like, and have used them while driving. This should increase the likelihood of participants engaging in the reference tasks.

Altogether 24 participants took part, nine women and 15 men. The mean age was 42 ± 5 years. The participants had held their licences for 23 ± 6 years on average. One of the participants drove below 5000 km a year, seven reported to drive between 5000 and 10 000 km yearly, twelve reported to drive 10 001 to 20 000 km yearly, and the remaining four reported to drive more than 20 001 km yearly.

VCC
VCC employees were invited to participate in the study. The test group comprised 12 participants in total, 4 women and 8 men, with a mean age of 34 ± 7 years. The participants had held their licences for 17 ± 7 years on average. All the participants worked with research and development at VCC.

4.5 Design
In each of the two simulators included in the study the same full within-subjects design was run. The factor reference task was varied on four levels (baseline, visual, cognitive, haptic), the factor traffic was varied on two levels (lead car, no lead car), and the factor task difficulty was varied on three levels (easy, medium, hard). This resulted in a 4x2x3 design, which was completed with an additional baseline in which the reference task was performed while the vehicle was parked. This was necessary to obtain a baseline for the maximum performance on the reference task only, allowing an estimation of the performance decrement caused by driving, and consequently, of the effort placed on the secondary task. Two baseline blocks, with and without lead car, were driven both in the beginning and in the end of each run, to be able to control for behavioural changes over the course of the experiment. Within each block the order for the sections with or without lead car was balanced between and within participants.

Further, each participant completed three blocks with one of the three reference tasks in each block. The order of the blocks was balanced between participants. Within each block, there were always three sections. In one the participant only completed the reference task without driving, in one the participant drove with a lead car ahead while executing the reference task, and in one the participant drove without lead car ahead while executing the reference task. The order of the three sections within each block was balanced across participants.

Each section was built up such that the participant first drove or waited 30 s without performing the reference task, depending on which section in the block it was. Then the reference task was executed on Level 1 for 30 s, followed by a break of 30 s. Next the reference task was executed on Level 2 for 30 s, followed by another break of 30 s. Finally the reference task was executed on Level 3 for 30 s, followed by a last break of 30 s. When the participant had to drive at the same time, the time counter was started as soon as the participant had reached a speed of 75 km/h. For a visualisation of a reference task block and a baseline block see Figure 5.
4.6 Procedure

The procedure is described in detail for VTI, which was first out to run the study. For VCC only those aspects that differ from the VTI procedure are taken up.

**VTI**

Each participant received information and instructions for the experiment via mail a few days before the actual experimental run. A link and a login to a web-based background questionnaire was sent as well, together with the request to fill in the questionnaire before the experiment was run.

Upon arrival the participant was welcomed and led to the simulator hall. The experimental procedures were gone through once more, whereupon the participant signed the informed consent form. Afterwards all three secondary tasks were explained in detail and practiced until the experimenter felt that the participant surely had understood the task. The participant was informed that each reference task would be practiced in the simulator as well, before the respective block (Figure 5) was started. Then the participant was informed about the performance rating scale. It was stressed that only the secondary task performance during the third level should be rated. Driving performance should be rated for the same time period.

The participant was then seated in the simulator. The safety routines were explained, and the necessary equipment like the haptic task box and the screen for the visual task was pointed out to the participant. The participant was instructed to drive as he or she usually would on a road of the present type, and that it was not allowed to overtake. It was stressed, however, that it was paramount that the secondary task should be performed as well as possible.

Then the motion system was started and the simulator was placed in driving position. The participant was asked to calibrate the Smart Eye cameras, whereupon the eye tracking profile for the participant was built.

At first the participant drove a training session without secondary tasks in order to get used to the simulator. The training lasted for five minutes, and the participant was asked to brake and

---

<table>
<thead>
<tr>
<th>ref. task block</th>
<th>30 s</th>
<th>30 s</th>
<th>30 s</th>
<th>30 s</th>
<th>30 s</th>
<th>30 s</th>
<th>30 s</th>
</tr>
</thead>
<tbody>
<tr>
<td>ref. task only</td>
<td>Level 1</td>
<td>Level 2</td>
<td>Level 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ref task + driving with lead car</td>
<td>Level 1</td>
<td>Level 2</td>
<td>Level 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ref task + driving without lead car</td>
<td>Level 1</td>
<td>Level 2</td>
<td>Level 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>baseline block</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>driving only with lead car</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>driving only without lead car</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Figure 5: Visualisation of one reference task block and one baseline block. The order of the sections within each block is balanced across and within participants, the levels of task difficulty always appeared in the same order.*
accelerate a couple of times in order to get used to the simulator motion. When the training was finished and the participant did not have any further questions, two baseline sections were driven. They lasted 3.5 minutes each, and the only difference between them was that there was a lead car in one of the baselines and no lead car in the other baseline. The order was balanced between participants.

After the two baseline sections one of the three secondary task blocks started. Each was made up of three sections, one where the participant executed the secondary task while parked (secondary task baseline) and two where the participant executed the secondary task while driving, once with and once without lead vehicle. The order of the secondary task blocks as well as the section order within each block was balanced across participants. Each block started out with a practice session of the secondary task in question. The participant practiced to reach proficiency in the task as long as the experimenter deemed necessary. Within each section the secondary task executions always started with the easiest level, followed by Level 2, and then Level 3. In the sections that involved driving Level 1 was started 30 s after the participant had reached a speed of 75 km/h for the first time. When all three secondary task blocks were finished the participant drove the two baselines once more, with the order balanced across participants.

After each section the participant was asked to rate his or her performance between 1 and 10. In the driving baseline sections only driving was rated, in the secondary task baseline sections only secondary task performance on Level 3 was rated, and in the sections where both driving and secondary task execution took place, both activities were rated. About halfway through the experiment the participants were also asked how they felt they had allocated their attentional resources between driving and secondary task engagement, where applicable.

After the last baseline had been driven the simulator was parked, and the participant left the simulator. The participant was thanked and accompanied to the door.

**VCC**

The procedure at VCC was identical to the procedure at VTI, except that the safety instructions were different and the eye tracking system was calibrated by the experimenter.

### 4.7 Research questions and analysis

The study design allowed for within-subjects comparisons for all factors except for the factor location, which was between-subjects. The research questions touched different levels of analysis and are listed in Table 4. They are taken up again in the general discussion.
Table 4: Generalised research questions for the study at hand, with motivations.

<table>
<thead>
<tr>
<th>Research Question</th>
<th>Motivation for testing this research question.</th>
</tr>
</thead>
<tbody>
<tr>
<td>The PI will change in a consistent manner across participants with driving situation and task level.</td>
<td>Subject effects are limited, results are consistent across participants.</td>
</tr>
<tr>
<td>The longitudinal PI and the eye-tracking related PI will be more consistent across simulators than the lateral PI.</td>
<td>Earlier research has shown that linear motion in lateral direction affects lateral PI more (Greenberg, Artz, &amp; Cathey, 2003).</td>
</tr>
<tr>
<td>Driving and driver parameters will not differ between the first and the second set of baselines for the free vehicle situation.</td>
<td>Learning or fatigue effects do not influence the results in a consistent manner.</td>
</tr>
<tr>
<td>Driving and driver parameters will not differ between the first and the second set of baselines for the following vehicle situation.</td>
<td>Learning or fatigue effects do not influence the results in a consistent manner.</td>
</tr>
<tr>
<td>The more difficult the secondary tasks are the more incorrect execution of them will occur.</td>
<td>Performance is influenced by the intrinsic task difficulty of the secondary task (Östlund et al., 2005).</td>
</tr>
<tr>
<td>The secondary task performance will differ for the baseline, lead-car- and no-lead-car situation.</td>
<td>The driving situation influences the secondary task performance, as resources have to be shared (Wickens &amp; McCarley, 2008).</td>
</tr>
<tr>
<td>The driving performance will differ depending on the secondary task type, and will agree with findings in the literature.</td>
<td>It has been shown that especially visual and cognitive tasks affect driving behaviour differently (Recarte &amp; Nunes, 2000). Purely haptic tasks have not been investigated extensively.</td>
</tr>
<tr>
<td>Driving performance will decrease with increasing secondary task difficulty.</td>
<td>It is assumed that driving and the secondary task compete for the same central resource, and increasing demands from one task will decrease performance on the other task (Wickens &amp; McCarley, 2008).</td>
</tr>
<tr>
<td>Cognitive task performance will be rated as better during the secondary task baseline than with concurrent driving.</td>
<td>It is easier to concentrate on the secondary task only as compared to synchronising it with driving.</td>
</tr>
<tr>
<td>Driving performance will be rated differently depending on the secondary task type (better for cognitive and haptic than visual).</td>
<td>The participants will feel the performance decrements more for the visual load than for the cognitive and haptic load.</td>
</tr>
<tr>
<td>Secondary task performance will be rated differently depending on the driving condition (better for baseline than with concurrent driving).</td>
<td>The hypothesis goes hand in hand with the expected higher performance during secondary task baseline. A better performance is expected to be rated as better.</td>
</tr>
</tbody>
</table>

