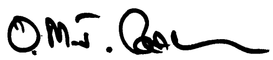
Human Machine Interface And the
Safety of Traffic in Europe
Project GRD1/2000/25361 S12.319626

Deliverable 3

- Validation of the HASTE protocol specification -

DOCUMENT CONTROL INFORMATION

Title	HASTE Deliverable 3 ***Validation of the HASTE protocol specification***
Author(s)	E. Johansson, O. Carsten, W. Janssen, S. Jamson, H. Jamson, N. Merat, J. Östlund, R. Brouwer, S. Mouta, J. Harbluk, V. Anttila, H. Sandberg, J. Luoma
Editors	Natasha Merat, Oliver Carsten
Reviewers	D. de Waard, K. Brookhuis, N. Ward
Reference Number	HASTE-D3
Version Number	1.1
Date	August 2005
Distribution	Project and European Commission
Availability	Restricted, until approved
File	H:\Safetygroup\HASTE\HASTE D3 v1-1.doc
Authorised	O.M.J Carsten
Signature	

Executive Summary

The aim of HASTE is to develop methodologies and guidelines for the assessment of In-Vehicle Information Systems (IVIS) and present an outline for a test regime which could be used both throughout the design process at IVIS manufacturers as well as in later stages for final verification and certification before the system is released on the market. The test regime would specify methods and tools which would:

- Be technology independent;
- Have safety-related criteria
- Be cost effective;
- Be appropriate for any system design; and
- Have been validated through real-world testing.

The objective of the Workpackage (WP) 3 experiments was to apply the methods developed earlier to real systems and recommend a test regime. The WP3 experiments included the most successful measures and scenarios found in WP2. The findings from WP2 were then validated by applying the protocol from WP2 to three real In-Vehicle Information Systems and tasks as well as one prototype system for traffic information.

In this deliverable, tasks within four In-Vehicle Information Systems have been assessed in field, simulator and laboratory experiments in six different countries. Besides different task types and a comparison between different test sites, dimensions of the experiments were methodology type (ranging from experiments in the field to low-scale laboratory experiments) and road complexity (straight, curved).

The TRL checklist was used to evaluate each selected system in order to assess the utility of using such checklists as part of the test regime.

The experimental results can be summarised as follows:

An a priori expert ranking of the tasks, with regard to overall difficulty level, was made prior to the experiments. However, this ranking does not generally correspond to the objective findings in the experiments.

A meta-analysis of the results of all studies was performed in order to bring out and grasp the common patterns in the experiments, to identify and select the most powerful parameters for detecting these patterns, and to check whether the conclusions drawn from the earlier S-IVIS studies (WP2) would uphold for real systems. If so, it would be possible to conclude that the necessary ingredients for a test regime for IVIS are available.

The analysis was done along the lines of the earlier analysis in WP2. It comprised a sequence of steps, each of them putting a statistical or methodological requirement on a measurement parameter that it could or could not fulfil. If not, the parameter was not retained for the next step. The outcome of this analysis was that a group of parameters were singled out that, taken together, are both necessary and sufficient in capturing the effects an IVIS has on driver behaviour. These were:

(1) The subjective rating of the quality of the driver's own driving performance with IVIS.
(2) The average speed when driving with the IVIS.

(3) The proportion of high frequency steering activity in the steering wheel signal spectrum when driving with the IVIS.
(4) The minimum time headway to a lead vehicle.
(5) Reaction time on the Peripheral Detection Task (PDT).

Reaction time in the PDT was not included in all WP3 experiments since the equipment was not available at all sites and thus the use of this parameter should be adopted with some caution. Another plausible candidate, Percent Road Centre (PRC), was also only included in a limited number of studies to provide conclusive results. However, this measure yielded sizable effects in those studies in which it was measured.

The effect sizes associated with all of the above parameters were so large that it would require no more than 15 subjects in an evaluation set-up to demonstrate the existence of IVIS effects. Further analysis of the results showed that the optimal environment in which to find IVIS effects would be the (full size) simulator, in which a rural road type should be used.

A comparison of results from the TRL checklist and the HASTE findings showed that the two should probably be considered as complementary, the HASTE behavioural measure assessing a system's effect on driving performance and safety, and the TRL checklist results identifying possible system design problems at a stage where they can still be corrected.

Based on the WP3 results, an outline of a test regime has been defined in which scenarios and measures are presented. It is hoped that the test regime will be developed further, in order to work for final verification and certification of systems soon to be on the market. However, if a final certification regime is introduced to the market it is important to develop the regime in a way that removes any possibilities of sub-optimisation of a specific system. Clear instructions on how to choose tasks to test, as well as what participants to recruit, need to be further specified. It is also important to stress that the test regime should work as a formative tool for use by system and vehicle manufacturers in order to improve the design throughout the iterative design process.

Disclaimer

The research reported herein was conducted under the European Commission Competitive and Sustainable Growth Programme. The project is being carried out by a consortium comprising: the Institute for Transport Studies, University of Leeds; TNO Human Factors Research Institute; the Swedish National Road and Transport Research Institute (VTI); Delft University of Technology; Volvo Technology Corporation; MIRA Ltd; Technical Research Centre of Finland (VTT); Universidade do Minho; and Transport Canada. The opinions, findings and conclusions expressed in this report are those of the authors alone and do not necessarily reflect those of the EC or of any organisation involved in the project.

Table of Contents

Index of Figures

Index of Tables

List of Abbreviations

IVIS	In Vehicle Information System, used in WP3 experiments
JAMA	Japan Automobile Manufacturers Association
S-IVIS	Surrogate In Vehicle Information System, used in WP2 experiments
ILv	IVIS Level, corresponds to Task number
RLv	Road complexity level
BL	BaseLine - refers to driving without IVIS
PDA	Personal Digital Assistant
PDT	Peripheral Detection Task
OEM	Original Equipment Manufacturer
LCT	Lane Change Task

1. Introduction

1.1. The HASTE Project

The aim of HASTE was to develop methodologies and guidelines for the assessment of In-Vehicle Information Systems (IVIS) and present it in an outline for a test regime which could be used both throughout the design process by IVIS manufacturers, as well as in later stages for final verification and certification before the system is released on the market. Ideally, the test regime would specify methods and tools which:

- Would be technology independent;
- Have safety-related criteria;
- Is cost effective;
- Is appropriate for any system design; and
- Is validated through real-world testing.

The theoretical schema behind the HASTE approach is illustrated in Figure 1. In-vehicle information systems can impose visual and/or cognitive loads on drivers. Such loads can be measured by glance behaviour and various indicators of workload. The effects of load can be manifested by changes in driving performance (e.g. reduced speed, greater lateral variability, decreased time-to-collision) and interference of perception and judgement of the traffic situation (i.e. reduced Situation Awareness). Reduced performance has a negative impact on safety, which could be measured, for example, by the increased risk of a conflict. All of this is influenced by the traffic environment and the current situation.

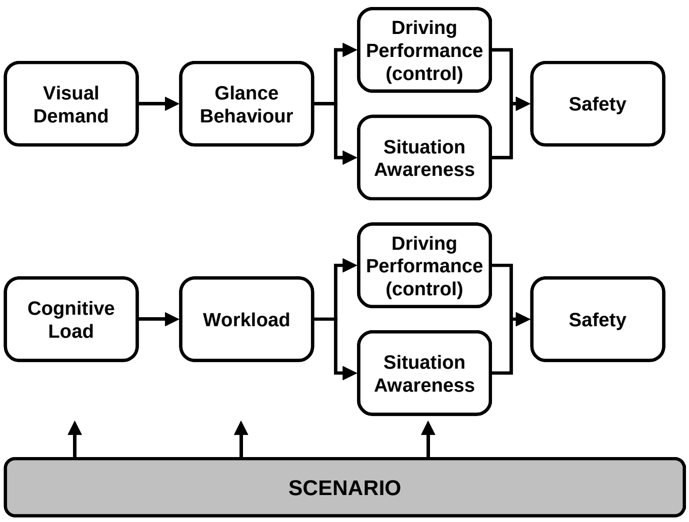


Figure 1 – The HASTE approach

The work of HASTE began by designing the experimental procedures to be used in Workpackage 1. This work is documented in Deliverable 1 (Roskam et al., 2002). The next stage, Workpackage 2, was the work examining the impact of distraction on driving performance. This work is covered in Deliverable 2 (Östlund et al., 2004). Here, two types of distraction, visual load and (non-visual) cognitive load were created and manipulated by means of specifically designed tasks. The relationship between each type of distraction and driving performance was then observed. Both levels of distraction and primary (driving) task load were manipulated. This work was designed to:

1. Identify the major impacts of distraction

2. Identify how driving performance changes as distraction increases
3. Test which features of the assessment are the most effective in diagnosis of (1) and (2).

The final stage of the work is to apply the test regime which has been developed from Workpackage 2 to the evaluation of real IVIS (Workpackage 3). The intention is that the test regime should be further refined at this stage.

1.2. Implications from WP2

Important conclusions from the studies in WP2 were:

- The effect of the S-IVIS visual task on driving was very clear: increased distraction led to problems in lateral control.

- The effect of cognitive task was more complex, in that some driving parameters, particularly related to steering control and lateral position appeared to improve. However, this improvement seemed to be an artefact of greater concentration on the road straight ahead at the expense of information acquired from the periphery. Therefore, it was suggested that thought needed to be given to tasks or tests that might capture this loss of information acquisition from the periphery.

- Motorway driving in the various simulators and the laboratory was generally less diagnostic.

- Elderly drivers exhibited very risky driving while performing IVIS tasks

- The field studies provided some information that was not provided by the simulator assessments. It was suggested that the subsequent work in the project should consider simulator tasks that could provide analogous information.

The results obtained confirmed some of the initial decisions made in formulating the HASTE approach. There was clear value to the focus on *dynamic* evaluation, i.e. of looking at interaction with an IVIS while driving and of identifying the effects of that interaction on driving. Static testing cannot predict how an IVIS will affect steering behaviour or interaction with other road users. The different road levels proved their worth, particularly levels 2 and 3 of the rural road. There was also clear value to the inclusion of events (road level 3), but there was thought to be some scope for improving the events so that the drivers would be less able to adapt to their occurrence, by for example slowing down as the lead vehicle comes closer to them.

It was suggested that there may be scope for the inclusion of peripheral detection tasks (e.g. PDT) in the driving, in order to gain a better understanding of drivers' ability to assimilate information in the periphery, which is crucial to safety maintenance. However, there were also some potential problems here: a PDT would become a tertiary task, in addition to the primary task of driving and the secondary task of interaction with an IVIS. Thus there might be the potential for the PDT to distort findings. This would require further investigation.

Workpackage 2 also used a very large number of indicators. Some of these indicators turned out to be non-diagnostic and therefore could be abandoned in WP3. Others turned out to be superfluous in that they revealed overlaps with the diagnosis provided by other indicators. The meta-analysis carried out for WP2 provided a capability to identify the most powerful indicators. This permitted a substantial reduction in the amount of data to be analysed and also permitted the number of participants in the WP3 experiments to be reduced as compared with WP2.

1.3. Objectives of the WP3 Experiments

The objective of Workpackage 3 experiments was to apply the devised methods to real systems and recommend a test regime. The WP3 experiments included the most successful measures and scenarios found in WP2. The findings from WP2 were then validated by applying the protocol from WP2 to three real In-Vehicle Information Systems and tasks as well as one prototype system for traffic information.

In this deliverable, tasks within four In-Vehicle Information Systems have been assessed in field, simulator and laboratory experiments in six different countries. Besides different task types and comparison between different test sites, dimensions of the experiments were methodology type (ranging from experiment in field to low scale laboratory experiments) and road complexity (straight, curved).

The main focus in the WP3 experiments was on the level of ‘task’ rather than ‘system’. The reason for this was simply that a system can consist of both “good” and “less good” tasks and thus the intention was to have a research focus where we would find out whether the HASTE test regime could specifically identify the bad tasks. If so, the test regime could later be used in order for e.g. the system manufacturer to re-design those particular tasks. Also, the scores could be weighted to provide final overall scores for a group of tasks or an overall system.

The TRL checklist was used to evaluate the selected system in order to assess the utility of using such checklists as part of the test regime.

Based on the WP3 results, an outline of a test regime has been defined where scenarios and measures are presented. It is hoped that the test regime will be developed further, in order to work for final verification and certification of systems soon to be on the market. However, if a final certification regime is introduced to the market it is important to develop the regime in a way that removes any possibilities of sub-optimisation of a specific system. Clear instructions on how to choose tasks to test, as well as what participants to recruit, need to be further specified. It is also important to stress that the test regime should work as a formative tool for use by system and vehicle manufacturers in order to improve the design throughout the iterative design process.

2. Common Methodology for IVIS Assessment

2.1. Methodological approach

The methodological approach was to impose load on the drivers using real systems and tasks.

The findings described in HASTE Deliverable 2 ***HMI and Safety-Related Driver Performance*** had a number of implications for the experiments in WP3. The scenarios in WP3 were limited to the ones considered most useful in a future test regime with the additional criteria of maintaining reliability between different simulators and between field, simulator and laboratory. Based on the WP2 findings, the road types were limited to rural road only in the simulators and laboratories. In the field experiments motorway was used, apart from the Leeds studies, where both rural and motorway were employed. The results from WP2 were validated in WP3 by the use of four real IVIS. For system and task descriptions, see section 2.2 below.

2.2. IVIS – systems and tasks

A range of systems currently available on the market were considered. Eight systems were selected for an initial preliminary assessment where general system descriptions were made along with more specific task descriptions. The tasks were classified into:

- Different complexity levels (how easy/difficult the task was thought to be when performed while driving)
- Main modality (e.g. mainly visual task) and
- System- or driver paced task.

Each task was also classified into real vs. simulated versions for the different methodologies (laboratory, simulator or field). This classification was made since some tasks would not work in the simulator and laboratory, since no GPS signal was available. Also, simulations were created for tasks which would not work in identical ways for all participants if the real system was used.

Four systems were then chosen for the validation experiments in Workpackage 3. System A is similar to many Original Equipment Manufacturer integrated in-vehicle systems (e.g. with regard to display size and placement). The display was fairly large and the main interaction was made with a remote control. Systems B and D were PDAs and therefore had smaller displays. They were both very common nomad devices and very similar to each other. These two systems were mainly chosen because (i) they had a small screen (thus different to System A), (ii) were able to perform an almost unlimited range of tasks which could often be quite demanding, and (iii) had different input devices as well as voice output. System C was a simulation of a possible system for RDS-TMC messages concerning accidents, congestion or roadworks. The text messages on this system were presented on a touch screen located on the "off side" of the steering wheel and the driver could accept new incoming messages as well as scroll in menus and select information. For more detailed system descriptions see sections below.

For each system the tasks were a-priori divided into overall complexity level. The ranking was based on number of modalities, number of button presses and manual difficulty level.

Simulations of tasks were made in Macromedia Director.

2.2.1. System A

System A is available on the market and consists of a removable 6.5 TFT colour display in 16/9 format with a remote control and hand controls. The cost of this system is €2000. Route guidance information is provided with symbols, a map and voice output. A variety of displays (simultaneous arrow/map, large map or arrow) and map alignments (north, automatic, zoom on junction) are possible with this system. Examples of on-board computer functions for this system include: display of arrival time and remaining distance from destinations, current speed, distance already travelled, total journey time and average speed. Route options can be pre-set (e.g. fast/short route; avoid motorway/ferry/toll). The task descriptions chosen for testing are shown in Table 1.

Table 1 – Task description of system A

Task no.	Task description	Main modality	Manual effort (low, medium, high)	Approximate time to complete (seconds)	Real/simulation
1	Auditory route guidance message incl. arithmetic information (distance to final destination)	Auditory	None	12.2	Simulation
2	Like Task 1 but with more information (distance to final and intermediate destination)	Auditory	None	16.1	Simulation
3	Auditory route guidance message incl. spatial information (turn by turn instructions)	Auditory	None	17.8	Simulation
4	Like Task 3 but with more information	Auditory	None	22.3	Simulation
5	Entering Destination by data entry – City	Visual-Manual	Medium/High	52.1	Real
6	Entering Destination by data entry – City, Street	Visual-Manual	High	61.3	Real
7	Change volume	Visual-Manual	Easy/Medium	10.4	Real
8	Change settings; add one displayed category (e.g. gas stations)	Visual-Manual	Easy/Medium	28	Real
9	Change settings; add six displayed categories (e.g. gas station, hotel etc)	Visual-Manual	Medium	44.2	Real

The following a-priori ranking of the tasks was made: 1, 2, 3, 4, 7, 8, 9, 5 and 6.