It was investigated whether the participants changed their behaviour over time, to ensure that later analyses are not influenced by artefacts of that kind. It was also investigated whether the performance on the secondary task differed while performing the secondary task only, as compared to performing it while driving. These analyses established the pre-conditions for the main analyses, which then compared the effect of the three levels of difficulty of the different secondary tasks on driver behaviour, driving behaviour and subjective judgements for driving.
with and without a car ahead. It was also investigated whether the performance indicator changed in a similar fashion between participants, which is important for its robustness.

Each participant drove one baseline with a car ahead and one baseline without a car ahead before the trials with the secondary tasks, and once more after the trials with the secondary task.

The comparisons (analyses) undertaken are shown in Figure 6. The statistical analyses used an $\alpha$-level of 0.05.

**Figure 6:** Graphical representation of the analyses. Light grey boxes represent baseline drives without lead car. Dark grey boxes represent baseline drives with lead car. Red, green and blue boxes represent the three different secondary tasks on three levels of difficulty (from unsaturated to saturated colours). The different shadings represent the driving condition under which the secondary task was executed (no black component: secondary task only; slight black component: driving without lead car; heavy black component: driving with lead car).

4.7.1 Performance indicator classification scheme

PIs for analysis were identified by building a PI classification scheme (Table 5). For the first phase of the project only control level PIs were used, that is, strategic and tactical decisions were not considered. This is also due to the study design used, which did not require any strategic or tactical decisions on the part of the driver.
Table 5: Classification of the PIs into different categories.

<table>
<thead>
<tr>
<th>Level 1</th>
<th>Level 2</th>
<th>Level 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Driver behaviour</td>
<td>Driver opinion</td>
<td>Driving behaviour</td>
</tr>
<tr>
<td>Level 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>driver internal</td>
<td>interaction with vehicle</td>
<td>interaction with secondary task</td>
</tr>
<tr>
<td>Level 3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>gaze direction</td>
<td>steering wheel</td>
<td>embedded nomadic</td>
</tr>
<tr>
<td>- PRC</td>
<td>- steering wheel</td>
<td>- estimation of own driving performance</td>
</tr>
<tr>
<td>- mean AttenD</td>
<td>- steering wheel reversal rate</td>
<td>- estimation of own driving performance</td>
</tr>
<tr>
<td>- AttenD warning</td>
<td>- steering wheel entropy</td>
<td>- estimation of own performance on secondary tasks</td>
</tr>
<tr>
<td>- energy radius</td>
<td>gas pedal</td>
<td></td>
</tr>
<tr>
<td>eyelid</td>
<td>eye tracking</td>
<td></td>
</tr>
<tr>
<td>head direction</td>
<td></td>
<td></td>
</tr>
<tr>
<td>electro-physiology</td>
<td></td>
<td></td>
</tr>
<tr>
<td>hand position</td>
<td></td>
<td></td>
</tr>
<tr>
<td>other</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

4.7.2 Performance indicator definitions

Each PI is calculated for each box in Table 6. For the different tasks, data from the 30-second segments are extracted and the PI is calculated based on data from this particular segment. In the baseline cases, both in the beginning and in the end of the experiment, the 30-second segments are extracted at corresponding locations along the road. In total, 26 different PIs are derived. These are defined in Table 6.

The PIs that are based on eye tracking are only derived if the gaze quality is above 0.2 for at least 70% of the data in the segment. The vertical and horizontal gaze directions were derived as a weighted mean of data from the left and the right eye. The dominant eye was given a weight of two plus the quality value of that eye unless no tracking was available. In that case
the weight was set to zero. The other eye was given a weight corresponding to the quality value of that eye. The fused gaze direction signal was subsequently filtered by a 0.5 sec wide interpolating median filter. A median filter was chosen since it does not smear out the edges associated with saccades. Interpolation was used to fill out short sequences of low quality data for example due to eye blinks. The interpolation was implemented by simply ignoring data with quality below 0.2 in the median calculation. Low quality data with duration longer than the window width were marked as missing values.

Table 6: PI (performance indicator) definitions.

<table>
<thead>
<tr>
<th>PI</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean speed</td>
<td>The mean velocity in the current segment.</td>
</tr>
<tr>
<td>max speed</td>
<td>The maximum velocity in the current segment.</td>
</tr>
<tr>
<td>std speed</td>
<td>The standard deviation of velocity in the current segment.</td>
</tr>
</tbody>
</table>
| speeding index            | The speeding index in the current segment. Speeding index is calculated according to the Uppåt project (Hjälmårdahl, Dukic, & Pettersson, 2009). Basically it quantifies how much the driver is violating the speed limit and large transgressions are punished more severely (\(v\) is the speed limit):

\[
\int_{\text{start segment}}^{\text{stop segment}} \left( \frac{|v(s)|}{v_l + 3} - 1 \right) ds / \text{total segment distance}
\]

| speeding area             | The area under the velocity curve where the speed exceeds the speed limit. In the sketch below this is indicated by the grey area where the velocity exceeds the speed limit. |
| mean THW                  | The mean time headway in the current segment. Time headway is defined as the distance from the front of the driven vehicle to the rear of the vehicle ahead divided by the velocity of the own vehicle. In the sketch below THW = d/v_{\text{own}}. |
| min THW                   | The minimum time headway in the current segment.                                                     |
| std THW                   | The standard deviation of the time headway in the current segment.                                  |
| **mean lat pos** | The mean lateral position in the current segment. The lateral position is measured from the centre of the lane. |
|**std lat pos** | The standard deviation of lateral position in the current segment (also SDLP). |
|**mod std lat pos** | Modified standard deviation of lateral position according to the AIDE project (Östlund et al., 2005). The lateral position time series is filtered with a 0.1 Hz 2nd order Butterworth high pass filter before the standard deviation is calculated. The high pass filter removes any variations of longer periods than 10 seconds. |
| **# lane crossings** | The total number of lane deviations where the outer margin of the left or the right wheel crosses the inside of the line marking for the lane boundaries. |
| **lane departure area** | The total area that any part of the vehicle is outside the lane boundaries in the current segment. The area is calculated as the integral of the lateral position time series according to the grey areas in the sketch below. |
| **lane departure time** | The total time that any part of the vehicle is outside the lane boundaries in the current segment. |
| **lane dept. distance** | The total distance that any part of the vehicle is outside the lane boundaries in the current segment. |
| **PRC 8 deg radius** | The percentage of gaze data that lies within a circle with 8° radius according to Victor et al. (2005). The data are centred around the median value in the horizontal and vertical gaze direction (separately). In the sketch below PRC is the number of gaze data points (dots) inside the circle divided by the total number of dots. |
| **mean AttenD** | The mean AttenD buffer value in the current segment. AttenD is an algorithm that estimates how attentive the driver is based on where he/she is looking. The algorithm is implemented according to Kircher and Ahlstrom (2012). In the sketch below the driver is considered attentive when the buffer is full. Every time the driver looks away the buffer is decreased and if it reaches zero (stars) the driver is considered inattentive. |
### # AttenD warning
The number of times that the driver is considered inattentive according to the AttenD algorithm (Kircher & Ahlstrom, 2012). Note that the driver is not actually given a warning, it is just the occasions where a warning should have been given that are indicated by this PI.

### energy radius
The radius of an ellipse covering 90% of the gaze data. The radius $r$ defines the horizontal radius of the ellipse whereas the vertical radius is defined as $r/2$. In the sketch below, the energy radius equals the $r$-value that defines an ellipse that covers 90% of the dots.

![Energy radius diagram](image)

### longest glance
The time duration of the longest consecutive sequence of gaze data that is directed outside a circle with 20° radius. In the sketch below there are four sequences of data outside the circle (light grey part of the scan path). The longest sequence outside the circle is the longest glance.

![Longest glance diagram](image)

### eyes-off-road glances
The number of glances that are directed outside a circle with 20° radius. In the sketch above there are four eyes-off-road glances.

### eyes-off-road time
The percentage of accumulated time that the gaze is directed outside a circle with 20° radius.

### throttle hold rate
Throttle hold rate is calculated based on the position of the accelerator pedal (measured in per cent) according to Green et al. (2007). If the throttle position deviates with less than 0.2% in a one-second wide sliding window, a throttle hold is marked at this particular window location. The throttle hold rate is then determined as the percentage of throttle hold instances within the current segment.

### throttle hold diff
Similar to throttle hold rate, but throttle hold diff is defined as when the differentiated throttle position is zero.

### stw reversal rate
The number of steering wheel reversals larger than 1° (Östlund et al., 2005). Before counting reversals, the steering wheel angle is low pass filtered with a 0.6 Hz 2nd order low pass Butterworth filter (also SWRR).

### stw entropy
Steering wheel entropy represents the predictability of steering wheel movements (Boer, Rakauskas, Ward, & Goodrich, 2005). The steering wheel signal is down-sampled to 4 Hz and the residuals from a one-step prediction based on autoregressive modelling are used to determine the entropy.
5 Results and interpretation

The results presentation in this report focus on the comparison of the PIs for the two different simulators used as well as on the usefulness of the PIs as sensors for secondary task difficulty. Further more detailed analyses will be presented within a follow-up project.

5.1 Baseline comparisons

Each participant completed four baseline drives, two with a lead car and two without. One of each was completed before the secondary task blocks, and one of each after the secondary task blocks at the end of the experiment.

For each PI a repeated measures GLM (general linear model) with the within-subjects factors baseline (beginning or end of session), level (from which 30-second segment of the baseline the data were collected) and lead car (present or not), and the between-subjects factor location (VTI or VCC) was computed.

Generally it can be noted that for the driving behaviour indicators the results differ between locations mostly for the lateral PIs (see Table 7). There are differences for almost all factor combinations, and in addition, most of the PIs differ during the first and second baseline completion for the VCC participants. For the longitudinal PIs the only difference between locations was related to lead car presence. Here, the presence of a lead car significantly influenced a number of PIs only for participants at VTI, who generally drove at higher speeds without a lead car present. Otherwise, the longitudinal driving PIs did not differ between locations. Only at VTI it was found that the participants drove somewhat faster during the second baseline, this was especially true for the condition without lead car. For the factor level and the lateral PIs, the second data segment (Level 2 for the factor level) was different from the others at both locations, the difference was especially clear during the late baseline runs at VCC.
Table 7: F-values and levels of significance for the driving behaviour PIs for baseline comparisons. Within each row the upper values indicate the VTI results and the lower values indicate the VCC results. When a box is filled with grey, there is a significant difference ($p < .05$) between VTI and VCC for the factor combination in question. Rows marked with red indicate PIs that relate to traffic code violations.

(*): $p < .10$; *: $p < .05$; **: $p < .01$.

<table>
<thead>
<tr>
<th>PI type</th>
<th>PI</th>
<th>Baseline</th>
<th>Level</th>
<th>Lead car</th>
<th>Baseline x level</th>
<th>Baseline x lead car</th>
<th>Level x lead car</th>
</tr>
</thead>
<tbody>
<tr>
<td>driving behaviour – longitudinal control – rel. to road</td>
<td>mean speed</td>
<td>3.7(*)</td>
<td>26.6**</td>
<td>3.4(*)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>driving behaviour – longitudinal control – rel. to road</td>
<td>max speed</td>
<td>4.0(*)</td>
<td>13.9**</td>
<td>7.8*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>driving behaviour – longitudinal control – rel. to road</td>
<td>std speed</td>
<td>26.5**</td>
<td>5.3(*)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>driving behaviour – longitudinal control – rel. to road</td>
<td>speeding index</td>
<td>4.5*</td>
<td>31.9**</td>
<td>3.2(*)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>driving behaviour – longitudinal control – rel. other road users</td>
<td>speeding area</td>
<td>6.7*</td>
<td>5.5*</td>
<td>3.5(*)</td>
<td>5.6*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>driving behaviour – longitudinal control – rel. to road</td>
<td>mean THW</td>
<td>19.4**</td>
<td>2.7(*)</td>
<td>2.6(*)</td>
<td>2.6(*)</td>
<td>2.6(*)</td>
<td>2.6(*)</td>
</tr>
<tr>
<td>driving behaviour – longitudinal control – rel. to road</td>
<td>min THW</td>
<td>18.5**</td>
<td>2.7(*)</td>
<td>2.6(*)</td>
<td>2.6(*)</td>
<td>2.6(*)</td>
<td>2.6(*)</td>
</tr>
<tr>
<td>driving behaviour – longitudinal control – rel. to road</td>
<td>std THW</td>
<td>21.7**</td>
<td>3.1(*)</td>
<td>5.6**</td>
<td>8.6**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>driving behaviour – longitudinal control – rel. to road</td>
<td>mean lateral position</td>
<td>21.7**</td>
<td>3.1(*)</td>
<td>5.6**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>driving behaviour – longitudinal control – rel. to road</td>
<td>std lateral position</td>
<td>13.0**</td>
<td>4.6*</td>
<td>15.3**</td>
<td>2.9(*)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>driving behaviour – longitudinal control – rel. to road</td>
<td>mod std lateral position</td>
<td>272.7**</td>
<td>12.0**</td>
<td>4.1(*)</td>
<td>6.2*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>driving behaviour – longitudinal control – rel. to road</td>
<td># lane crossings</td>
<td>15.0**</td>
<td>7.4**</td>
<td>5.6*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>driving behaviour – lateral control – rel. to road</td>
<td>lane departure area</td>
<td>9.9*</td>
<td>6.2*</td>
<td>3.6(*)</td>
<td>4.5(*)</td>
<td>4.8(*)</td>
<td></td>
</tr>
<tr>
<td>driving behaviour – lateral control – rel. to road</td>
<td>lane departure time</td>
<td>10.1*</td>
<td>6.8*</td>
<td>4.7(*)</td>
<td>4.2*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>driving behaviour – lateral control – rel. to road</td>
<td>lane departure distance</td>
<td>10.5*</td>
<td>6.9*</td>
<td>4.5(*)</td>
<td>4.1(*)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
For the driver behaviour related indicators the picture was not quite as clear-cut (Table 8). The PIs that were most consistent between the two locations were energy radius and longest glance, as well as the throttle hold variables. Some of the other eye tracking variables differed between the locations, with VCC participants generally being more likely to look away from the forward roadway. The PIs related to lateral control differed as well, with VCC arriving at higher values for Steering Wheel Reversal Rate and Steering Wheel Entropy.