2.2.2. System B

System B is available on the market and consists of a PDA and a GPS unit. The system is attached to the windscreen coupled to the vehicle power via a cable. The PDA has a colour touch display and data entry is made either by a stylus for most functions or by hardware keys. The approximate cost is €750. Data is mainly presented as visual information in a range of ways (icons, text etc) but also via voice output (e.g. route guidance information). The user has a range of possibilities to alter settings and enter information. The tasks are described below (see Table 2).

Table 2 – Task description of system B

Task no.	Task description	Main modality	Manual effort (low, medium, high)	Approximate time to complete	Real/simulation
1	Check visual information: distance and suggested action (e.g. turn left in 250 metres)	Visual	None	11.6	Simulation
2	Read list of directions: (e.g. turn x on street y in z metres)	Visual	None	23.4	Simulation
3	Close the navigation program and re-open	Visual-Manual	Low	11.1	Real
4	Change settings from large to small and back to large keys	Visual-Manual	Medium	21.2	Real
5	Set destination by choosing a pre-set destination	Visual-Manual	Medium/High	18.9	Real
6	Set destination (City, Street, Nr) by data entry	Visual-Manual	High	47.6	Real
7	Zoom in to 10 metres, out to 10 km and back to 100 metres	Visual-Manual	Low/Medium	32.5	Real
8	Change settings: route options	Visual-Manual	Medium/High	24.3	Real
9	Create a waypoint by pointing to a road on map with stylus	Visual-Manual	Medium	16.5	Real

The following ranking of the tasks was made: 1, 2, 3, 7, 9, 4, 8, 5 and 6.

2.2.3. System C

System C was a system simulation of a traffic information system used in Finland (with mobile phones). In this simulation, the display consists of a removable black and white touch screen. Traffic information is provided with written messages, together with an auditory presence sign. The information is presented in menus that allow the reading of the message and a search for previous messages in a menu. This search can be made using the number of the message, or the road name. The system requires a manual action by the driver. The drivers accept and select the messages by pressing a button on the touch screen and use a scroll function to read the entire message. All tasks have a maximum presentation period of 60 seconds. The major difference between the message types is the use of different menus and the order of presentation of the messages. All the messages are visual/manual, but are divided into three task types: for the first type (e.g. task 1) participants were only required to read the message, without using any menu. In task type 2 (task 4) participants had to search for a message in a menu by a task number. The number of messages was ordered by this same number in the menu, so it was an easier search. In task type 3 (task 7), the search for messages was made by road name. The order of the roads in the system menu is random (because the messages are ordered by number), so it is the hardest search task since it requires more attention to identify the correct message (see Table 3 for more details).

Table 3 – Task description of system C

Task no.	Task description	Main modality	Manual effort (low, medium, high)*	Approximate time to complete	Real/simulation
1	Accept new message, read message, answer question	Visual/manual	Low	21.1	Simulation
2	Like T1	Visual/manual	Medium	25.3	Simulation
3	Like T1	Visual/manual	High	27.2	Simulation
4	Search message number X from list of messages, read message, answer question	Visual/manual	Low	24.8	Simulation
5	Like T4	Visual/manual	Medium	34.1	Simulation
6	Like T4	Visual/manual	High	36.3	Simulation
7	Search message “Road X…” from list of messages, read message, answer question	Visual/manual	Low	27.5	Simulation
8	Like T7	Visual/manual	Medium	31.8	Simulation
9	Like T7	Visual/manual	High	44.6	Simulation

The following ranking of the tasks was made: 1, 2, 3, 4, 5, 6, 7, 8 and 9.

2.2.4. System D

System D is available on the market and is the first PDA to include integrated GPS technology. It consists of a 54 x 81mm display in a 72 x 128 x 20.3 mm PDA unit.
It is operated using a stylus. The approximate cost is $750 CAN (approx €465). Route guidance information is provided with symbols, a map and voice output. A variety of displays and map alignments (north, automatic, zoom on junction) are possible. Other functions include an MP3 player, appointments and contacts information, voice recorder and an SD expansion slot for flexible memory and additional software.

The on-board computer functions include a display of arrival time and remaining distance from destinations, current speed, distance already travelled, total journey time and average speed. Route options can be pre-set (e.g. fast/short route; avoid motorway/ferry/toll). For task descriptions see Table 4).

Table 4 – Task description of system D

Task no.	Task description	Main modality	Manual effort (low, medium, high)	Approximate time to complete (s)	Real/simulation
1	Check visual information: Read two lines of text	Visual	Low	9.9	Simulation
2	Read list of directions: (e.g. turn x on street y in z metres) and calculate	Visual	Medium	15.7	Simulation
3	Auditory route guidance message incl. spatial information (turn by turn instructions)	Auditory	Difficult	12.8	Simulation
4	Participant pan map north/south three times and read out street name from the display aloud	Visual-Manual	Medium	16.7	Real
5	Locate and display special destinations on map (e.g. drink/lodging locations) and read out third item in list	Visual-Manual	Medium/ Difficult	28.9	Real
6	Set destination by data entry using the stylus: Street, Nr	Visual-Manual	Difficult	45.6	Real
7	Zoom on the map from 200 metres to 1.2 km or 2km metres and out again to 200 metres	Visual-Manual	Medium	5.8	Real
8	Change settings: route options	Visual-Manual	Difficult	17.3	Real
9	Create a waypoint by pointing to an intersection on map with stylus	Visual-Manual	Medium	32.8	Real

The following ranking of the tasks was made: 1, 2, 3, 7, 9, 4, 8, 5 and 6.

2.3. Participants

"Average" drivers were included in all experiments and they were defined in the following way:

- Age: 25-50 years old
- Gender: male and female
- Driving experience: to have a minimum total driving experience of 10,000 km.
- Driving license for a minimum of 5 years.
- Non-professional drivers

Based on effect size calculations for the dependent measures in the WP2 experimental results (Östlund et. al, 2004) the number of participants were greatly reduced from n=48 to n=20 (minimum) in simulator and laboratory experiments and n=15 in the field experiments. The University of Leeds conducted simulator, laboratory and field experiments and selected a smaller group of 18 drivers all of whom performed all three experiments.

2.4. Experimental design

2.4.1. Overview of experimental design

Ten experiments were carried out in order to cover all four systems and the three methodologies (laboratory, simulator, field).

All simulator and laboratory experiments contained a rural road. In the field experiments the participants drove on motorways. See
Table **5** for an overview.

Table 5 – Distribution of road types and in-vehicle systems

Systems	Field (motorway)	Simulator (rural)	Laboratory (rural)
System A	Volvo, VTT	VTI, TC	-
System B	Volvo, Leeds*	VTI, Leeds, TNO	Leeds
System C	VTT	-	Minho
System D	-	TC	-

**In the field experiment by Leeds the participants also drove along a rural road.*

In the simulator experiment by VTI and in the field experiment by Volvo Technology the PDT task was included in the experiment and added as an extra condition. In the TC experiment the PDT was not an extra condition. See Table 6 for an overview.

Table 6 – Distribution of PDT task used by partners in the experiments

Systems	Field (motorway)	Simulator (rural)	Laboratory (rural)
System A	Volvo	TC, VTI	-
System B	Volvo	VTI	-
System C	-	-	-
System D	-	TC	-

Baseline drives without the IVIS were included within the experimental drive in the laboratory, simulator and field experiments except at VTT. Three baseline sections were collected in the beginning, middle and end of each drive with each system. At VTT, an extra baseline drive was driven in order to collect expert assessment of driver performance with the Wiener Fahrprobe protocol (e.g. interaction with vulnerable road users).

For partners assessing one system the design was as presented in Table 7 below.

Table 7 – Design for 1 system

All particip ants	System x	BL	½ the number of tasks	BL	½ the number of tasks	BL

A within-subjects design was used for partners who assessed more than one system. In these cases, half the group of participants started with one of the systems followed by the second system and vice versa (see Table 8).

Table 8 – Design for 2 systems

½ group of participants	System x	BL	½ the number of tasks	BL	½ the number of tasks	BL	System y	BL	½ the number of tasks	BL	½ the number of tasks	BL

At VTI and Volvo Technology the PDT was included as an extra condition and the PDT runs were not balanced. The reason for this was so that the results from the no-PDT condition would be comparable with the rest of the experiments where no PDT was included. In the VTI experiment only half the group of subjects ran the experiment with the PDT. In the Volvo experiment all subjects drove with the two systems and with the PDT.

The tasks were presented in randomised order for each participant in all experiments.

2.4.2. Impact of IVIS on experimental design

The IVIS task was a within-subjects factor. The IVIS level (ILv), had a different number of levels depending on system and experimental environment (field, simulator or lab). For an overview of levels see Table 9 below. Each of the four systems had nine tasks. Due to the fact that some tasks were assumed be too difficult to perform in the field, Task 5 and 6 for System A and Task 6 for System B were removed for this environment. These levels were randomized in one single drive for each driver.

Table 9 – Number of levels for each IVIS

Systems	Field (motorway)	Simulator (rural)	Laboratory (rural)
System A	7	9	9
System B	8	9	9
System C	9	9	9
System D	9	9	9

2.4.3. Impact of Road Complexity on experimental design

In the simulator and laboratory experiments, road complexity was a factor with two levels: straights and curves. In the field experiments only one single level of road complexity was used.

Rural roads were selected for the simulator environments because these were shown to produce the best discriminative performance in WP2, in terms of making IVIS effects visible. For the same reason, motorways were chosen as the environment for the field experiments.

2.4.4. Factors and Levels

The experimental factors and number of included levels are listed below (Table 10). Experiments on different road types, with and without PDT, are considered separate experiments. Road type and PDT are thus not included as factors. A mixed effect model was chosen with Road complexity and IVIS level as the main factors which were rationally and systematically chosen and therefore considered as fixed. Subject was included as a random factor.

Table 10 – Factors and Levels

Factor	Levels	Type
IVIS level (ILv)	See table above + 1 for baseline	fixed
Road complexity level (RLv)	2 (in Sim and Lab rural road) 1 (in field trials)	fixed
Study object	N (number of participants, at least 20 in sim and lab and 15 in field)	random

2.5. Assessment Methods

In WP3, two fixed base and two moving base simulators were used. Also, at Leeds and Minho, PC-based driving simulators with a PC-monitor, and low cost steering and speed controls were used and will be referred to in the report as laboratory experiments. Three field experiments with instrumented vehicles were conducted at Leeds, Volvo Technology and at VTT.

2.6. Road and Traffic Environments

For the reasons given in section 2.4.3 the experiments in WP3 were narrowed down in their selection of road types on the basis of WP2 results.

A number of critical events were implemented in WP2 to examine the effect of an IVIS on hazardous or unexpected road situations. However, these critical events were not included in the WP3 experiments since they were shown to introduce a strong learning effect and were problematic in terms of producing real-life critical situations. Also, since the signals ideally needed to be repeated more often and in order to keep the idea of reaction time for a final test regime, the PDT was introduced. Thus, the detection task in WP3 differed a lot from WP2 experiments since it was a simple reaction time task not incorporated in the driving environment. The use of PDT meant that the event rate was continuous with a single source complexity. Also, the signal discrimination type was simple sensory. However, this task has a higher signal rate and higher expectancy which might lower the face validity of the method.

2.6.1. Rural Road in Simulator and Laboratory experiments

The WP3 simulator/lab drives were to be based on WP2 as much as possible. Hence, a rural environment was used and the individual road segments used were identical to WP2. However, the exact road layout of these segments was different. Furthermore, since interactions of driving difficulty (straight and curved road segments) and IVIS performance were detected in WP2, it was decided to keep driving difficulty as a factor in WP3. Hence, two levels of driving difficulty were present:

- Driving on straights
- Driving on s-shaped curves

The car following task used in WP2 was also retained since it provided an extra source of data on longitudinal driver behaviour and performance. However, to simplify the behaviour of the lead vehicle, the driving speed of the lead vehicle was fixed in the WP3 experiments. Rather than simply choosing the same fixed lead vehicle speed for each participant, it was also decided to attempt to create a following scenario that was sensitive to the fact that different drivers adopt different driving speeds. To achieve this, the first section of the virtual drive included both a straight and curved section in which no lead vehicle was present. This was to allow the measurement of *"free speed"* choice for each individual driver. This value was measured separately on both straight and curved sections. Once this was achieved, the lead vehicle was introduced at an intersection, building up to a *driver-dependent speed*. The driver dependent speed was defined as either:

- Each driver's "free speed" minus 10%, or
- Speed limit of the road (96km/h) minus 10%

The minimum of these two values was used as a separate *driver-dependent speed* for both straight and curved sections.

The rural simulator and laboratory roads were divided into *test sections* (either *straight* or *curved*) which were long enough to allow for up to 90 seconds of current IVIS and driving performance to be measured (based on a speed limit of 96km/h). Within the 90 seconds the subjective rating of driving performance was also to be included and thus the actual task performance had an upper limit of 80 seconds. After 80 seconds, a verbal instruction 'RATE' was announced and the participant had 10 seconds to state their subjective rating. In order to allow drivers a short rest between tasks, the test section also included a *filler section* (i.e. a 10 second period of "relaxed" driving when no concurrent IVIS interaction and driving took place). Test sections were also added for baseline data collection.

The road layout essentially consisted of a series of *straight* and *curved test sections* laid end-to-end. The number of test sections were flexible depending on the IVIS system and the number of tasks investigated, but allowed for at least nine tasks (i.e. 9 *straight* and 9 *curved test sections*) plus six *baseline sections* (straight and curved) positioned at the start, the middle and the end of each drive. Two sections (one curved and one straight) were required at the beginning of the road to allow for *free-speed* measurement. At the end of the *free-speed* sections, there was a cross-road intersection to allow for the introduction of the lead vehicle and around 1km of road to allow the lead vehicle to reach its desired speed (free speed – 10%, maximum speed 96km/h speed limit). In addition, there were the 10 seconds *filler sections* and 10 seconds *task description sections* between each *test section*. The width of the road was 7.3 metres made up of two carriageways of 3.65 metres each.

2.6.2. Motorway in Field experiments

In the motorway experiments, the direction of travel was divided with at least two lanes per direction (plus hard shoulder). Speed limits were in general between 110 and 130 km/h. The lane width was generally between 3.6 – 4.0 metres. The following variations across test sites also existed:

- In the Volvo Technology experiment there were two lanes per direction plus hard shoulder and the speed limit ranged between 90 and 110 km/h.

- In the Leeds field experiment, the motorway consisted of three lanes in each direction, with a speed limit of 112 km/h.

- In the VTT experiment the tests were carried out on a motorway section in the Helsinki capital area. The posted speed limit was 120 km/h.

2.7. Dependent measures and Analysis

2.7.1. Dependent measures

In order to validate the results in WP2, a sub group of the WP2 measures were chosen for WP3 experiments. The WP2 measures which had large effect sizes and were especially sensitive to S-IVIS difficulty were used in WP3.

Additional headway measures were added for the simulator and laboratory experiments since it was of interest to investigate changes in different parts of the headway distribution. In WP2, mean headway was shown not be the optimal headway measure and so in WP3 it was interesting to establish if drivers spent more or less time at different headways. Therefore, the added measures were hwt_0_1, hwt_1_2, hwt_2_3, hwt_3_4, hwt_4_5, hwt_5_6 and hwt_6.

Peripheral Detection Task measures were added (pdt_hit (%), pdt_rt (s), pdt_miss (%), pdt_cheat(%)).

A large difference between the S-IVIS in WP2 and the real tasks assessed in WP3 was that the real tasks varied in task length. Therefore, some of the measures were biased by length. To overcome this bias, a 'sliding window' technique was used, creating a number of new measures (e.g. st_lp15, st_lp30). For more information on this matter see Appendix 2.