Table 8: F-values and levels of significance for the driver behaviour PIs for baseline comparisons. Within each row the upper values indicate the VTI results and the lower values indicate the VCC results. When a box is filled with grey, there is a significant difference (p < .05) between VTI and VCC for the factor combination in question.

(*): p < .10; *: p < .05; **: p < .01.

<table>
<thead>
<tr>
<th>PI Type</th>
<th>PI</th>
<th>Base-line</th>
<th>Level</th>
<th>Lead car</th>
<th>Base-line x level</th>
<th>Base-line x lead car</th>
<th>Level x lead car</th>
</tr>
</thead>
<tbody>
<tr>
<td>PRC 8 deg radius</td>
<td>14.5**</td>
<td>6.8*</td>
<td>7.4(*)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mean AttenD</td>
<td>4.7*</td>
<td>4.0(*)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td># AttenD warning</td>
<td>no AttenD warnings occurred at VTI during baseline</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>energy radius</td>
<td>8.3**</td>
<td>4.1*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>longest glance</td>
<td>7.8(*)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>eyes-off-road glance</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>eyes-off-road time</td>
<td>3.5(*)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>throttle hold rate</td>
<td>6.5**</td>
<td>30.2**</td>
<td>11.6**</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>throttle hold diff</td>
<td>25.2**</td>
<td>10.5*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Steering wheel reversal rate</td>
<td>84.9**</td>
<td>87.8**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Steering wheel entropy</td>
<td>6.6**</td>
<td>2.7(*)</td>
<td>4.2*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
5.1.1 Interpretation of results for the baseline comparison

The situation would have been ideal if there had been no differences between the locations for any factor combination, and if the factor baseline and the factor level had had no significant influence on the results. The null hypothesis was formulated based on this ideal situation. For both driving situations it was assumed that driving parameters will not differ between the first and the second set of baselines. This would have meant that the simulators (and participants) were completely comparable, that the participants did not change their behaviour over the course of the experiment, and that the test route was consistent in terms of demand. This latter aspect was controlled for by the factor level, which, in the case of the baseline analysis, only indicated from which 30-second segment of the route the log data were sampled – as there were no secondary tasks present during baseline, everything else was equal.

The actual results look different. The most striking difference is the dissimilarity of results for the lateral driving behaviour variables. It has already been shown by Greenberg et al. (2003) that the presence of motion cueing has a significant effect especially on heading control, with fixed-base simulators requiring a larger effort on the part of the driver. The strength of the influence of the absence or presence of motion cueing on performance depended on the secondary task type. Therefore, even the relative comparisons based on findings from a fixed-base simulator are questioned. The present study extends Greenberg et al.’s findings such that, for the curvy road used in the experiment, lateral control related differences can already be found for baseline driving.

In addition to the differences found between locations, for VTI the longitudinal results differed between the early and the late baseline drives, and for VCC the lateral results differed. The increase in speeds in the second baseline drive at VTI can be interpreted as the participants’ getting used to the setting. The anomaly for the lateral PIs, however, for which the second data segment differed markedly from the first and the third segments, is more difficult to explain. One possibility might be that the curvature of the road differed in the second segment. A driver covers 750 m within 30 s when keeping a speed of 90 km/h. At VTI the curve radius was constant at 1000 m, but changed sign after having covered a stretch of 90 degrees corresponding to a curve length of 1570 m, i.e. one of the three level segments might have had more changes of direction than the other two segments. On the other hand, data collection was started only when the drivers had reached a speed above 75 km/h. The time needed to reach this speed varied between participants. At VCC the curve radius and curve length was not constant, making the similarity of the result anomaly across locations even harder to explain.

5.2 Secondary task performance

Secondary task performance was investigated for all three tasks. The effect of location (VTI versus VCC simulator), driving condition (whether the participants drove with a lead car, without a lead car, or sat in a stationary vehicle), and task level (whether the secondary task was easy, medium or difficult) on secondary task performance was investigated. Statistical analyses were made based on the general linear model (GLM) with repeated measures, with location as between-subjects factor.
5.2.1 Visual task

For the visual task three performance indicators were calculated across all participants. Per condition the percentage of answered stimuli and the percentage of correct answers were counted, and the reaction time for correct answers only was computed.

The participants at VCC answered significantly more stimuli than the VTI participants (F(1, 306) = 5.0; p < .05), and the number of answered stimuli was smaller for the most difficult task level (F(2, 306) = 4.0; p < .05). The condition under which the stimuli appeared did not significantly influence the likelihood to answer (see Figure 7).

![Figure 7](image)

Figure 7: Percentage answers to stimuli for the visual task across all participants, for both locations, all driving conditions and all task difficulty levels.

As shown in Figure 8, the VCC participants made very few errors in general for those stimuli that were answered, almost regardless of difficulty level and condition. At VTI, however, the difficulty level clearly influenced the number of correctly answered stimuli, and the interaction between location and level was significant (F(4, 306) = 4.3; p < .05). Driving condition did not influence the results at any of the locations, or for any level.
VTI

VCC

Figure 8: Percentage correct answers of all answered stimuli for the visual task across all participants, for both locations, all driving conditions and all task difficulty levels.

VCC participants were also significantly faster with their correct answers (F(1, 289) = 36.3; p < .05) than the participants at VTI (Figure 9). The driving condition did not influence task completion times, but there was a very clear effect of task difficulty level (F(2, 289) = 94.6; p < .05), with task completion times being doubled from the easiest to the most difficult level. An interaction of level with location could be observed, showing that the task completion time for VTI participants was particularly long for the most difficult level (F(2, 289) = 3.7; p < .05.

Figure 9: Mean task completion times (s) for the correct answers on the visual task across all participants, for both locations, all driving conditions and all task difficulty levels.
5.2.2 Interpretation of results on the visual task

The VCC participants answered more stimuli, answered them more correctly and faster. For both test sites, generally speaking, it did not matter whether the participants were driving or not while executing the task. This indicates that the participants followed the instructions and tried to maximise their performance on the secondary task. The reduced number of correct answers at VTI for higher difficulty levels, and the prolonged task completion times for more difficult tasks, which could also be observed in the baseline condition, indicate that the task difficulty itself was the limiting factor, and not the interaction with driving.

5.2.3 Cognitive task

The results for the cognitive task are shown in Figure 10. The statistical analysis showed that location had a significant influence on the number of correct answers on the cognitive task (F(1, 306) = 17.7; p < .05). The VCC participants answered correctly more often than the VTI participants. No interaction effects of other factors with location existed.

The factor task level had a significant influence on the number of correct answers (F(2, 306) = 6.2; p < .05). A higher task difficulty led to a lower number of correct answers, and post-hoc tests with Bonferroni correction showed that the only non-significant level difference was between easy and medium task difficulty in both locations.

![Figure 10](image_url)

*Figure 10: Number of correct answers on the cognitive task across all participants, for both locations, all driving conditions and all task difficulty levels.*

5.2.4 Interpretation of results on the cognitive task

While the most difficult task level led to significantly fewer correct answers, the driving condition did not play a significant role in task performance. Like for the visual task, the main difficulty appears to lie within the task, and not in the addition of the driving task. Again, the VCC participants were better on average at solving the task than the VTI participants.
5.2.5 Haptic task

To determine the performance on the haptic task, the number of correctly placed bricks was counted. From this number the number of incorrectly placed bricks was subtracted. An incorrectly placed brick is a brick that was placed in the correct hole, but in the wrong order. The number of incorrectly placed bricks was 12, with 1418 correctly placed bricks, equalling a wrong placement quote of 0.8%.

As before, VCC participants were clearly more proficient at executing the task (F(1, 306) = 68.7; p < .05). The more difficult the task level, the lower was the number of correctly placed bricks (F(2, 306) = 16.7). Again, the driving condition did not have a significant influence on the secondary task performance (Figure 11).

![Figure 11: Number of correctly placed bricks minus the number of incorrectly placed bricks across all participants, for both locations, all driving conditions and all task difficulty levels.](image)

5.2.6 Interpretation of results on the haptic task

The performance on the haptic task was quite equal to the performance on the cognitive task. The inherent task difficulty had a greater effect on performance than the addition of the driving task, but the driving condition did not influence the secondary task performance.

5.2.7 Discussion and interpretation of secondary task performance

Even though not all differences were statistically significant, there was a clear effect of task difficulty. This could be observed across all driving conditions, including baseline secondary task performance, where the participants could focus all of their attention on the “secondary” task. This implies that the secondary tasks were appropriately selected, and neither floor nor ceiling effects were observed on average. As for all results, the individual variations were quite large, however, so for testing purposes it might be worth considering to adapt the difficulty level of the secondary task to the performance capabilities of each individual driver. In this case a training to asymptote including a calibration of task difficulty would have to
precede the actual testing. While this is time consuming and demands secondary tasks that are freely adjustable, it would help cancelling out inter-individual differences.

Generally, the data show that task level had a much higher influence on secondary task performance than driving condition, which did not lead to any significant changes in any of the secondary task performance indicators. While it is necessary for a full interpretation to consider how driving behaviour was affected, it can be stated that executing the driving task in addition to the secondary task did not affect secondary task performance substantially, either because it did not interfere with the secondary task to begin with, or because the participants increased the effort invested in the secondary task, while they were still able to perform the driving task. It has to be kept in mind that the driving task was a very simple tracking task, which is likely to have reached a high level of automation in experienced drivers. Therefore, the results should not be generalised to more difficult driving tasks.

Interestingly, the participants at VCC were more proficient at all secondary task types, and for all PIs computed. As VCC participants were recruited in-house from the research and development department, it can be assumed that their average level of education was higher than that of the VTI participants, which could explain the offset in performance. In most cases the offset between the participant groups was relatively linear for the different task levels and driving conditions, but for some PIs even the most difficult task levels were apparently not difficult enough for the VCC group to affect task performance. Within a range of task difficulties it should be possible to draw relative conclusions about task level based on the performance of different sub-groups.

One important conclusion from the secondary task performance results is that the participants appeared to have followed the instruction to give priority to the secondary task. The performance decrements seem to depend on task inherent factors and not on the addition of the driving task. This is promising, as it will allow an interpretation of the driving data based on the assumption that the participants did what they could to maximise secondary task performance.

5.3 Performance ratings

The participants rated their performance after each section. Secondary task performance was self-rated for the most difficult level, and driving performance was rated for the time during which the secondary task was executed on the most difficult level. During the sections in which the vehicle was stationary only secondary task performance was rated, and during the driving baseline sections only driving performance was rated.

Self-rated driving performance (Figure 12) was influenced significantly by which type of secondary task that was executed concurrently (VTI: F(4, 84) = 50.0, p < .05; VCC: F(4, 40) = 19.0, p < .05). For the VTI-participants, concurrent performance of the visual task made the participants rate their driving performance significantly worse than for other concurrent tasks or no secondary task. There was a tendency for the participants to rate their driving performance as better when a lead car was present (F(1, 21) = 3.7, p < .10). At VCC the order of ratings equalled the VTI order, but with the driving performance rated as somewhat better during the visual task as compared to VTI. The lead car tendency did not exist at VCC.
Figure 12: Self-rated performance on the driving task, depending on the concurrently executed secondary task (most difficult level) and lead car presence.

The subjective ratings of secondary task performance always refer to the most difficult task level. At VTI, the main effect of driving condition ($F(2, 40) = 3.8, p < .05$), the main effect of secondary task type ($F(2, 40) = 22.5, p < .05$), and the interaction between the two ($F(4, 80) = 4.0, p < .05$) were all significant. As illustrated in Figure 13 (VTI), the participants rated their performance on the visual task as best while driving with a lead car, followed by baseline, followed by driving without a lead car. For the other two tasks the ratings were quite similar to each other, with the performance in the lead car condition receiving the highest values, followed by the condition without lead car. Participants gave themselves the lowest grades for secondary task performance while the vehicle was stationary. For VCC the overall rating pattern was similar, but with smaller differences between driving conditions and secondary task types. There were no significant differences between the ratings, except for a trend for an interaction between the two factors ($F(4, 44) = 2.5, p < .10$).

Figure 13: Self-rated secondary task performance on the most difficult level, depending on secondary task type and driving condition.
For those sections in which driving and secondary task performance occurred concurrently, the last ten VTI participants in the study as well as all VCC participants were asked to rate how they distributed their efforts between the two, given that they had 100% effort to distribute. For all task types the secondary task received more effort than the driving task, but at VTI with significant differences between task types (F(2, 16) = 3.9, p < .05). The visual task was given the highest mental effort, followed by the cognitive task, and then the haptic task. Also driving condition had a significant effect on invested effort (F(1, 8) = 7.0, p < .05), which has to be differentiated by task type. When no lead car was present, the visual and the haptic task received a higher share of the total effort than during lead car presence (see Figure 14). This difference could not be observed for the cognitive task. At VCC the mean effort distribution ratings were very similar, but without being significantly different. There was a trend for the secondary task type (F(1, 22) = 3.1, p < .10), also here with the visual task receiving the largest perceived effort.

5.3.1 Interpretation of subjective ratings

The subjective ratings of task performance and invested effort follow a logical pattern, and the ratings are very similar between the two test sites, even though significances are not always equal. The visual task received the highest mental effort, which led to a perceived relatively high performance on the secondary task, while the perceived performance on the driving task was rather bad. Corresponding reasoning holds for the cognitive and the haptic tasks.

It is remarkable, however, that the secondary task performance ratings were lower for baseline performance, that is, without driving, than when the participants drove at the same time. The only exception is the visual task, where baseline performance was rated above secondary task performance while driving without lead car at VTI, and better than both driving conditions at VCC. One interpretation might be that the participants set their task performance in relation to what they thought was achievable in a given situation. They might have expected to perform better when they did not have to drive at the same time. In this case, the same absolute performance would be rated differently depending on the surrounding circumstances.
There is some support for this assumption in the actual secondary task performance scores, where baseline performance usually is equal to or above secondary task performance with concurrent driving.

5.4 Driving behaviour

The PIs based on continuous measures were classified into driving behaviour and driver behaviour, with several sub-categories (see Table 5). In Table 9 an overview of the F-values and significances of the driving behaviour related PI categories is presented.
Table 9: F-values and significances for all tested driving behaviour PIs for main effects and 2-way interactions. Within each row the upper values indicate the VTI results and the lower values indicate the VCC results. When a box is filled with grey, there is a significant difference (p < .05) between VTI and VCC for the factor combination in question. Red bars indicate PIs related to a traffic rule violation.

(*): p < .10; *: p < .05; **: p < .01.