The dependent measures were grouped in the following way:

- Lane-position and Time-to-line-crossing measures
- Steering wheel measures
- Speed and Headway-related measures
- Eye Movements
- PDT measures
- Subjective ratings

The measures were further divided into 'Mandatory' and 'Optional' measures. The 'Optional' measures were mainly the ones that either had shown weaker effects in WP2 or relied on sensors not available to all partners.

The measures calculated in the Simulator and Laboratory experiments are presented in Table 11 and the ones for field experiments in Table 12. Implementation specification is presented in Appendix 2.

Table 11 – Dependent measures for Simulator and Laboratory experiments

Lane-position and Time-to-line-crossing measures	
Mandatory	mn_lp(m), st_lp(m), st_lp15, st_lp30
Optional	lnx(%), pr_tlc(%)
Steering wheel measures	
Mandatory	rr_st1 (1/minute), rr_st3 (1/minute), mn_tlc (s), hi_st
Optional	rswt_5 (1/minute), rswt_10(1/minute), rswt_20(1/minute), rswt_40(1/minute), rswt_70(1/minute), hi_st2
Speed and Headway-related measures	
Mandatory	mn_sp(km/h), u_sp(km/h), st_sp(km/h), st_sp15, st_sp30, d_sp(km/h), n_hwd(m), sd_hwd(m), sd_hwd15, sd_hwd30, u_hwt(s), mn_hwt(s), sd_hwt(s), sd_hwt15, sd_hwt(30), hwt_0_1, hwt_1_2, hwt_2_3, hwt_3_4, hwt_4_5, hwt_5_6, hwt_6
Optional	-
Eye Movements	
Mandatory	-
Optional	n_gl, tot_gl (s), st_ga(deg), PRC(%), mn_gd(s), tot_gl_t (s)
PDT measures	
Mandatory	-
Optional	pdt_hit (%), pdt_rt (s), pdt_miss (%), pdt_cheat(%)
Subjective ratings	
Mandatory	subj_r,
Optional	-

Table 12 – Dependent measures for Field experiments

Lane-position and Time-to-line-crossing measures	
Mandatory	-
Optional	mn_lp(m), st_lp(m), st_lp15, st_lp30, pr_tlc(%), mn_tlc(s)
Steering wheel measures	
Mandatory	rr_st1(1/minute), rr_st3(1/minute), hi_st
Optional	rswt_5(1/minute), rswt_10(1/minute), rswt_20(1/minute), rswt_40(1/minute), rswt_70(1/minute), hi_st2
Speed and Headway-related measures	
Mandatory	mn_sp(km/h), st_sp(km/h), st_sp15, st_sp30, u_sp(km/h), d_sp(km/h)
Optional	-
Eye Movements	
Mandatory	-
Optional	n_gl, tot_gl, st_ga(deg), PRC(%), mn_gd(s), tot_gl_t (s)
PDT measures	
Mandatory	-
Optional	pdt_hit, pdt_rt, pdt_miss, pdt_cheat
Subjective ratings	
Mandatory	subj_r, compl_t
Optional	obs_r

2.7.2. Common analysis method

The analysis method designed at VTI for WP2 experiments was also adopted for the WP3 experiments. In order to compare results between the experiments, all partners used the same analysis method. Leeds conducted both a laboratory and a simulator experiment and thus adapted the design by adding methodology type (Simulator Type: LabSim (the Leeds scaled down driving simulator) vs their full scale Driving Simulator) as a within subject factor in their design.

For simulator and laboratory experiments, the rural road had two difficulty levels (RLv; curved and straight sections). Univariate Analysis of Variance (ANOVA) was used with a 5% level of significance. For more details on the analysis plan, see Appendix 4. Along with RLv, IVIS level (IVt; i.e. IVIS task) was included as a factor (with a maximum of 9 levels). For partners running experiments with more than one IVIS, these were analyzed in separate analyses.

In order to illustrate the ANOVA results, we presented model-based estimates of the mean and confidence intervals in the graphs. If the groups are equally large, the confidence intervals should be of the same size since this is a basic assumption in the ANOVA model.

For each of the individual reports there are summary tables in which significant main effects are indicated with '✓'. A '✗' means that no main effect was found. When the values representing the tasks differ from the 'No Task' condition, the cell is highlighted.

3. The MINHO laboratory experiment

3.1. Test site

The laboratory experiment was performed in a simple scaled-down driving simulator (DriS) at the Faculty of Engineering of the University of Porto. The main core of DriS runs on a PC Pentium IV 3.2 Hz with a graphic boarder Nvidia GeForce 4 Ti 4200. This PC holds the scene database, and performs the simulation and the computer graphics tasks. In these experiments, the driver saw the image on a 21” monitor at a distance of 80 cm. The horizontal visual angle under these conditions was 27°. Experiments were performed with a spatial resolution of 1280x1024, and a temporal resolution of 30 frames per second. The driver interface was composed of a low cost kit of steering-wheel and pedals (brake and accelerator). Audio and dynamic feedback were not provided in these experiments. All the experimental work was recorded by a video camera.

3.2. Scenarios and participants

The sample included 20 average drivers, who held a licence for at least 5 years and were aged between 23 and 38 years. The average age of participants was 27 years (SD = 3.977) and there were 16 male and 4 females in the group. The average driving experience of the group was 8 years (SD = 3.127).

The experimental route only included a rural scenario. The speed limit was 90 km/h. The road had a length of just over 81 km, divided into 24 sections: 9 straight and 9 curved sections and 6 sections were used as baseline (i.e. without the system).

3.3. IVIS included

A simulation of System C (already described in this document) was included. The 9 tasks were adapted to the Portuguese road environment and a tactile screen was used for message display.

3.4. Experimental design

The experimental design followed the description already presented in this report.

Before the experiment, all the participants underwent a learning period (driving, additional task and performance rating), while they drove on a specific road circuit with the same instructions of the experiment. All the experiments were completed with one ride on the rural road. After the system display and at the corresponding points of the baseline sections (6 points), the participants were asked to rate their driving performance on a scale from 1 to 10. This assessment was triggered verbally by the test leader.

3.5. Measures and analysis method

The simulator data were sampled at 10 Hz. The collected data allowed the calculation of all the mandatory measures – longitudinal control, lateral control and self-reported driving performance related measures, already indicated in this document. Some optional measures were also collected:

- Percentage of lane exceedence (lnx %)
- minimum distance headway
- time exposed to time to collision (TET)

In the lab studies, the variables related to lateral and longitudinal position were measured by taking the body of the driver as the reference point (instead of a part of the vehicle as in most simulators).

The driving and S-IVIS data obtained in the lab experiments were analysed with the common method already described.

3.6. Results

Before the procedures for the analysis of the system effects, the hypothesis that all baseline sections have the same effect on driving data was tested to allow a later aggregation of the three sections. Significant differences between the six sections were not found. The assumption that the six events were equivalent was assumed within the subsequent data analysis.

3.6.1. Effects of System C

3.6.1.1. Self-reported driving performance

Analyses of variance revealed an effect of the different System C tasks on self-reported driving performance (subj_r). Drivers rated their performance better in the baseline, compared to the sections where the system tasks were presented. Although they did not differ significantly in their performance between most of the nine tasks, task 7 resulted in a significantly higher evaluation of driving performance than tasks 8 and 9.

An increase in difficulty of driving scenario was seen by drivers as a factor that reinforced their workload and reduced their driving performance. Drivers rated their performance as worse in the negotiation of the curves, when compared to the straight sections (see Figure 2).

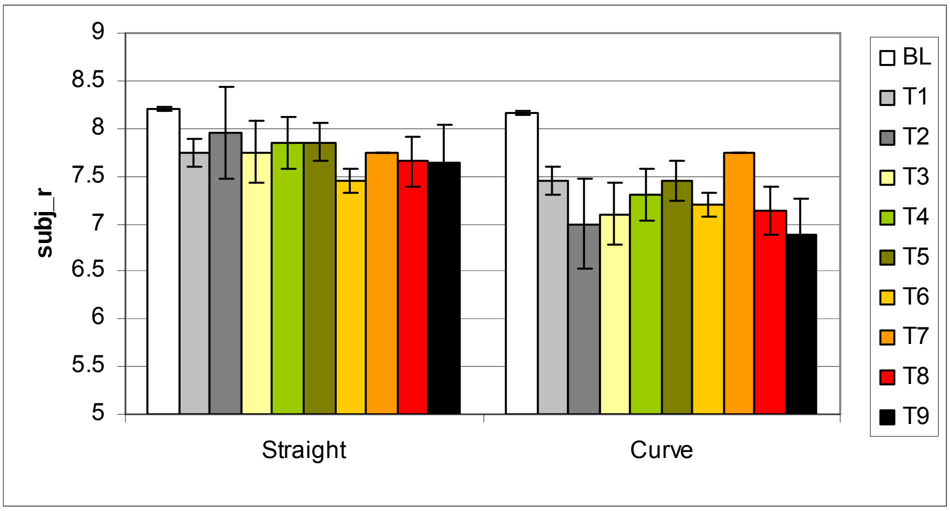


Figure 2 – Subjective ratings (subj_r) for system C and tasks on rural road

3.6.1.2. Longitudinal control

When driving the rural road, the presence of System C tasks was found to reduce mean speed (mn_sp). This effect was significant for all the tasks, except task 1. It appears that when confronted with the simultaneous negotiation of the two tasks (driving and system operation), drivers tended to reduce their speed in order to try and control the rise in workload of the situation.

Analysis of variance also showed that drivers reduced their mean speed of travel in the curves, compared to the straight sections (see Figure 3).

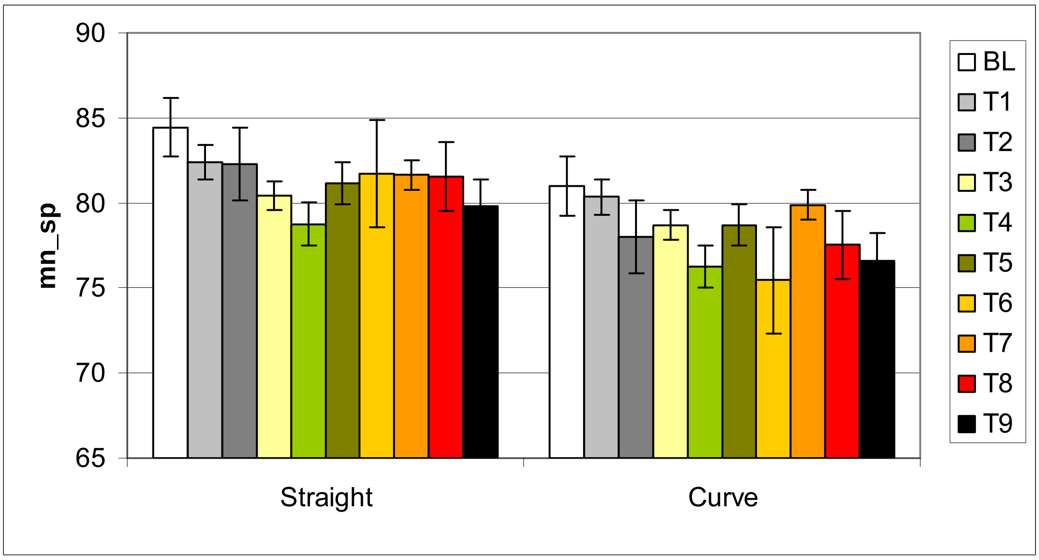


Figure 3 – Mean speed (mn_sp; km/h) for system C tasks on rural road

This tendency to reduce speed was supported by the data obtained for Time exposed to Collision, which also decreased with the presence of System C, when compared to the baseline.

Distance headway related measures indicated a significant effect of the presence of the System C tasks. The mean and minimum distance headway to the lead vehicle increased during the presence of tasks 6 and 9, suggesting that the visual and manual demand induced drivers to reduce speed and thus “drop back” from the vehicle in front.. The effect of these tasks was significantly different from the baseline driving performance and the other system tasks.

Analysis of results showed that the variation of the distance time headway increased with the increase of task difficulty. The presence of demanding tasks during the driving elicited a higher instability in the longitudinal control of the car. This was supported by the lower results of baseline and task 1 when compared with the tasks 2, 3, 5, 6, 8 and 9.

Although the effects of the road levels in the driving measures are not significant, Figure 4 shows an increase in the variation of distance headway in some tasks during the curved sections.

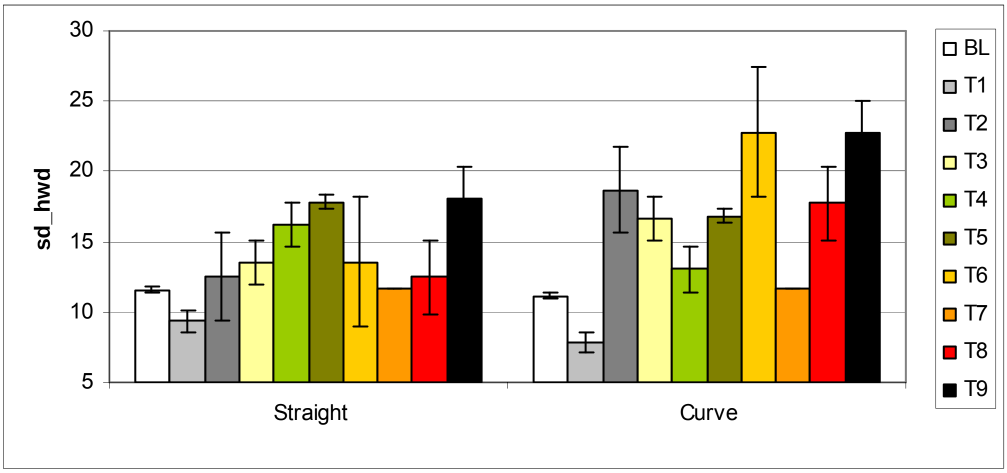


Figure 4 – Distance headway variation (sd_hwd) for system C tasks on rural road

As with the analysis outlined above, the results of time headway related measures indicate a significant effect of increasing task demand on driving performance. Analysis of variance showed that drivers increased their mean and minimum time headway from the lead car when performing tasks 6 and 9. At the same time, they failed to maintain stability, by increasing variation during the performance of the tasks and subsequently reducing longitudinal control.

A significant effect of road levels was not found for the time headway related measures, but an increase in time headway variation for tasks 6 and 9 can be observed in Figure 5.

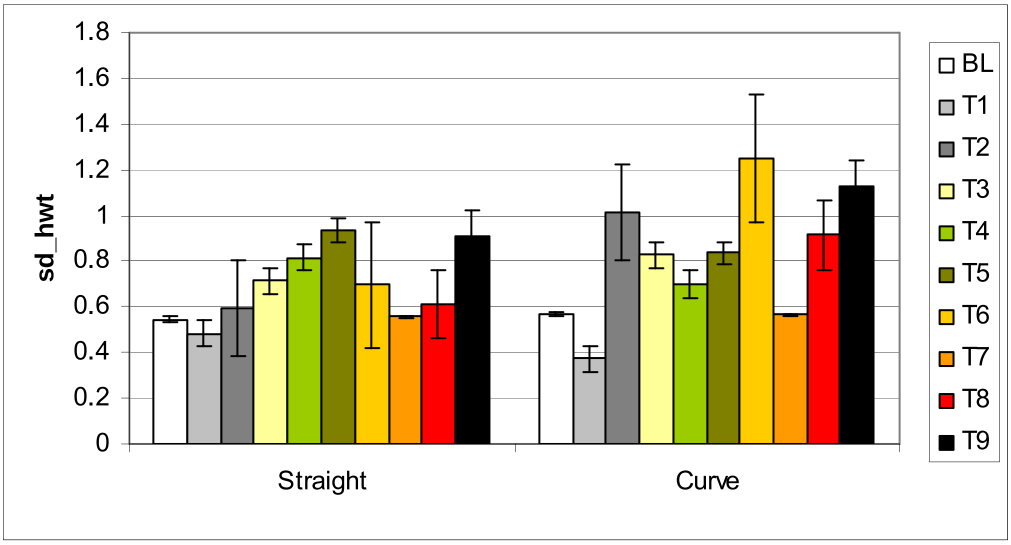


Figure 5 – Time headway variation (sd_hwt) for system C and tasks on rural road

The headway distributions by task duration are shown Figure 6. In the curved sections, drivers performing tasks 1 and 7 travelled closer to the lead vehicle for the majority of the task duration. On the other hand, the tasks with the hardest level of difficulty were distributed along the categories, with a significant incidence in the longest time headway categories. This fact reflects the speed reduction and increase in headway when the driver was required to manage a high workload generated by the road and task difficulty (as reported above).