<table>
<thead>
<tr>
<th>PI Type</th>
<th>PI</th>
<th>Modality</th>
<th>Level</th>
<th>Lead car</th>
<th>Modality x level</th>
<th>Modality x lead car</th>
<th>Level x lead car</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mean speed</td>
<td>17.0**</td>
<td>29.0**</td>
<td>6.2*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>max speed</td>
<td>10.5**</td>
<td>23.1**</td>
<td>5.8*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>std speed</td>
<td>3.3*</td>
<td>5.9*</td>
<td>4.9*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>speeding index</td>
<td>16.1**</td>
<td>31.8**</td>
<td>6.5*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>speeding area</td>
<td>3.6*</td>
<td>8.6**</td>
<td></td>
<td></td>
<td></td>
<td>2.9(*)</td>
</tr>
<tr>
<td></td>
<td>mean THW</td>
<td>n/a</td>
<td></td>
<td></td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td></td>
<td>min THW</td>
<td>n/a</td>
<td></td>
<td></td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td></td>
<td>std THW</td>
<td>4.9*</td>
<td>3.9*</td>
<td></td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td></td>
<td>mean lateral position</td>
<td>3.5*</td>
<td>19.7**</td>
<td></td>
<td>6.0*</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>std lateral position</td>
<td>27.2**</td>
<td>10.0**</td>
<td>11.5**</td>
<td>4.1*</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>mod std lateral position</td>
<td>49.9**</td>
<td>9.1**</td>
<td>12.7**</td>
<td>8.3**</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td># lane crossings</td>
<td>20.3**</td>
<td>22.9**</td>
<td>2.5(*)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>lane departure area</td>
<td>11.6**</td>
<td>3.4(*)</td>
<td></td>
<td>17.8**</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>lane departure time</td>
<td>18.2**</td>
<td>11.4**</td>
<td>2.3(*)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>lane departure distance</td>
<td>15.9**</td>
<td>11.4**</td>
<td>2.3(*)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Instead of reporting the mean values for each individual PI an analysis of patterns is made. For longitudinal control in relation to the road, that is variables related to speed, there is a clear pattern that the modality of the secondary task as well as the lead car have an effect. The presence of a lead vehicle reduces the mean and max speed with approximately 4 km/h, regardless of modality. The secondary task modality, on the other hand, influences speed such that a reduction with about 5 km/h is found for the visual secondary task, while for the other modalities, including baseline driving, the mean and max speed values are similar to each other. The same is true for the speeding index and the speeding area.

The picture looked slightly different for the standard deviation of speed, which was lower for baseline driving and concurrent cognitive task performance (ca. 3 km/h), and higher for concurrent visual and haptic task performance (ca. 4 km/h), but the only significant difference was between baseline driving and driving with the concurrent haptic task. Especially at higher task difficulty levels the presence of a lead car helped to reduce the speed variance significantly.

During baseline driving and with the visual task active, participants drove about 10 cm further to the outer road edge than with the cognitive and the haptic task. On the easiest task level they drove the furthest to the right on average, on the medium task level they drove the furthest to the left, and on the most difficult task level in between the two. On the other hand, the standard deviation of lateral position (SDLP) was much larger during the visual task (11-17 cm) than in the other three conditions (6-8 cm), and with the visual task SDLP increased markedly with increasing task difficulty. For all modalities the SDLP was slightly attenuated by the presence of a lead car.

The PIs connected to the traffic violation “lane departure” show that, the number of lane departures per 30 s of secondary task performance, the distance driven with the left wheel pair in the oncoming lane, the time spent with the left wheel pair in the oncoming lane, and the area between the left wheel pair and the lane marking, were significantly larger for concurrent visual task performance than for baseline driving, cognitive or haptic task performance. The time and distance spent in the oncoming lane was largest for the medium secondary task level, with almost five out of 30 s and around 110 m out of approximately 670 m in the oncoming lane for the visual task. For the visual task the most difficult task level produced similar values, while the most difficult task level produced values in the range of baseline driving for the other modalities. A glance at the lane departure area reveals that only the visual task, on medium and difficult task level, leads to more than one square metre covered in the oncoming field, with the most difficult level leading to the highest average value. The spread between participants is large, and while most of them lie far below one square metre of lane departure area, there is one occurrence where more than 30 square metres are covered in the oncoming lane, and two occurrences where this value lies at above 15 square metres.

With regard to task difficulty, for the PIs lane departure distance, number of line crossings and lane departure time it was found, somewhat unexpectedly, that the medium task level shows the largest deviations from baseline, while the values for the highest task level are closer to baseline values again, with the medium task level being significantly different from the easy task level, the most difficult task level, or both.

The only PI related to other road users that was investigated was time headway. Here, comparisons were only made for modality and level as a lead car must be present for time headway to be meaningful. No significant influence of either modality or level of task difficulty could be found for the time headway related PIs under investigation.
5.5 Driver behaviour

Table 10 shows an overview of the driver behaviour related PIs. They describe gaze behavior, the interaction with the gas pedal and with the steering wheel. Generally, a number of gaze related PIs differed between locations, with VCC participants typically showing longer and more frequent glances away from the forward roadway, leading to a lower mean AttenD value and a higher number of “would-be” AttenD warnings. This was especially noticeable for the visual task. While the throttle hold related PIs did not differ substantially between locations, there were major differences for this factor for the steering wheel related PIs.

Table 10: F-values and significances for all tested driver behaviour PIs for main effects and 2-way interactions. Within each row the upper values indicate the VTI results and the lower values indicate the VCC results. When a box is filled with grey, there is a significant difference (p < .05) between VTI and VCC for the factor combination in question. (*) = p < .10; * = p < .05; ** = p < .01.

<table>
<thead>
<tr>
<th>PI Type</th>
<th>PI</th>
<th>Modal-ity</th>
<th>Level</th>
<th>Lead car</th>
<th>Modal-ity x level</th>
<th>Modal-ity x lead car</th>
<th>Level x lead car</th>
</tr>
</thead>
<tbody>
<tr>
<td>Driver behaviour - driver internal – gaze direction</td>
<td>PRC 8 degree radius</td>
<td>67.6**</td>
<td>3.9*</td>
<td>8.7*</td>
<td>6.0**</td>
<td>12.4**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>mean AttenD</td>
<td>80.0**</td>
<td>25.0**</td>
<td>48.8**</td>
<td>5.0*</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td># AttenD warning</td>
<td>24.2**</td>
<td>23.7**</td>
<td>23.7**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>energy radius</td>
<td>138.7**</td>
<td>7.0**</td>
<td>4.1*</td>
<td>4.6**</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>longest glance</td>
<td>277.8**</td>
<td>12.6**</td>
<td>13.9**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>eyes-off-road glance</td>
<td>89.2**</td>
<td>11.3*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>eyes-off-road time</td>
<td>148.6**</td>
<td>44.5**</td>
<td>45.8**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>throttle hold rate</td>
<td>9.0**</td>
<td>3.5*</td>
<td>26.4**</td>
<td>4.4**</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>throttle hold diff</td>
<td>5.9**</td>
<td>3.7*</td>
<td>27.8**</td>
<td>2.8*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>driver behaviour - driver internal – steering wheel</td>
<td>stw reversal rate</td>
<td>30.3**</td>
<td>4.7*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>stw entropy</td>
<td>4.8**</td>
<td>2.9(*)</td>
<td>5.0**</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

ViP publication 2014-1
The longest single glance duration was on average above 2 s for the medium task and the difficult task, and the mean eyes-off-road time within the 30 s of dual task performance lay above 10 s for those two task levels. One participant looked away for 20 s out of 30 s on the most difficult level. Generally it was obvious that the visual secondary task led to much longer glances and more intensive visual time sharing than the other two task types. With respect to the gaze related PIs, the task level had its strongest effect for the visual task, leading to significant interactions between modality and level for almost all PIs in both locations. The significant effect of level alone for the VTI participants can be explained by the strong effect of the visual task.

The steering wheel reversal rate was influenced mostly by modality and location, and at VCC also by the task level. At VTI it was significantly lower for baseline driving than when driving with a concurrent secondary task, while at VCC the reversal rate especially during baseline was much higher than at VTI. The presence of a lead car led to a higher steering wheel reversal rate.

While steering wheel reversal rate indicates how often the steering wheel is moved back and forth, steering entropy gives information about the size, the speed and the randomness of the movements. A higher steering entropy is viewed as a sign of decreased control. Interestingly, baseline driving and the cognitive task produced the highest steering entropy, mainly on the medium and difficult secondary task levels, while values were lower for the visual task and the haptic task. The absolute differences between the values were quite small, however, lying between 1.53 and 1.57, which corresponds to a maximum difference of 2 %. For steering entropy, just as for a number of driving behaviour PIs related to lateral control, it was found that there was a tendency for the medium task level to produce the greatest value, with the easy and the difficult task levels both leading to smaller values.

### 5.6 Interpretation of driver and driving behaviour results

**Location.** For the lateral PIs there were substantial differences between locations, while the longitudinal PIs did not differ that much. Gaze behaviour showed some differences between the locations. The differences in lateral behaviour are similar to what has already been found in the baseline comparisons, which supports the explanation suggested by Greenberg et al. (2003). However, the performance on the secondary tasks also indicates that the participants at VCC generally were more proficient at those tasks. This can either mean that they generally have a higher level of capability, or that they put more effort into the secondary task. The former would lead to expectations of a better or equal driving performance compared to the VTI participants, the latter to a degraded performance as compared to VTI participants.

Driving performance may, as mentioned, be influenced by the type of simulator also, which makes it difficult to resolve the confounding variables. It shows the importance of selecting participants from comparable pools when running experiments on different platforms.

**Modality.** In general the visual task seems to have the greatest impact on the different PIs by far. This is the only task where drivers actually have to look away from the forward roadway for substantial periods of time in order to be able to solve the task, which is also reflected by the PIs related to glance behaviour. Lateral vehicle control was clearly diminished, not only such that the SDLP increased, but for a number of participants also such that the vehicle ended up in the oncoming lane for substantial amounts of time, distance and area. For the longitudinal PIs the visual task led to a reduced mean speed (at VTI) and a slightly increased standard deviation of speed. Whether the decrease in speed was voluntary or involuntary cannot be determined from the data at hand. Both possibilities are thinkable. Either, drivers
lose control, just as they quite obviously do for the lateral positioning, or they compensate for having to focus on something else, as they know that a lower speed will provide greater margins for correcting other mistakes.

Driver behaviour PIs related to vehicle handling did not show clear effects, except that the steering wheel reversal rate clearly increased for all secondary tasks, an effect that had already been observed in the HASTE project for visual and cognitive tasks (Östlund et al., 2004).

While a number of PIs showed effects of the visual task, the picture was less clear for the cognitive and the haptic tasks. For those modalities most PIs did not show significant differences from baseline at all. Still, participants felt that their driving performance during concurrent performance of the haptic and cognitive tasks decreased slightly as compared to baseline, even though not as much as for the visual secondary task. This indicates that an experienced decline in performance might not immediately be measurable in the PIs used here. It might well be measurable in other PIs, however, or it might only show in more critical situations.

Level. The level of task difficulty had a clear effect on driving PIs related to lateral control, and on driver performance with respect to eye movements and, to a lesser extent, the handling of the vehicle controls. In most cases an interaction effect with modality was present. For some variables that meant that particularly the visual task affected driving and driver performance more with increasing task difficulty, while driving and driver performance were not affected for the other modalities. In other cases it meant that, while growing performance decrements could be found for the visual task, for the other tasks the lowest performance occurred at the medium task level.

Lead Car. It is not surprising that the lead car influenced PIs that are connected to longitudinal control. The lead vehicle at VTI held a lower mean speed than the lead vehicle at VCC, which made its effect on mean speed and max speed stronger at VTI.

5.7 Inter-individual differences

It is very common to compare the mean values when investigating differences between groups, just as was done in the present report. There are, however, several caveats with doing so, some of which we illustrate here. Especially when the values are not normally distributed outliers can have a big influence on the mean value as compared to the median. In addition, actual values can be widely spread, and it is not always the mean that is of highest interest. In Figure 15 (showing lane departure distance) all of these aspects are illustrated. With a concurrent visual task on the easy level half of the participants or more did not leave their lane at all. However, in the case of no lead vehicle two participants left their lane for distances of 300 m and more, and another two had departure distances of more than 100 m. With a lead vehicle present, one participant exceeded his lane for almost 600 m on the easiest level, another with almost 200 m. It is mainly because of these participants that the mean value ends up at around 50 m, but it does not really represent the average behaviour.

While the inclusion of “participant” as a factor allows keeping track of whether one person has a systematic offset from the mean or not, for this report we analysed each PI separately. Therefore, the analyses do not show whether there are certain “types” of participants that generally tend to be outliers, or otherwise perform in a systematically different way. That would be of great interest, however, especially if a driver’s perceived invested effort in the
secondary task, the performance on the secondary task and the driving performance can be viewed in combination.

In Figure 15 it is also illustrated that the data can look different for different measuring platforms. Here, lane departure distance was measured both in a moving-base simulator (VTI, red colour) and a fixed-base simulator (VCC, blue). For the visual task, especially with a lead vehicle present, the data patterns look different for the two test sites, and for the haptic task the sticking out effect of the medium difficulty level is much stronger at VCC than at VTI. This is a reminder of how important it is to consider the constraints of the given test scenario when interpreting data.

![Figure 15: Lane departure distance (m) for individual VTI (red) and VCC (blue) participants. Mean (fat solid line) and median (fat dashed line) values for all secondary task conditions, with and without lead car, and for all task difficulty levels (1, 2, and 3).](image)

Another example is the PI PRC (Figure 16). This PI is based on the same sensor at the two test sites, the Smart Eye Pro eye tracking system, and does not use any simulator data. Still, a systematic offset is visible between the two locations, which indicates more extreme mean values for VCC than for VTI. While the VCC participants have higher PRC values for the baseline and the cognitive and haptic tasks, they have even lower values than the VTI participants for the visual task. This divergence may have to do with tracking quality, but it is also possible that the differences between the two test groups, discussed above, have an influence on the results.
Figure 16: Percent road centre (PRC) with a radius of 8° for individual VTI (red) and VCC (blue) participants. Mean (fat solid line) and median (fat dashed line) values for all secondary task conditions, with and without lead car, and for all task difficulty levels (1, 2, and 3).
6 Discussion

The goal of this study was to investigate commonly used PIs for their sensitivity to different levels of difficulty of secondary tasks loading on different modalities, and to evaluate their robustness across platforms.

Due to the physical distance of almost 300 km between the two simulators it was not feasible to run the same drivers on both platforms, which resulted in a confounding between the drivers and platforms. This is aggravated by different recruiting procedures; the drivers driving the moving-base simulator were recruited from the general public, while the drivers in the fixed-base simulator were recruited in-house and consisted of a group of engineers.

Generally speaking, the longitudinal PIs behaved more similarly across platforms, while the lateral control was much more impaired in the fixed-base simulator. It has already been indicated by Greenberg et al. (2003) that motion cueing may have a strong impact on lateral driving PIs when disturbances are present. The present study suffers from similar limitations as the Greenberg study, with a between-subjects design for platforms and a smaller number of drivers in the fixed-base condition. But the accumulated evidence still demonstrates that lateral motion cues seem to have a big impact on the drivers’ tracking capabilities. While a validation of the moving-base simulator against real driving is still required, caution should definitely be exerted in the interpretation of outcomes based on lateral PIs in fixed-base simulators. It can be assumed that the lane tracking part of driving is relatively highly automated for experienced drivers. Therefore it is unlikely that tasks that do not require extended glances away from the forward roadway should lead to decreased lateral control of the extent observed here. The road layout may have encouraged drivers to cut corners in left curves, which may have led to some planned lane departures, but the other lateral PIs also indicated a decrease in control which cannot necessarily be attributed to planned behaviour.

In the literature evidence can be found that secondary task load influences driving behaviour differently, depending on the affected modality. Cognitive and visual load have been reported to lead to less respectively more variable lane tracking (Engström, Johansson, & Östlund, 2005). Not unexpectedly, eye movement based PIs have also been found to be affected by modality (Recarte & Nunes, 2000; Victor, Harbluk, & Engström, 2005). In the current study the effects on lateral position based PIs could not reliably be replicated due to the overwhelming platform effects. Lane departure distance, the modified SDLP and the number of lane crossings were, in fact, higher for the visual task than during baseline for both platforms, but for cognitive load the results were not as distinct. A reduction of lateral variation under cognitive load could not be observed for all affected PIs and for all task difficulty levels, and the effects differed between platforms. Also, the unexpected direction of the interaction with task difficulty made interpretations more difficult.