During the straight sections, all tasks are distributed in a similar way, but it can be observed that the drivers kept the shortest headway to the lead vehicle during baseline and task 1.

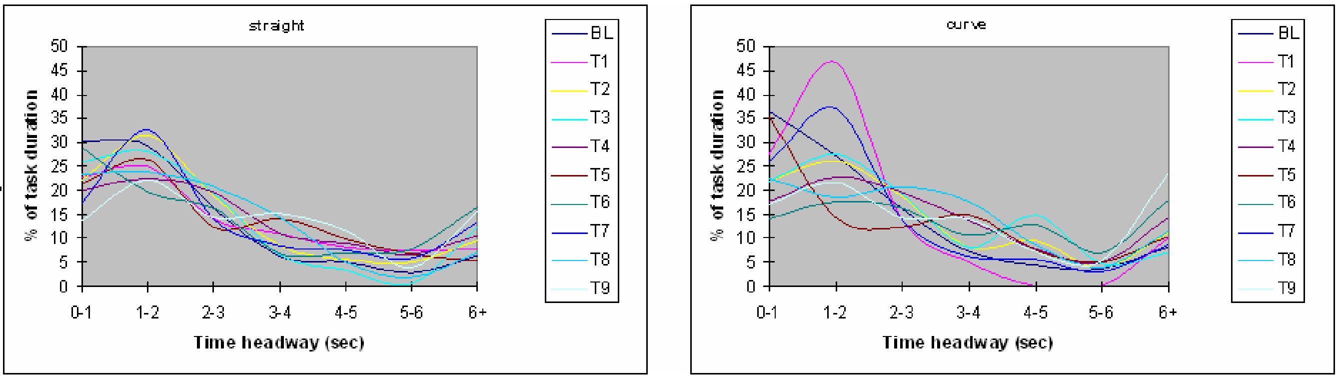


Figure 6 – Percentage of task duration in the different categories of time headway for system C tasks on rural road

3.6.1.3. Lateral control

In terms of lateral control measures, the presence of system C was found to have a consistent effect on the mean lateral position (mn_lp) and standard deviation of lateral position (st_lp). The mean lateral position increased in task 2 and 9, when compared to the baseline and there was reliably more deviation in lateral position when driving was performed in conjunction with tasks 2, 3, 5, 6, 8 and 9 of the system (see Figure 8). Lateral control of the car was actually worse in presence of system C, especially with the hardest tasks, when compared to baseline.

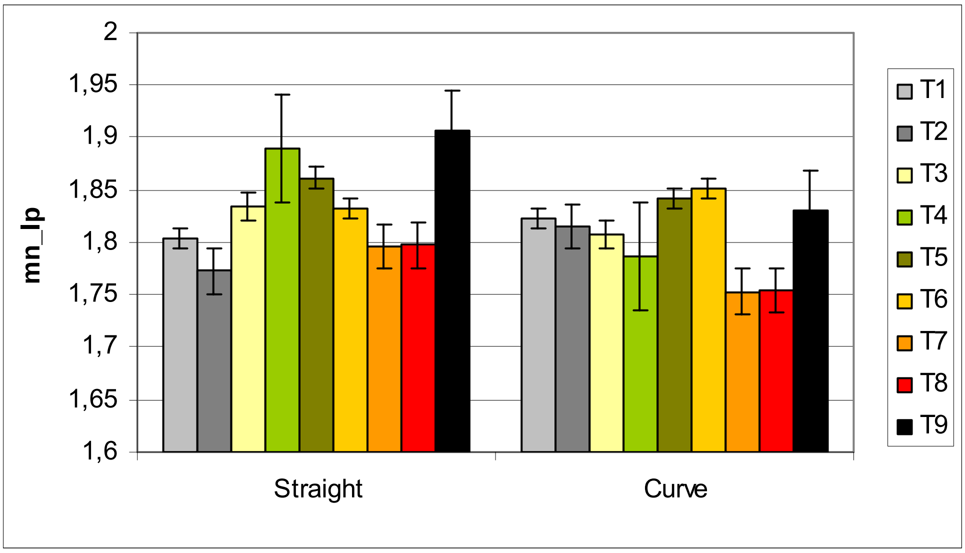


Figure 7 – Mean lateral position (mn_lp) for system C and tasks on rural road

Standard deviation of lateral position (sd_lp) was found to vary reliably across the two road scenarios, with an increase of this measure in the curved sections (see Figure 8). This scenario constitutes a more demanding driving task, and when combined with the hardest tasks of System C, a peak in difficulty of maintaining the lateral control of the car was seen.

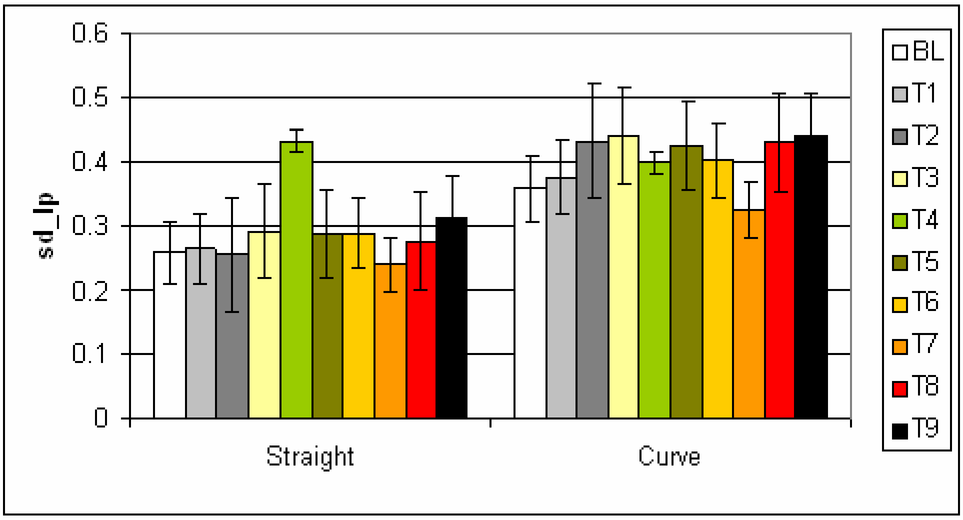


Figure 8 – Standard variation lateral position (sd_lp) for system C tasks on rural road

This reduction in lateral control, as a result of the presence of the tasks of System C, was also shown by measures such as the number of steering reversals (rr_st1) - see Figure 9. This measure increased significantly in all the tasks, when compared to the baseline. Whilst drivers were aware that their performance with the system tasks might affect their control of the car (as indicated by the self-reported measures), and reduced their speed and headway, they were less aware or unable to deal with a loss in lateral deviation control.

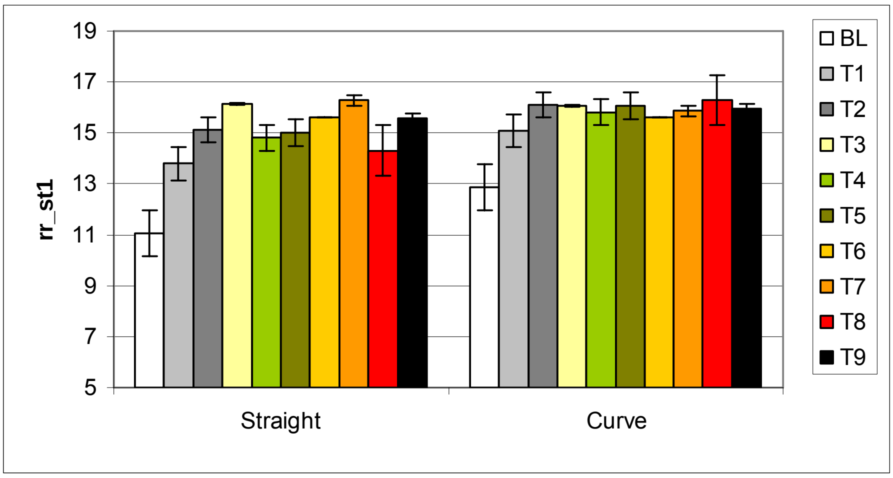


Figure 9 – Reversal rate (rr_st1) for system C and tasks on rural road

3.6.1.4. IVIS performance

All participants experienced thorough training session on system C tasks. All participants performed the tasks correctly. The only exception was for tasks 6 and 9 (the hardest tasks of type 2 and 3 respectively) where they needed to spend more than 40 seconds completing the task, up to a maximum of 60 seconds.

The increase in task completion time along the difficulty levels supports not only the fact that the hard tasks demanded more time for completion, but also that drivers felt the need to deal more carefully with the driving task. When drivers were required to attend to the IVIS task and driving at the same time, there was a tendency to delay completing the IVIS task to prevent the loss of vehicular control (adopting one or more of the strategies already described – reduction of mean speed, increase of headway).

The observed increase in task completion time in the curved sections (Figure 10) supports the above interpretation. In the presence of a more demanding road environment, task completion time increased. This suggests that an increase in completion time is due not only to the increased demand of the task, but also to the increase in workload of the combination of the two tasks (system and driving).

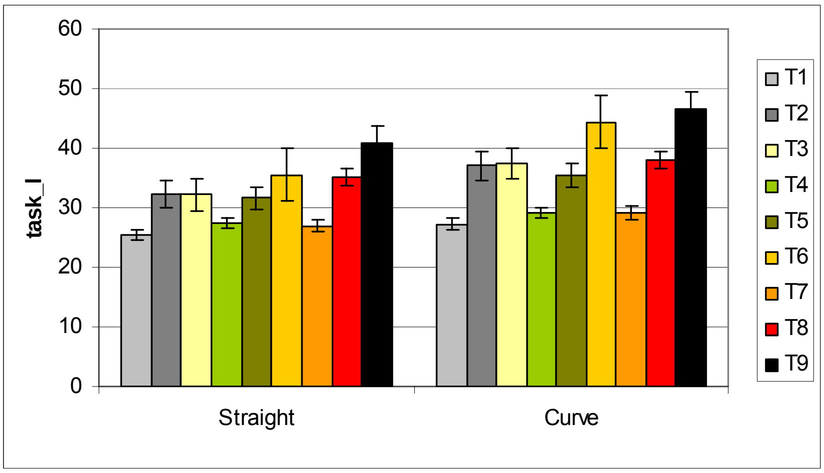


Figure 10 – Task length (task_l; seconds) for system C tasks on rural road

3.6.2. Comparison between groups of tasks

The type and difficulty level of tasks were created *a priori* with reference to the structure and dimension of the message and its instructions (see section 2.2.3).

Type of task was related to the goal of the task, and consequently, to the complexity of manipulation and interaction with the system display. In type 1 a new message was presented and the task just needed to be read; in type 2 the driver had to search for a specified message amongst a list of messages organised by number; type 3 was also a search task, but more difficult as the messages were presented in a random order. Results showed an increase in visual and manual demand, and consequently, an increase in task completion time across the three task types (see Figure 11).

Difficulty level was a factor directly related to the format of the sentences presented in each message: an increase in the number of words in a sentence corresponded to an increase in the presumed demand of the task. In addition, the number of visual and manual interactions with the system displays increased task completion time (see Figure 11).

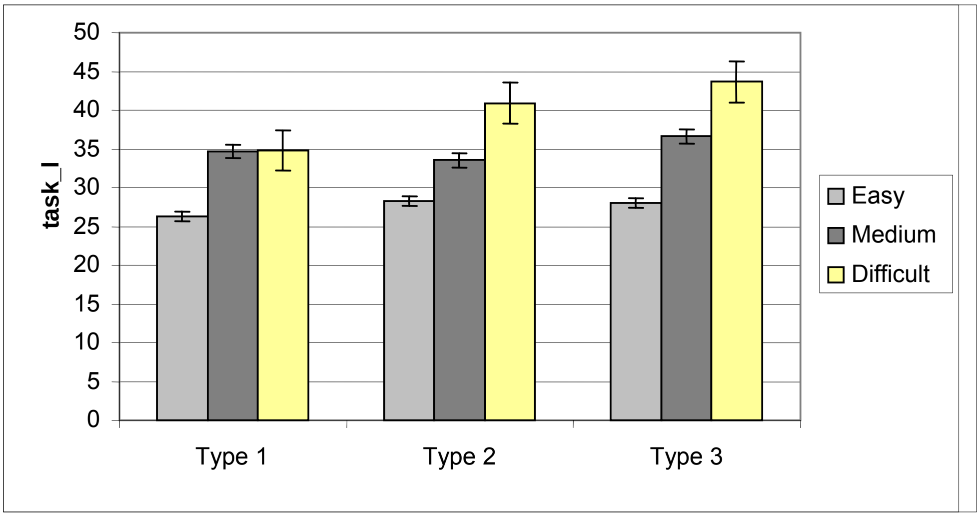


Figure 11 - Task length (task_l; seconds) distributed by difficulty level and type of task for system C on rural road

Regarding the analysis of the results already described, significant differences were found between baseline and performances in tasks belonging to different types and difficulty levels of System C tasks.

To assess the reliable effects of these different ways of grouping the tasks of System C and to analyse its influence on driving performance, an analysis of variance was run using type and difficulty level as factors. The results are presented in the following sections.

3.6.2.1. Task type

Effects of task type were only found in longitudinal related measures. Performance in task type 1 was significantly different from the other two types. Mean time and distance headway were found to increase for task type 2 and 3, when compared to task type 1. In addition, variation in these measures also increased. On the other hand, the minimum speed and time exposed to collision decreased in task types 2 and 3.

A behavioural effect was also observed when drivers believed it was becoming difficult to maintain control of the vehicle. The tasks included in categories 2 and 3 were search tasks, so they required a higher demand and therefore increased drivers' workload. This resulted in a decrease in speed of travel and an increase in distance to the lead car.

3.6.2.2. Difficulty level of the task

Analysis of results showed that the difficulty level of the task had effects on driving performance in terms of longitudinal control of the vehicle and the subjective reports of driving performance.

Drivers decreased the subjective appreciation of driving performance with an increase in difficulty level of the tasks. Significant differences between easy and hard levels were observed. This awareness of driving performance was highlighted by the difficulty in maintaining the vehicle control – with hard tasks the time and distance headway variation increased.

In the presence of more demanding conditions, drivers tried to develop strategies to compensate for this loss of control. They tended to reduce their mean and minimum speed trying to deal in an adequate way with the high workload of the situation. Analysis of variance showed an increase in mean distance and time headway from the lead car while drivers were performing the hardest tasks.

The results of this section highlighted that the organization of the tasks relative to their goals and structures was a suitable way of categorizing and assessing System C. Although this analysis helps in an overall assessment of the influence of the system in driving performance, a careful evaluation of all tasks and their effects will contribute to a more integrated knowledge about the equipment and measure sensitivity.

3.7. Result summary and conclusions

For System C, driving performance was seen to deteriorate in the presence of tasks and with an increase in task complexity. The analysis of the created groups of tasks highlighted the effects of the system not only with the simple presence of tasks but also showing how different goals and complexity of a task are influential in this matter. The results supported the a priori task categorization during the conceptualisation and selection of the system. But this analysis does not substitute a task-by-task evaluation.

The results show that the performance of the tasks was also influenced by road characteristics, with more detrimental effects from the curved sections. With the increase in road difficulty, the task completion time increased too, augmenting the difficulty in maintaining the lateral and longitudinal control of the car.

The effect of each task on driving performance did not differ too much, because all tasks were visual and manual. This explains the reliability in the identification of the effects of the tasks in the driving performance and the sensitivity of the same measures for all tasks.

The visual and manual demand required for System C produced more lateral deviation and consequently more frequent steering corrections, when compared to baseline driving. Thus the measures of steering reversals (rr_st) and standard deviation of lateral position (sd_lp) are sensitive measures for lateral control.

The hazardous effects of the tasks during driving, especially regarding the lateral control of the vehicle and the awareness of a more demanding secondary task induced a cautious approach from drivers. In this way, the presence and complexity of the system tasks were accompanied by a decrease in mean speed (mn_sp), mean distance headway (mn_hwd) and mean time headway (mn_hwt). Although this behaviour had a conservative function, during the task presentation the variation of headway increased (sd_hwd, sd_hwt), revealing a reduction in longitudinal control of the vehicle.

3.8. Measures summary tables

3.8.1. System C

		Task type 1			Task type 2			Task type 3			Effects		
Measure	**BL**	**T1**	**T2**	**T3**	**T4**	**T5**	**T6**	**T7**	**T8**	**T9**	**SLv**	**RLv**	**SLv*RLv**
taskl_l		26.35	34.71	34.80	28.33	33.55	40.91	28.04	36.61	43.67	✓	✓	×
subj_r	8.18	7.60	7.48	7.43	7.58	7.65	7.30	7.73	7.39	7.27	✓	✓	✓
mn_sp	82.71	81.37	80.15	79.56	77.53	79.93	78.59	80.76	79.55	78.21	✓	✓	×
st_sp	22.92	25.59	26.13	22.26	23.49	24.58	25.91	25.10	23.86	26.22	×	×	×
st_sp15	2.49	2.50	2.42	2.4894	2.25	2.75	2.27	2.25	2.22	2.27	×	×	×
st_sp30	3.77						3.37		3.37		×	×	×
d_sp	0.04	0.06	0.05	0.06	0.04	0.11	0.07	0.06	0.06	0.05	×	×	×
mn_hwd	52.51	55.80	61.86	59.63	67.44	59.24	76.52	60.59	62.77	79.58	✓	×	×
u_hwd	34.50	43.49	39.11	37.35	45.32	35.34	49.85	42.29	39.55	47.07	✓	×	×
sd_hwd	11.36	8.61	15.60	15.03	14.65	17.34	18.16	11.66	15.13	20.39	✓	×	×
sd_hwd15	4.345	4.658	5.518	4.65	5.6355	5.22	4.88	5.24	5.00	5.41	×	×	×
sd_hwd30	7.31						10.05		9.24		✓	×	×
u_hwt	1.48	1.95	1.78	1.71	2.13	1.54	2.30	1.92	1.86	2.20	✓	×	×
mn_hwt	2.35	2.54	2.93	2.85	3.30	2.79	3.73	2.80	3.01	3.81	✓	×	×
hwt_0_1	33.40	25.18	21.98	23.90	18.83	28.35	21.53	21.65	22.85	15.47	✓	×	×
hwt_1_2	28.33	35.85	28.80	27.86	22.63	20.30	18.70	34.80	21.23	21.99	✓	×	×
hwt_2_3	16.19	14.13	18.95	19.51	19.73	13.23	16.33	13.93	20.71	14.38	×	×	×
hwt_3_4	6.84	8.05	8.05	7.33	12.43	14.40	8.70	7.30	15.98	14.51	✓	×	×
hwt_4_5	4.867	4.05	7.475	9.08486	8.5	8.67	10.03	6.675	6.8	9.41	×	×	✓
hwt_5_6	3.23	3.85	4.40	2.54	5.90	5.88	7.40	4.53	2.87	4.65	×	×	×
hwt_6	7.33	8.90	10.35	9.77	12.50	7.90	17.33	11.13	9.34	19.60	✓	×	×
sd_hwt	0.55	0.43	0.81	0.77	0.76	0.89	0.97	0.56	0.76	1.02	✓	×	×
sd_hwt15	0.21	0.35	0.27	0.34	0.30	0.28	0.37	0.27	0.26	0.28	×	×	✓
sd_hwt30	0.36						0.544		0.468		×	✓	✓
mn_lp	1.78	1.81	1.79	1.83	1.83	1.85	1.84	1.77	1.77	1.87	✓	×	×
st_lp	0.31	0.32	0.34	0.37	0.3228	0.36	0.35	0.28	0.35	0.38	✓	✓	×
st_lp15	0.24	0.25	0.23	0.24	0.2342	0.24	0.24	0.22	0.23	0.25	×	✓	×
st_lp30	0.29						0.28		0.28		×	✓	×
rr_st1	11.96	14.44	15.62	16.09	15.31	15.53	15.63	16.07	15.30	15.76	✓	×	×
mn_tlc	-8.38	-50.14	15.78	-409.20	-237.47	-15.99	161.73	341.71	211.15	140.73	×	×	×
lnx	0.08	0.10	0.12	0.71	0.00	0.37	0.10	0.02	0.12	0.54	×	✓	×
tet	14.09	9.38	7.70	9.56	6.61	7.15	7.66	7.13	9.47	8.09	✓	×	×

(Significant difference from BL indicated by grey background)

4. The TNO driving simulator experiment

4.1. Test site

The experiment was conducted in the driving simulator of TNO at Soesterberg, The Netherlands. The TNO driving simulator consists of a mock-up of a BMW 318I, with normal controls, which is placed on a motion platform with six degrees of freedom. Different subsystems are used to run the vehicle model, generate images (computer generated image system, CGI; Evans & Sutherland SIMFusions), sounds, etc. Participants had a forward view of 120° horizontal (3 x 40°). The images were projected by three '3D Perception DLP Compact View SX15I' projectors.

4.2. Scenarios and participants

In total, twenty drivers participated in the experiment. The average age for the participants was 40.5 years (range: 27-49). They all had their driving license for at least five years and drove more than 10,000 kilometres a year.

4.2.1. Rural road

All participants drove on a rural road with a total width of 7.3 m, which included two carriageways of 3.65 m (lane markings included). The speed limit was 80 km/h.

4.3. IVIS included

Each participant performed tasks from System B (for task descriptions see section 2.2.2). For all partners performing a simulator or laboratory experiment, nine tasks per system were designed. However, one task (Task 5) could not be performed in this experiment, since the system broke down each time this task was initiated. Therefore, participants performed only eight tasks.

4.4. Experimental design

The experimental design does not differ from the general experimental design (see section 2.4). However, due to an error, three participants had all baseline conditions at the end of their drive.

4.5. Procedure

Participants were first instructed by a test leader about the experiment. Each participant practised the tasks in static mode and while driving in the simulator. The experiment started after the participant was familiar with all tasks. The entire experiment took about one and ah half hours to complete.

4.6. Measures and analysis method

The mandatory measures for the driving simulator were computed for this experiment (see section 2.7), with the exception of high steering frequency measure hi_st (hi_st2 was calculated instead) and the measure registering whether participants performed the task correctly (com_t). The subjective rating was obtained for all tasks with the exception of the baseline conditions. This was due to an error in the rating list. Because of too many missing data, not all mandatory measures could be statistically analysed or analysed for all tasks. For certain measures, a time window was used to overcome problems with the different task lengths (see section 2.7 and Appendix 3). When the task duration was less than the used time window, these tasks could not be analysed for the measures that depended upon the time window. Table 13 shows whether or not a particular measure was analysed.

Table 13 – The list of mandatory measures and whether or not they were analysed

Lane-position and TLC measures	
Variable short name	
mn_lp(m)	analysed
st_lp(m)	analysed for window 15s; available for 30s window but not analysed because of too many missing data
mn_tlc(s)	analysed
lnx(%)	analysed, but not normally distributed; analysed using Friedman ANOVA
pr_tlc(%)	analysed, but not normally distributed; analysed using Friedman ANOVA
Steering wheel measures	
rr_st1(1/minute)	analysed
rr_st3(1/minute)	analysed
hi_st2	analysed
Speed and Headway-related measures	
mn_sp(km/h)	analysed
u_sp(km/h)	analysed
st_sp(km/h)	analysed for window 15s; available for 30s window but not analysed because of too many missing data
d_sp(km/h)	analysed
sd_hwd(m)	available for both window 15s and 30s but not analysed because of too many missing data
mn_hwt(s)	analysed
mn_hwd	available but not analysed because of too many missing data
sd_hwt(s)	analysed for window 15s; available for 30s window but not analysed because of too many missing data
u_hwt(s)	analysed
hwt_0_1	analysed, but not normally distributed; analysed using Friedman ANOVA
hwt_1_2	analysed, but not normally distributed; analysed using Friedman ANOVA
hwt_2_3	analysed, but not normally distributed; analysed using Friedman ANOVA
hwt_3_4	analysed, but not normally distributed; analysed using Friedman ANOVA
hwt_4_5	analysed, but not normally distributed; analysed using Friedman ANOVA
hwt_5_6	analysed, but not normally distributed; analysed using Friedman ANOVA
hwt_6	analysed, but not normally distributed; analysed using Friedman ANOVA

As stated, Task 5 was not performed due to problems with the tested system. However, Task 5 is presented in the legends of all figures, in order to compare the results with other experiments of this report. In general, all tasks are presented in the legend even if they are not presented in the figure itself. Whiskers represent the +/- 95% confidence limits.

4.7. Results of System B

4.7.1. Task length

Average task length varied between 5.88 s (Task 1, straight road) and 67.73 s (Task 6, winding road; see Figure 12). A total of 90 seconds was analysed for baseline. The most difficult task took on average the longest time to complete (Task 6) while the easiest task took at average the shortest (Task 1).

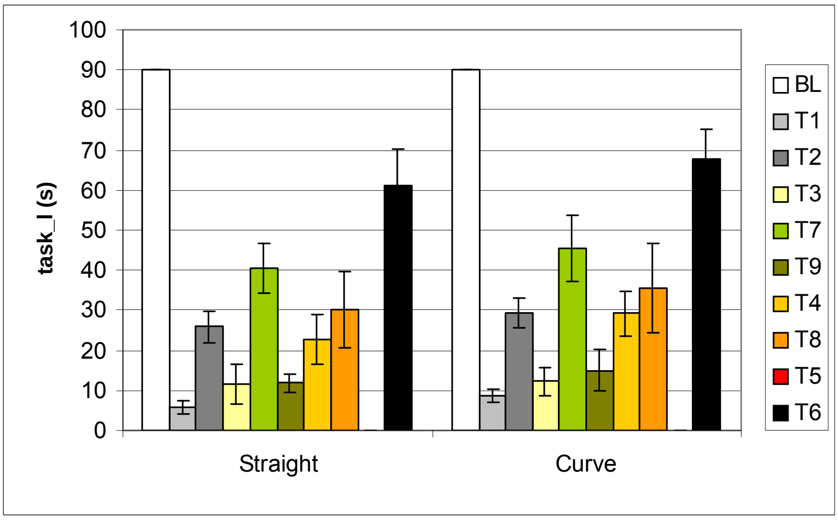


Figure 12 – Average task length for the different road and task levels

4.7.2. Self-reported driving performance

As stated, due to an error in the reporting list of the subjective rating, no subjective rating of driving performance was given by the participant during baseline conditions. It is reasonable to assume however that the average rating in the baseline condition would be higher than in the conditions in which participants performed a task (see Haste Deliverable WP2). However, we have made no assumptions about this and analysed only the ratings that were given. The results are presented in Figure 13.

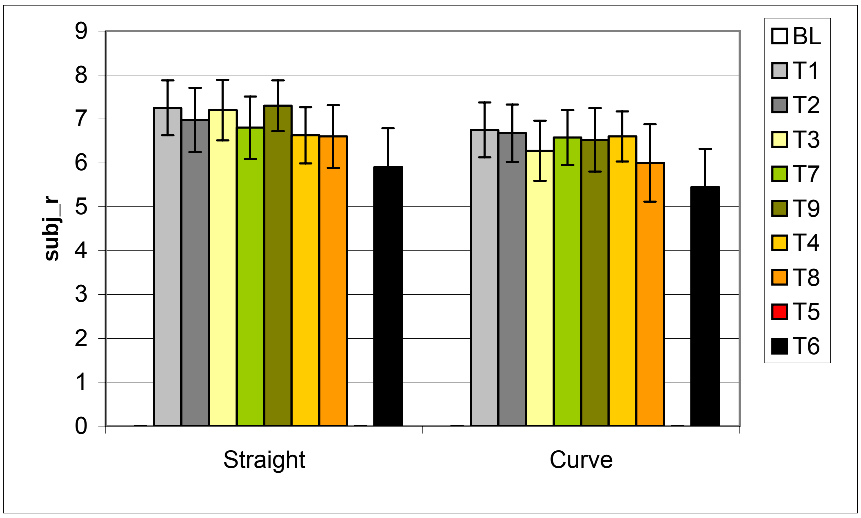


Figure 13 – Average subjective ratings for the different road and task levels

A main effect for road level and task level was found while the interaction was not significant. Driving performance on a straight section was rated higher than on a winding section. A post-hoc analysis showed that driving performance was rated lower during Task 6 (the most difficult task) than during the other tasks (with the exception of tasks four and eight).

4.7.3. Longitudinal control

An analyses of the average speed showed a main effect for road level and a main effect for task level (see Figure 14). On a straight section, participants drove on average two kilometres faster than during a winding section (70.79 km/h vs 68.84 km/h). The highest average speed was measured during Task 8 on a straight section (72.65 km/h) and the lowest speed was measured during Task 3 on a winding section (65.55 km/h). Post-hoc analysis showed that only the driving speed during Task 3 on a winding section differed from the winding baseline section.

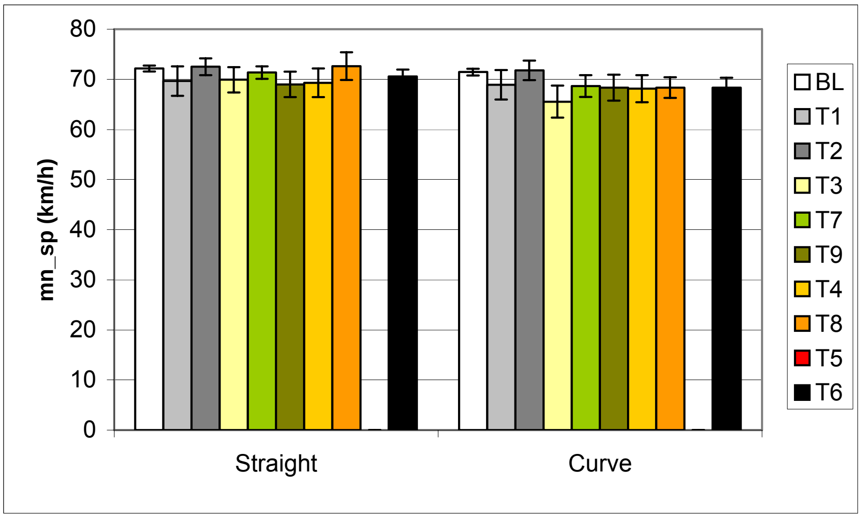


Figure 14 – The average speed for the different road and task levels

Most tasks lasted less than 30 seconds (see Figure 12). So, the standard deviation of speed was only analysed with a time window of 15 seconds and only for tasks two, seven, eight and six (and for the baseline conditions of course). The results showed no statistical effects for road level and task level (although the level of significance for the task level was just above 0.05).

A main effect for road level and task level was found for minimum speed. The results showed that the average minimum speed was lower on the winding sections than on the straight road sections (60.98 km/h vs 63.35 km/h). Post-hoc analyses showed that none of the minimum speeds during a task differed from the baseline conditions. The average minimum speed for each combination of task level and road level is presented in Figure 15.