As pointed out by one of the reviewers, it is of great importance to be able to quantify the characteristics of the tasks on the four dimensions “visual”, “auditory”, “cognitive” and “psychomotor” demands. This could not be done within the budget for the current study, but is highly recommended for a follow-up study. Further information can be found in Yee et al. (2007).

It has become common to bundle modalities into “visual/haptic” and ”talking/listening”, that is, rather cognitive tasks. While it is true that haptic tasks often involve eye glances to aid the hand movements, the results shown here do not fully support a split along those lines. Here the haptic task was conducted without visual support, the eyes could therefore remain directed at the road. Most PIs showed more similarities between the cognitive task and the haptic task than between the visual task and the haptic task. Interestingly, the PIs that were affected most
similarly by the visual and the cognitive load were the lateral PIs in the fixed-base simulator, whose validity possibly is most questionable. Therefore, it is suggested that haptic tasks without visual component be treated separately, and if they prove not to be very detrimental their increased deployment in HMI solutions should be investigated.

The presence of the lead car had its biggest influence on speed, which is quite obvious as the mean speed of the lead vehicle was lower than the posted speed limit, thereby effectively limiting the drivers’ speed. Effects on other PIs were spurious at best, which indicates that in such a scenario a lead car does not influence behaviour as compared to driving in a free vehicle condition. The advantage of a lead car is that headway related PIs can be included in the driving performance assessment.

There are some concerns with the results that need to be highlighted. It is unexpected that the lateral PIs should show the most extreme values for the medium level of task difficulty, especially as this could also be observed for the baseline condition, in which no task at all was conducted. For this condition the log data from the corresponding time window were sampled, while the drivers did nothing but drove throughout the section. One possible explanation might be that the curvature of the road differed in the second segment. In 30 s, which is the duration of the secondary task execution, a driver covers 750 m when keeping a speed of 90 km/h. For the moving-base simulator the curve radius was constant at 1000 m, but changed sign after a stretch of 90 degrees, which corresponds to a curve length of 1570 m. Thus, one of the three level segments might have had more changes of direction than the other two segments. On the other hand, data collection was started only when the drivers had reached a speed above 75 km/h. The time needed to reach this speed varied between drivers. For the fixed-base simulator the curve radius and curve length were not constant, making the similarity of the result anomaly across platforms even harder to explain. For future work of the same kind it is therefore recommended to keep the curve radius and direction constant during secondary task activity to avoid possible confounding with road layout.

A further concern is the large intra-individual differences found across most of the investigated PIs, both between driving situations and task difficulty levels. While significant effects could be found for a number of factors, which shows that the variance between groups is larger than the error variance, there is still a substantial error variance for many PIs. This has several consequences. There may be drivers who can perform even the difficult secondary tasks while not degrading their driving performance, while others already falter at the easiest level. The data need to be analysed further to determine whether there are certain distinct groups of drivers, so-called “supertaskers” (Watson & Strayer, 2010), or whether lapses and successes are more idiosyncratic. It has to be noted, however, that in the Watson & Strayer study the frequency of supertaskers amongst a population of 200 university students was 2.5%, so for the present sample of 36 drivers roughly one supertasker could be expected. Still, a more individual analysis focusing on whether drivers adopt differing strategies when dividing their resources between the driving task and the secondary task would be of interest. This might shed some light on the question whether some people have greater abilities at multitasking and handle this in a qualitatively different way (true supertaskers), or whether they have more resources, and thus, only reach their limits for performance decrements later.

It may, however, also be the case that a driver is very good in one level, but fails in another, possibly due to lack of prolonged concentration, or for being in a more difficult road section during a certain task level, or for any other reason. In some cases the behaviour is quantitatively different, with some drivers having much higher values than others but with the general trend being equal, while for other PIs the differences appear to be of a qualitative nature. An example is lane departure distance, showing that some drivers do not leave their lane at all, while some drive outside of their lane for a substantial amount of time. This
behaviour could be recoded into “manages the task” vs. “does not manage the task”. When the goal is to investigate whether a system leads to potentially unsafe behaviour, the PI should possibly be the percentage of drivers who did not manage the task, instead of reporting for example the mean lane departure distance. Using mean values can be very misleading when dealing with distributions that have two or more peaks, or are otherwise not suitable to be represented by a mean value.

It should also be discussed what certain PIs actually mean. Here, just as in many studies in the literature, the major part of PIs describes the control level of driver behaviour (Michon, 1989). The reason for this is probably simple, i. e. those parameters can be logged in practically all driving scenarios, and they have been widely used. When evaluating the safety impact of a new system, however, what does it mean if drivers increase their SDLP with a few centimetres, even though the difference to baseline might be significant? It may be more important to know whether a new system will make drivers commit more violations, or whether a behaviour surfaces that definitely is safety critical. For this type of indicator it may be better to look at the tactical level of driving behaviour, where interactions with other road users play an important role. On this level it is also less likely that behaviour is highly automated, and a negative impact might be discovered earlier.

In the present study the road layout was very simple, and except for a car following situation no interactions with other road users were necessary. In a scenario like this the tactical judgements are limited and may be restricted to the decision whether to cut a corner or not, or whether to speed or not. A scenario of such limited complexity is easier to keep consistent, but may not be able to produce the desired situational complexity. When the goal is to study interactions and tactical behaviour, it might be more advisable to employ more complex scenarios, with the risk of losing some of the repeatability.

In a simple environment it may be the case that secondary tasks do not have an effect on driving behaviour, not even when task difficulty is increased. It is thinkable that the drivers still have enough resources to deal with the secondary task and drive satisfactorily nevertheless. There is a risk that those results may be generalised to more complex scenarios, in which the secondary task very well may have an influence on driving performance. For further studies that are intended to systematically investigate the effect of different secondary tasks on driving behaviour it is suggested to not only vary secondary task difficulty, but also the complexity of the situation, which can be achieved by manipulating the road layout, the traffic, the weather or all of the above.
7 Conclusions

Assessing the effects of secondary tasks on traffic safety via PIs that are based on variables describing the control level of driving behaviour is a difficult undertaking. Results may vary widely between drivers, and the implications of changes in some PIs on safety are not always clear. It is therefore suggested to investigate the possibility of using PIs on the tactical level of driving behaviour. This can include interactions with other road users, like overtaking manoeuvres and gap acceptance, but also with the infrastructure, for example handling a traffic light turning to amber or choosing the correct lane for turning in an upcoming junction.

Instead of reporting mean values, that potentially are not very expressive, it may be worth exploring the notion that a certain safety envelope may not be breached by any or x % of the drivers. This may make the interpretation of results much more straightforward. It will also avoid the setting of almost arbitrary thresholds based on mean values.

Follow-up studies should include several levels of situational complexity, as performance may break down suddenly and in a non-linear fashion. Several attempts to classify situational complexity have already been made (Fastenmeier, 1995; Schweitzer & Green, 2007), which could be used as starting point to build appropriate scenarios.

The secondary task complexity has to be appropriate, meaning that neither ceiling effects nor floor effects occur, and that the drivers do not give up trying to solve the secondary task. Subjective demand ratings as well as effort distribution ratings should be obtained for all secondary task levels, not only for the most difficult task level, as was done in the current study for practicality reasons. As a further subjective measure it would be thinkable to include the “perceived safety” of the situation. The usage of further questionnaires to assess driver characteristics is also thinkable, in case that individual analyses show that there are distinct driver groups that differ qualitatively in their behaviour.
References


ISO. (under development). Road vehicles - measurement of driver visual behaviour with respect to transport information and control systems, Part 2: Equipment and procedures. ISO/DTS 15007-2.