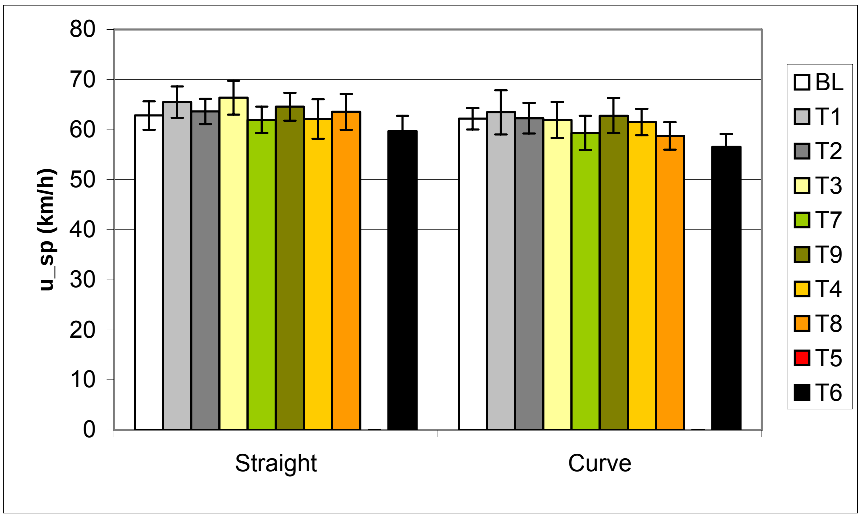


Figure 15 – The average minimum speed for the different road and task levels

The speed change per minute was also calculated. A negative number indicates that the speed at the beginning of the task was higher than at the end of the task while a positive number indicates the opposite. The average values are presented in Figure 16.

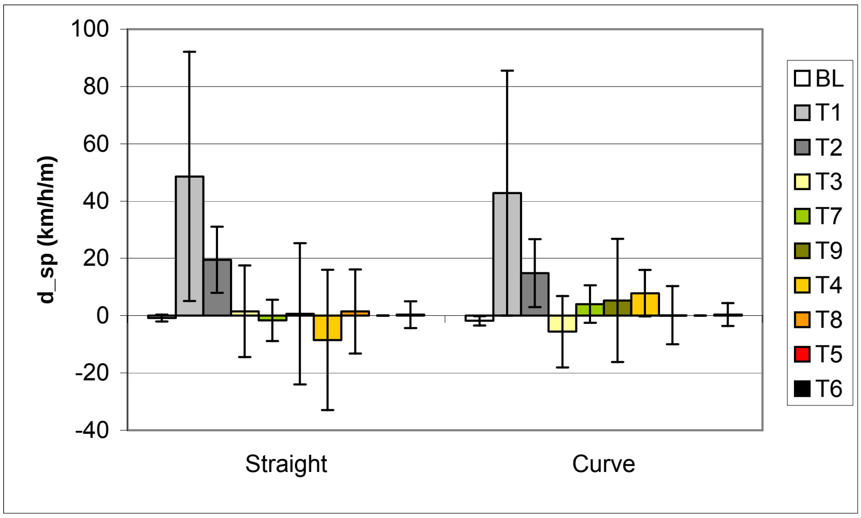


Figure 16 – The average speed change for the different road and task levels

The results showed a main effect for task level. Post-hoc analyses indicated that Task 1 in both the straight and winding conditions differed from the corresponding baseline conditions. The speed change seems large for Task 1; however these are speed changes per minute. On average, participants drove on the straight section at the end of Task 1 a bit faster (4.76 km/h) than at the beginning.

The analyses of mean time headway showed a main effect of road level and of task level (see Figure 17). On average, on the straight sections, participants followed closer than on the winding sections (4.11 s vs 5.27 s). Post-hoc analyses showed that on the straight sections none of the tasks differed significantly from the baseline condition, while on the winding sections almost all tasks (not Task 2 and 9) differed from the baseline condition. As can be seen in the figure, time headway was larger during these tasks than during the baseline condition. Participants also increased their distance with the leading vehicle (for most tasks the speed did not increase). It is however important to note that, on average, the distance between participants and the lead vehicle was already very large. With an average speed of around 70 km/h (which equals 19.44 m/s) and time headway of 3s, the average distance between the two cars is about 58 metres.

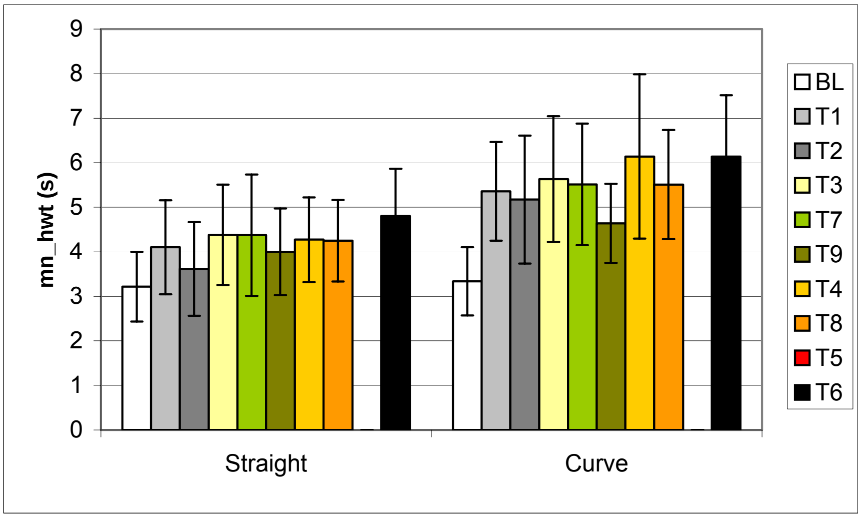


Figure 17 – The average mean time headway for the different road and task levels

To see whether there were any 'unsafe' time headways the minimum time headway was calculated. The averages are presented in Figure 18. Also in these cases, the average minimum time headways were quite long. Analyses showed a main effect for road level and task level. On straight sections the minimum time headway was shorter than on winding sections (3.25s vs. 4.20s). Post-hoc analyses showed that on the straight section Task 3 differed from the baseline condition, while on the winding section almost all tasks (not Tasks 2 and 9) differed from the baseline condition. The minimum time headway during these tasks was longer than during the baseline conditions. Nevertheless, the average minimum time headways were still very long. It seems that participants followed the lead vehicle at a large distance to ensure that they would not be too close when they had to perform a task.

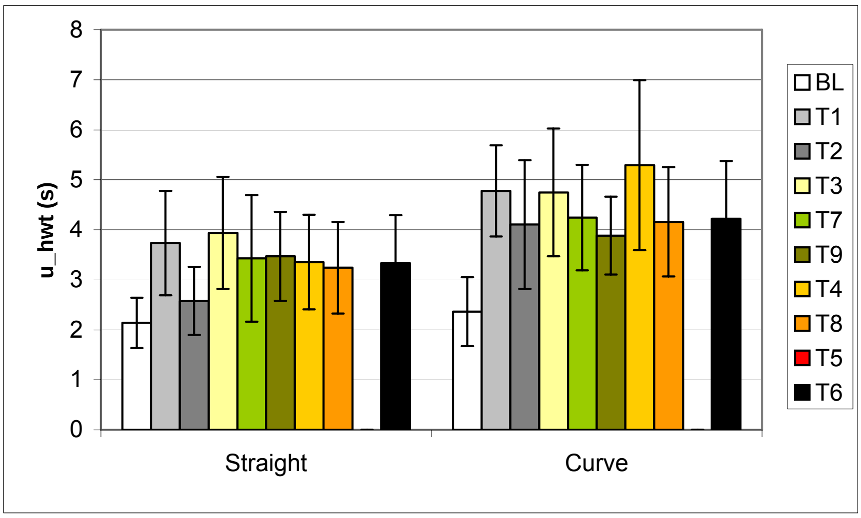


Figure 18 – The average minimum time headway for the different road and task levels

The distribution of time headway was also calculated per task and for each baseline condition. The distribution for the straight road sections is presented in Figure 19 and for the winding sections in Figure 20. In the baseline condition, on the straight road sections, participants mostly followed at 1-2 s time headway. This also applies for Task 2 while for the other tasks the top of the distribution is at time headway categories higher than 1-2 s. During the two easiest tasks (1 and 2) participants drove closer to the leading vehicle most of the time, compared to when more difficult tasks were performed.

On the winding sections, participants mostly followed at time headways of 1-2s during the baseline drive. However, during the tasks, participants mostly followed at a time headway larger than 6 seconds. A comparison of Figure 19 Figure 20 shows that participants followed closer on straight sections than on winding sections.

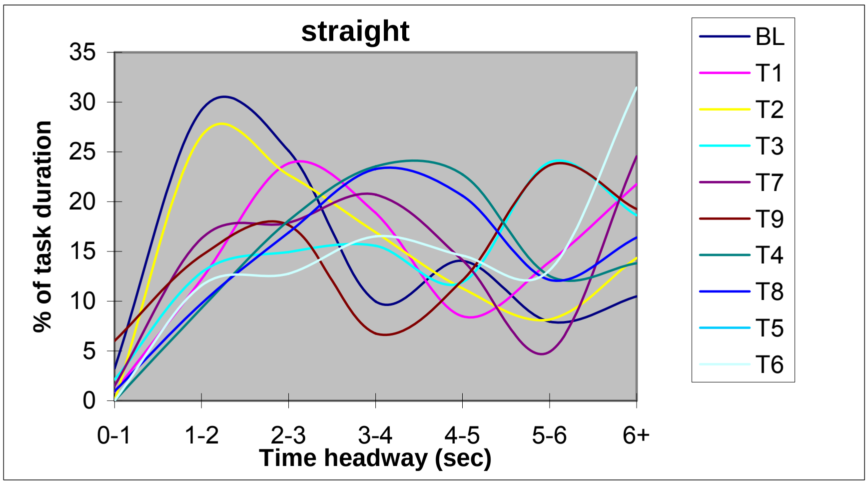


Figure 19 – Distribution of time headways for the different task levels on the straight sections

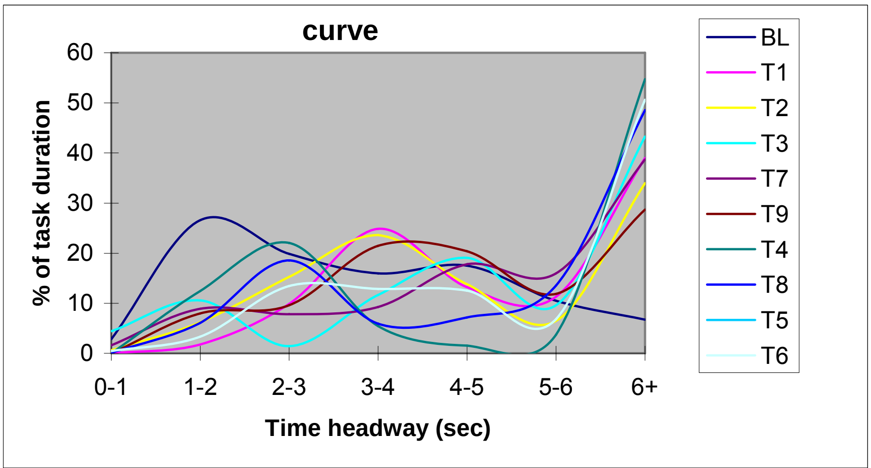


Figure 20 – Distribution of time headways for the different task levels on the winding sections

The different time headway categories were analysed with a Friedman ANOVA, since they were not normally distributed. This analysis only tests for main effects (thus for an effect between straight and winding road and between tasks). The results are presented in Table 14. The outcome of the Friedman ANOVA showed main effects for both road and task level on the time headway category 1-2 s, and 6s and larger, and a road level effect on 2–3s. Because post-hoc analysis was not possible, it is not clear which of the tasks differed from baseline. However, the large difference between hwt_6 in baseline (9.90) and the next task with the lowest percentage (T9: 23.97) seems to suggest that all tasks differed from baseline in this category.

Table 14 – The outcome of the Friedman ANOVA of the different time headway categories

Time headway category	Main Effect and Average Road level	Main Effect, Minimum, and Maximum Task level
hwt_0_1	✗ Straight: 1.61 Curve: 1.10	✗ Min: T4: 0.00 Max: T9: 3.26
hwt_1_2	✓ Straight: 15.90 Curve: 9.33	✓ Min: T1: 6.97 Max: BL:28.53
hwt_2_3	✓ Straight: 18.65 Curve: 13.02	✗ Min: T3: 8.18 Max: BL:21.55
hwt_3_4	✗ Straight: 16.84 Curve: 14.47	✗ Min: BL: 12.65 Max: T1:21.88
hwt_4_5	✗ Straight: 14.41 Curve: 13.56	✗ Min: T1: 10.83 Max: T9: 16.23
hwt_5_6	✗ Straight: 13.52 Curve: 9.91	✗ Min: T2: 7.33 Max: T9: 17.78
hwt_6	✓ Straight: 19.08 Curve: 38.61	✓ Min: BL: 9.90 Max: T6: 41.04

The variation in time headway was analysed with a time window of 15s (see section 4.6). Only a number of tasks could be analysed (tasks 2, 7, 8, and 6 and the baseline conditions). The results are presented in Figure 21.

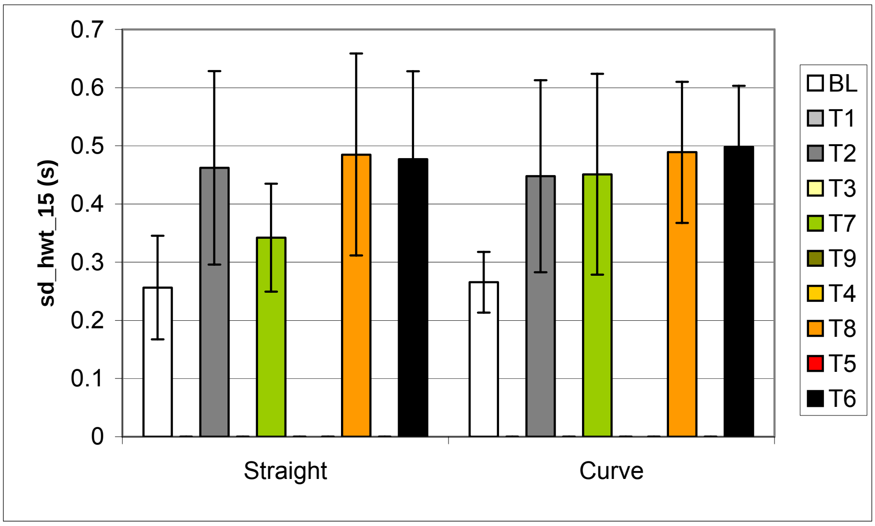


Figure 21 – The standard deviation time headway analysed with a window of 15s for the different road and task levels

An ANOVA showed a main effect for task level. A post-hoc analysis indicated that both on winding and straight sections tasks 8 and 6 differed from the baseline.

In general, the analyses regarding longitudinal control of the vehicle do not show clear-cut results with respect to effects of task level. With respect to speed, most tasks did not differ from baseline driving, suggesting that participants did not adjust their speed while performing a task. Effects were found with respect to following behaviour (time headway). Both minimum and mean time headway on winding sections showed an effect between tasks and baseline performance (with the exception of tasks 2 and 4). During task performance on winding sections, participants followed at larger time headways than in the baseline condition. With respect to the standard deviation of time headway, some task effects were found for both the winding and straight sections. These effects indicated a lower standard deviation during baseline driving, meaning that participants kept their 'distance' more constant. However, the time headway in the baseline condition was already large (more than three seconds). So, the importance of these effects is questionable.

4.7.4. Lateral control

The analysis of the average lateral position showed a main effect of task level. However, a post-hoc analysis showed no significant differences between a task and its corresponding baseline condition. The averages are presented in Figure 22.

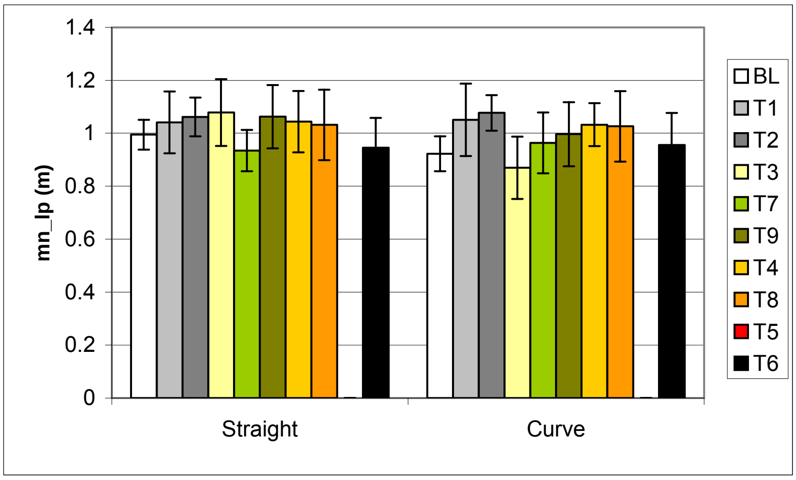


Figure 22 – The average lateral position for the different road and task levels

The amount of swerving as indicated by the standard deviation of the lateral position was analysed with a time window of 15s for tasks 2, 7, 8, and 6. The average amount of swerving is presented in Figure 23.[1] The results of the ANOVA showed a main effect for road level and for task level. Participants swerved more on winding sections than on straight sections. Post-hoc analyses indicated that on both the winding and straight sections, Task 6 differed from the corresponding baseline. Also, on the straight sections, Task 8 differed from the baseline and on the winding sections Task 7 differed from the baseline.

The variable time-to-line crossing (TLC) indicates the remaining time for the vehicle to leave its lane when speed and heading remain unchanged. Of this value, the average of the minima was calculated and is presented in Figure 24 (tasks 1 and 3 were not included in the analyses because of too many missing data). The results of the analyses showed a main effect for road level and for task level. The TLC on straight sections was longer than on winding sections. Post-hoc analyses showed that on straight sections the TLC decreased when participants had

[1] It is important to indicate that in the analysis of the lateral position no limits were imposed upon the lateral position. This means that even if a participant was completely out of its lane the data were still included in the calculation of the average and standard deviation of the lateral position. Consequently this lateral behaviour does not necessarily reflect driving behaviour within a lane.

to perform a task (with the exception of Task 7). On the winding sections none of the tasks differed from the baseline condition.

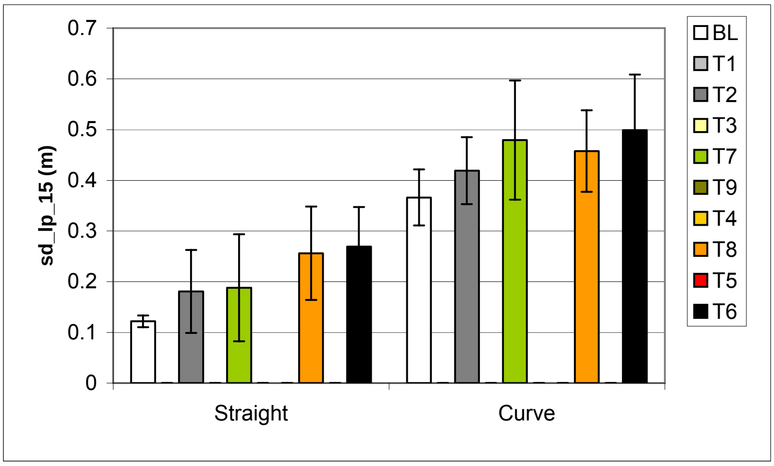


Figure 23 – The standard deviation of lateral position ('swerving') analysed with a window of 15s for the different road and task levels

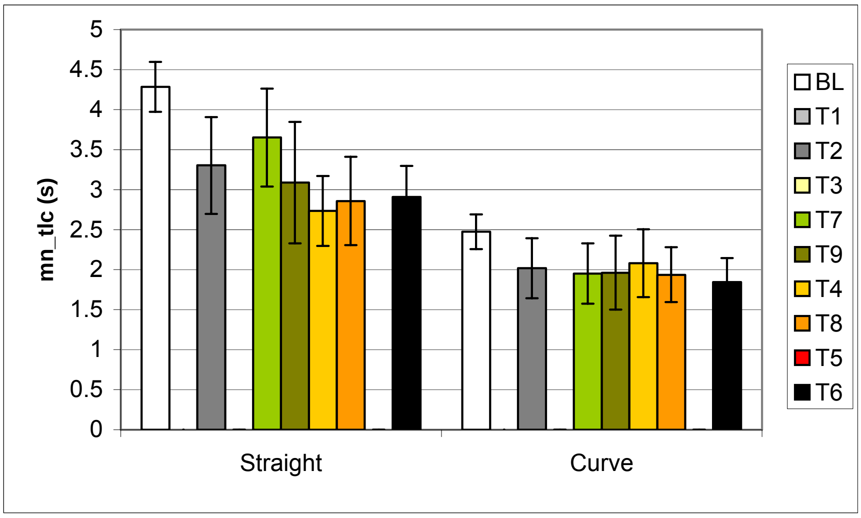


Figure 24 – The average minimum time-to-line crossing for the different road and task levels.

As stated earlier, the percentage of time that the time-to-line crossing was shorter than 1 second and the percentage of time that the participant drove outside the lane were not normally distributed. Therefore, the 'regular' ANOVA could not be performed. Instead a Friedman ANOVA was performed which is a non-parametric analysis method. This test only provided main effects, not interaction effects and no post-hoc analyses were possible.

The average values for percentage of time that TLC was shorter than 1 second are presented in Figure 25 (tasks 1 and 3 were excluded because of too many missing data). The analyses showed a main effect of road level and task level. The percentage of time that TLC was shorter than 1 second was found to be higher for the winding sections than for the straight sections. With respect to the main effect of task level, it is impossible to say which tasks (if any) differed from the baseline conditions. However as illustrated in Figure 25 all percentage values were higher during task performance than in the baseline condition. It seems reasonable to assume that if tested, tasks with a larger difference from the baseline would reach significance (e.g., tasks 2, 8, and 6 on the winding sections and tasks 4, 8, and 6 on the straight sections).

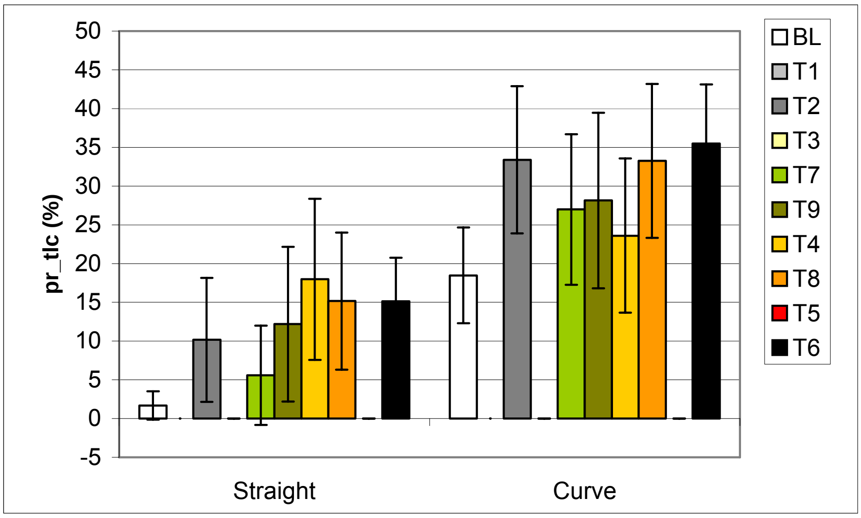


Figure 25 – The average percentage of time during a task that the TLC was shorter than 1 second for the different road and task levels

The average percentage time that participants drove (partly) outside their lane during task performance is presented in Figure 26. Analyses showed only a main effect of road level. On winding sections, participants drove a higher percentage of time (partly) outside their lane than on straight sections.

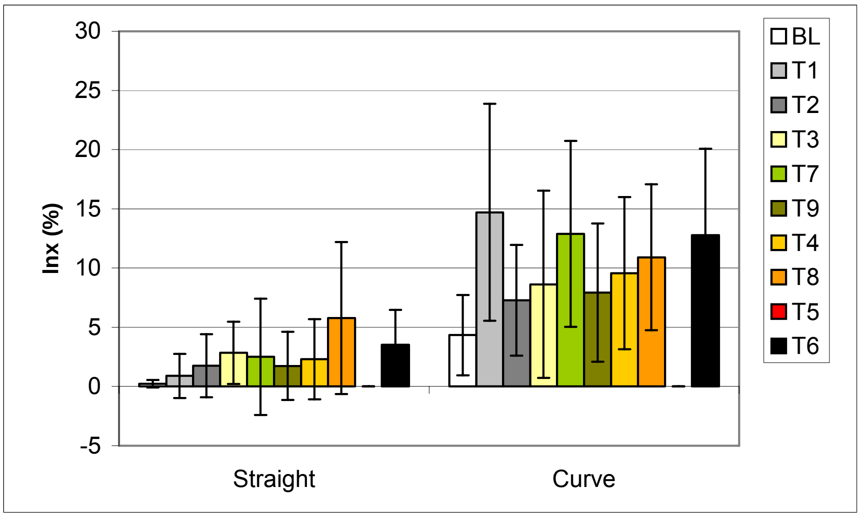


Figure 26 – The average percentage of lane crossings for the different road and task levels

Different measures were calculated to indicate steering effort. These were high steering frequency and steering reversals. Average values for high steering frequency area (hi_st2) are presented in Figure 27. Tasks 1, 3, and 9 were not analysed because of too many missing data.[2] The ANOVA showed a main effect for road and task level. On straight sections, the proportion of high frequency steering movements were larger than on winding sections. The post-hoc analysis for the main effect of task level showed that on the straight sections tasks 7, 8, and 6 differed from the baseline condition, while on winding sections only Task 7 differed from the baseline condition (tasks 8 and 6 were almost significant, $p < 0.07$).

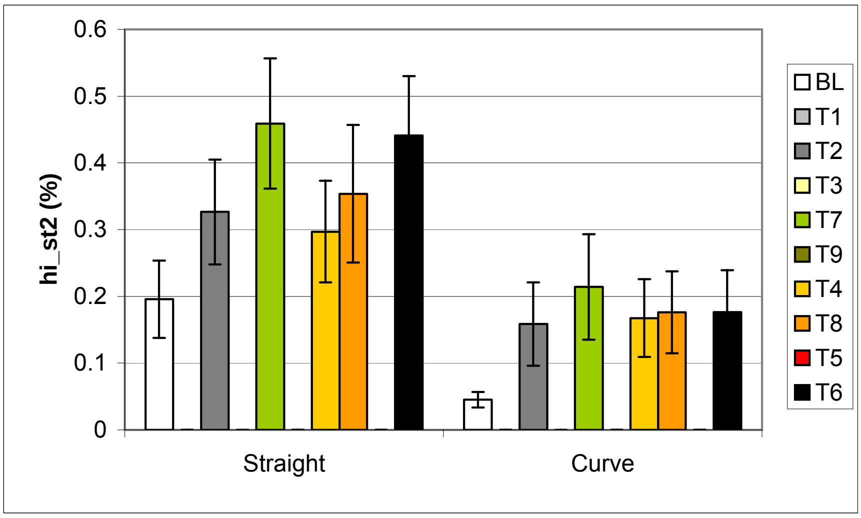


Figure 27 – The high frequency steering area for the different road and task levels

The steering reversals were analysed with a gap of one and three degrees. The average values with an amplitude of one (rr_st1) are presented in Figure 28 and those with an amplitude of three (rr_st3) are presented in Figure 29.

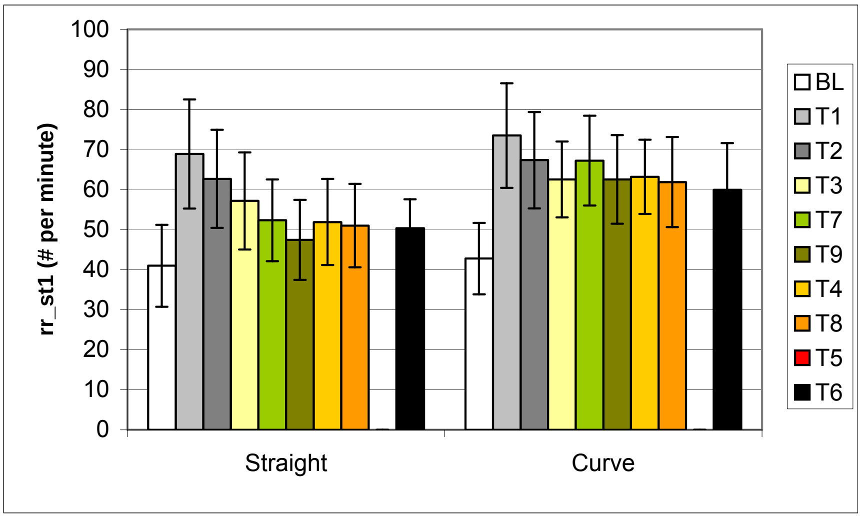


Figure 28 – The average steering reversal rate with a gap of one degree for the different road and task levels

[2] With a short task duration, the high frequency steering analysis becomes rather meaningless. In tasks 1, 3, and 9 there were too many participants with a task duration too short for a meaningful analysis, resulting in missing data.

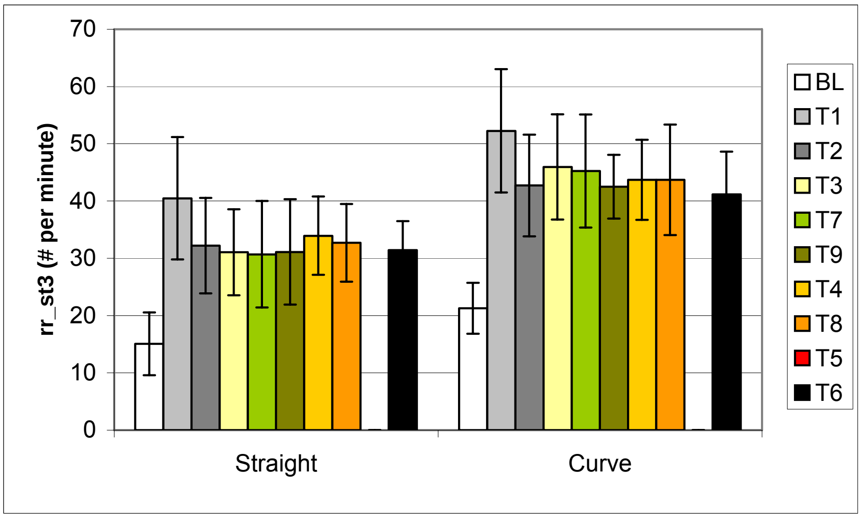


Figure 29 – The average steering reversal rate with a gap of three degrees for the different road and task levels

The results of the ANOVA of rr_st1 data showed main effects for road and task level. On straight sections the 'number of steering reversals' were lower than on winding sections. The post-hoc analysis indicated that on straight sections the relatively easy tasks differed from the baseline condition while on the winding sections all tasks differed from the baseline condition. A similar pattern of results was found for rr_st3, with the exception that for this measure on the straight sections also all tasks differed from the baseline condition.

With respect to the lateral control performance, a limited number of effects were found with respect to task level. Only the steering reversal rates showed clear effects of task performance on both straight and winding sections (rr_st3) or only on winding sections (rr_st1). With respect to average time-to-line crossing, some effects were found for certain tasks on straight sections.

4.8. Result summary and conclusions

This driving experiment investigated the relation between secondary task performance (operating a route guidance system) and driving behaviour. Participants performed secondary tasks while driving on a straight road or on a winding road. The relevant questions were (i) whether driving behaviour during secondary task performance would differ from driving in a baseline condition (no secondary tasks), (ii) whether there would be an effect of task difficulty and (iii) whether there would be a difference between driving along a straight road or on a winding road.

The results of the analyses showed different effects of secondary task performance on driving behaviour compared to driving in the baseline condition. On the straight sections, the effect of secondary task performance was apparent on mean time-to-line crossing and the steering reversal rate with a gap of three degrees.[3] On the winding sections, effects were found on mean time headway, minimum time headway and both steering reversal measures.[3] Table 15 shows which measures indicated an effect of task performance on the different road levels and

[3] For the other measures there are too few effects of some tasks on driving behaviour to generally speak of influence of task performance.

for the different tasks. It shows that on the straight sections, an effect of all tasks was found on the steering reversal rate with amplitude of three, whereas for the winding sections an effect of all tasks was found for both calculated reversal rates. Furthermore, Table 15 also shows that more effects were found on winding road sections.

Table 15 – Measures that showed an effect of secondary task performance on driving behaviour

Tasks	Straight	Curve
T1	d_sp, rr_st1, rr_st3	d_sp, mn_hwt, u_hwt, rr_st1, rr_st3
T2	mn_tlc, rr_st1, rr_st3	rr_st1, rr_st3
T3	u_hwt, rr_st1, rr_st3	mn_sp, mn_hwt, u_hwt, rr_st1, rr_st3
T7	hi_st2, rr_st3	mn_hwt, u_hwt, hi_st2, rr_st1, rr_st3
T9	mn_tlc, rr_st3	rr_st1, rr_st3
T4	mn_tlc, rr_st3	mn_hwt, u_hwt, rr_st1, rr_st3
T8	st_hwt_15, st_lp15, mn_tlc, hi_st2, rr_st3	mn_hwt, st_hwt_15,u_hwt, rr_st1, rr_st3
T5	-	-
T6	st_hwt_15, st_lp15, mn_tlc, hi_st2, rr_st3	mn_hwt, st_hwt_15,u_hwt, st_lp_15, rr_st1, rr_st3

The post-hoc analyses hardly ever showed a significant difference among the different tasks. Therefore, the tasks could not really be categorised by their different effects on driving performance. It is clear from the table above that the more difficult tasks (8 and 6) produced effects more often than the easier tasks. However, on the winding sections tasks 8 and 6 show a number of effects but so do tasks 1, 3, and 7. So, one can conclude that the more difficult tasks show effects irrespective of the road level.

It is important to note that the type of effects differ with road level. The exception is the steering reversal rate with amplitude of three degrees. This measure showed an effect both on straight and winding sections and for all tasks, which suggests that it is rather sensitive to (any type of) secondary task performance. However, the direct relation with traffic safety is less obvious (compared for example with time headway).

The results of the present experiment show that, to test the effect of an IVIS on driving performance, it does not really matter whether straight or winding road segments are used, since effects were found on both road levels. However, if a winding road is chosen, one can either use easy or difficult tasks, while with a straight road one should use difficult tasks. Among the variables that can be measured one should certainly include the steering signal and calculate the steering reversal rate with an amplitude of three degrees.

4.9. Measures summary tables

4.9.1. System B

Averages for the separate road levels.

	Straight										Curve											
		Visual		Visual-Manual								Visual		Visual-Manual								
	BL	T1	T2	T3	T7	T9	T4	T8	T5	T6	BL	T1	T2	T3	T7	T9	T4	T8	T5	T6	effect RLv	effect SLv
task_l	90.00	5.88	25.83	11.50	40.46	11.80	22.78	30.15	-	61.16	90.00	8.77	29.34	12.27	45.45	14.99	29.22	35.60	-	67.73		
subj_r	-	7.25	6.98	7.20	6.80	7.30	6.63	6.60	-	5.90	-	6.75	6.68	6.28	6.58	6.53	6.60	6.00	-	5.45	✓	✓
mn_sp	72.17	69.65	72.52	69.91	71.35	68.98	69.30	72.65	-	70.58	71.45	68.91	71.80	65.55	68.68	68.35	68.12	68.35	-	68.35	✓	✓
st_sp15	2.83	-	4.70	-	3.83	-	-	4.24	-	3.84	3.01	-	4.49	-	4.12	-	-	4.17	-	3.79	✗	trend
u_sp	62.82	65.46	63.64	66.37	61.95	64.56	62.11	63.55	-	59.70	62.17	63.45	62.26	61.94	59.36	62.80	61.52	58.76	-	56.57	✓	✓
d_sp	-0.84	48.60	19.50	1.53	-1.67	0.68	-8.47	1.48	-	0.35	-1.77	42.79	14.86	-5.55	4.05	5.28	7.83	0.15	-	0.43	✗	✓
mn_hwt	3.22	4.10	3.62	4.38	4.37	4.00	4.27	4.25	-	4.80	3.34	5.36	5.18	5.63	5.51	4.64	6.14	5.51	-	6.14	✓	✓
st_hwt_15	0.26	-	0.46	-	0.34	-	-	0.49	-	0.48	0.27	-	0.45	-	0.45	-	-	0.49	-	0.50	✗	✓
u_hwt	2.14	3.73	2.58	3.94	3.43	3.47	3.36	3.24	-	3.33	2.36	4.78	4.10	4.75	4.24	3.88	5.29	4.16	-	4.22	✓	✓
hwt_0_1*	3.24	0.90	0.00	2.11	1.50	6.01	0.00	0.96	-	0.00	2.82	0.00	0.64	4.42	1.62	0.00	0.00	0.12	-	0.49	✗	✗
hwt_1_2*	29.20	12.19	26.62	12.90	16.31	14.58	9.30	9.75	-	11.57	26.60	1.76	6.21	10.58	8.91	7.96	12.48	6.09	-	3.31	✓	✓
hwt_2_3*	25.11	23.85	22.68	14.95	17.89	17.64	18.09	16.90	-	12.76	19.85	9.89	15.26	1.41	7.79	9.61	22.08	18.52	-	13.43	✓	✗
hwt_3_4*	9.97	18.87	16.93	15.55	20.73	6.76	23.54	23.26	-	16.53	15.94	24.88	23.62	11.63	9.24	21.45	5.45	5.94	-	12.84	✗	✗
hwt_4_5*	14.06	8.48	11.29	11.89	14.13	12.09	22.73	20.59	-	14.57	17.54	13.19	13.78	19.05	17.78	20.38	1.53	7.19	-	12.51	✗	✗
hwt_5_6*	7.95	13.97	8.14	23.97	4.91	23.67	12.53	12.14	-	13.12	10.54	11.32	6.52	9.60	15.97	11.90	3.75	13.56	-	6.81	✗	✗
hwt_6*	10.48	21.74	14.35	18.65	24.54	19.24	13.82	16.40	-	31.45	6.72	38.97	33.98	43.30	38.68	28.71	54.72	48.58	-	50.62	✓	✓
mn_lp	0.99	1.04	1.06	1.08	0.93	1.06	1.04	1.03	-	0.95	0.92	1.05	1.08	0.87	0.96	1.00	1.03	1.03	-	0.96	✗	✓
st_lp15	0.12	-	0.18	-	0.19	-	-	0.26	-	0.27	0.37	-	0.42	-	0.48	-	-	0.46	-	0.50	✓	✓
mn_tlc	4.29	-	3.30	-	3.65	3.09	2.73	2.86	-	2.91	2.47	-	2.02	-	1.95	1.96	2.08	1.94	-	1.85	✓	✓
pr_tlc*	1.69	-	10.16	-	5.59	12.18	17.96	15.16	-	15.12	18.47	-	33.39	-	26.97	28.14	23.61	33.26		35.47	✓	✓
lnx*	0.23	0.89	1.75	2.84	2.50	1.73	2.30	5.78	-	3.51	4.34	14.71	7.28	8.63	12.89	7.93	9.56	10.91	-	12.78	✓	✗
hi_st2	0.20	-	0.33	-	0.46	-	0.30	0.35	-	0.44	0.05	-	0.16	-	0.21	-	0.17	0.18	-	0.18	✓	✓
rr_st1	40.97	68.86	62.63	57.16	52.32	47.42	51.88	51.02	-	50.34	42.77	73.49	67.33	62.52	67.22	62.51	63.16	61.85	-	59.96	✓	✓
rr_st3	15.06	40.49	32.24	31.07	30.71	31.13	33.96	32.71	-	31.43	21.28	52.26	42.72	45.96	45.25	42.51	43.71	43.72	-	41.18	✓	✓

Averages over Road levels

		Visual		Visual-Manual						
	BL	T1	T2	T3	T7	T9	T4	T8	T5	T6
task_l	90.00	7.33	27.59	11.89	42.96	13.40	26.00	32.88	-	64.45
subj_r	-	7.00	6.83	6.74	6.69	6.92	6.62	6.30	-	5.68
mn_sp	71.81	69.28	72.16	67.73	70.02	68.67	68.71	70.50	-	69.47
st_sp15	2.92	-	4.60	-	3.98	-	-	4.21	-	3.82
u_sp	62.50	64.46	62.95	64.16	60.66	63.68	61.82	61.16	-	58.14
d_sp	-1.31	45.70	17.18	-2.01	1.19	2.98	-0.32	0.82	-	0.39
mn_hwt	3.28	4.73	4.40	5.01	4.94	4.32	5.21	4.88	-	5.47
st_hwt_15	0.27	-	0.46	-	0.40	-	-	0.49	-	0.49
u_hwt	2.25	4.26	3.34	4.35	3.84	3.68	4.33	3.70	-	3.78
hwt_0_1*	3.03	0.45	0.32	3.26	1.56	3.01	0.00	0.54	-	0.24
hwt_1_2*	27.90	6.97	16.42	11.74	12.61	11.27	10.89	7.92	-	7.44
hwt_2_3*	22.48	16.87	18.97	8.18	12.84	13.63	20.08	17.71	-	13.10
hwt_3_4*	12.96	21.88	20.27	13.59	14.98	14.11	14.49	14.60	-	14.68
hwt_4_5*	15.80	10.83	12.53	15.47	15.96	16.23	12.13	13.89	-	13.54
hwt_5_6*	9.24	12.64	7.33	16.78	10.44	17.79	8.14	12.85	-	9.96
hwt_6*	8.60	30.35	24.16	30.97	31.61	23.97	34.27	32.49	-	41.04
mn_lp	0.96	1.05	1.07	0.98	0.95	1.03	1.04	1.03	-	0.96
st_lp15	0.25	-	0.30	-	0.34	-	-	0.36	-	0.39
mn_tlc	3.38	-	2.66	-	2.80	2.53	2.41	2.40	-	2.38
pr_tlc*	10.08	-	21.78	-	16.28	20.16	20.79	24.21	-	25.30
lnx*	2.29	7.80	4.52	5.74	7.70	4.83	5.93	8.35	-	8.15
hi_st2	0.13	-	0.25	-	0.34	-	0.24	0.27	-	0.31
rr_st1	41.87	71.18	64.98	59.84	59.77	54.97	57.52	56.44	-	55.15
rr_st3	18.17	46.38	37.48	38.52	37.98	36.82	38.84	38.22	-	36.31

* No post-hoc analysis could be performed on these measures.

5. The Transport Canada simulator experiment

5.1. Test site

Drivers were tested using Transport Canada's DriveSafety TM500c fixed-based simulator located in Ottawa Canada, which consists of seven inter-connected Pentium computers, including five graphics display computers, an authoring station, and a vehicle dynamics computer (see Figure 30). Graphics were displayed using five 81.3 cm x 61.0 cm LCD rear projection monitors in front of the driver, each providing a 50 degree field of view. A rear-view mirror was located on the top right corner of the centre panel, and the side view mirrors were located on the bottom of the right and left screens adjacent to the centre panel. The simulator housing consisted of the drivers' portion of a Saturn sedan cab, with an adjustable car seat, steering wheel, gas/brake pedal, instrument cluster and gearshift. Auditory cues included throttle-linked engine noise and wind noise when the subject vehicle passed oncoming vehicles.

Figure 30 – Transport Canada's DriveSafety driving simulator

5.2. Scenarios and participants

The design of the scenario was based on the rural road design described in the HASTE internal deliverable "WP3 Experimental Designs". The DriveSafety simulator scenario authoring system is based on pre-designed tiles that are linked together to form the basic route. This resulted in some minor deviations in segment length from the WP3 general route design.

This research was reviewed and approved by a Human Participant Use Review Board. In the assessment of System A, there were twenty paid participants (13 males, 7 females) with a mean age of 23.35 who completed the route. One participant did not complete due to the onset of mild simulator sickness; consequently an additional participant was run. Of the twenty participants who drove using System D, there were 15 males and 5 females with a mean age of 22.70. All drivers held a valid G-Class Driver's license, and drove regularly (minimum three years; on average approximately 15,000 km annually). All participants had normal or corrected-to-normal vision.

5.3. IVIS included

Transport Canada tested two systems. For System A, a screen was mounted on the dashboard above the centre stack and drivers interacted with the system by using a remote control. System D was in a PDA format and required the use of a stylus.

5.4. Experimental design

IVIS system was run between subjects, with twenty subjects using each system. Road Complexity Level (straight or curved) and Tasks were within subject variables. All subjects performed the PDT task while driving.

5.5. Procedure

Upon their arrival at the laboratory, participants' demographic information was collected and instructions regarding the experiment were given. The participant then practiced the chosen in-vehicle navigation tasks with the experimenter. For some tasks, the participant simply listened to recorded information, and responded with a verbal answer. For other tasks, they interacted with the in-vehicle navigation system using a remote control or stylus to get desired information.

The participants' first in-car practice scenario involved driving the simulator to become familiar with its features. At this time, the participant also practiced performing the perceptual detection task (PDT) which consisted of the visual presentation of a red square in the forward view of the scenario presented on screen. The driver's task was to respond to the squares as quickly as possible using a finger switch worn on his or her left hand.

The final practice session required the subject to complete the navigation tasks and perform the perceptual detection task while driving the simulator. At the end of each in-vehicle navigation system task, the participant heard a prompt "please rate your workload", and was required to say aloud a number between one and ten to reflect the difficulty of performing the task while driving. When the subject felt that he or she was comfortable with the tasks, a small break was taken before data collection began.

Data were collected for a total of 18 tasks; nine of which occurred on straight road segments, and nine similar tasks that occurred on curved road segments. Additionally there were three baseline segments during which the subject was required only to drive the route and perform the PDT, without having to complete any additional tasks. The subjects' chosen speed during the initial baseline segment was used to calculate the speed of a slower-moving lead vehicle (small sedan) that appeared in front of the subjects' vehicle after some time. As a result, the lead vehicles' speed was programmed to be 10% less than that of the drivers' own speed in the baseline section. Participants were required to follow the slower-moving lead vehicle for the remainder of the session, without passing, or driving too far behind it. Four video cameras were installed in the cab, providing a view of the subjects face, over their right shoulder, a view of the driving scene, and a view of the navigation system's interface with which the subject was interacting. Video recording was completed using a VCR to record camera feed and the audio signal provided from the microphone installed to record participants' oral responses. Data collection concluded when the subject reached a traffic light at the end of the route. Each scenario took between 60 to 80 minutes to complete, depending on the drivers' chosen speed. Finally, participants were asked a short series of questions, and were debriefed on the purpose of the present study.

5.6. Measures and analysis method

All mandatory variables described earlier were collected and analysed with the exception that Transport Canada collected Workload ratings rather than subjective driving ratings. In addition, the PDT variables were collected and analysed from the optional measures list. All the dependant variables are presented in Table 16. Only the PDT measures out of the optional measures were analyzed.

Table 16 – Measures collected by Transport Canada in the WP3 simulator experiments

Mandatory_sim/lab	Optional_sim/lab
subj_wl	lnx(%)
mn_sp(km/h)	pr_tlc(%)
u_sp(km/h)	rswt_20(1/minute)
st_sp(km/h)	rswt_40(1/minute)
d_sp(km/h)	rswt_70(1/minute)
mn_hwd(m)	hi_st2
sd_hwd(m)	en_st(-)
u_hwt(s)	st_ga(deg)
mn_hwt(s)	n_gl
sd_hwt(s)	tot_gl
mn_lp(m)	PRC(%)
st_lp(m)	mn_gd(s)
rr_st1, rr_st3 (1/minute)	pdt_hit
mn_tlc(s)	pdt_rt
hi_st	pdt_miss
com_t(y/n)	pdt_cheat
st_sp15, st_sp30	
sd_hwt15, sd_hwt30	
sd_hwd15, sd_hwd30	
st_lp15, st_lp30	
hwt_0_1, hwt_1_2, hwt_2_3, hwt_3_4, hwt_4_5, hwt_5_6	
hwt_6	
task_t (s)	

5.7. Results

5.7.1. Effects of System A

5.7.1.1. Drivers' reported Workload

In the Transport Canada study, drivers rated their perceived workload for each of the tasks on a scale from 1 (low) to 10 (high). The mean workload ratings for tasks and road complexity are presented in Figure 31. A significant main effect was found for Task. The pattern of workload ratings is consistent with the expected level of difficulty for the tasks within the auditory and visual-manual groupings of the tasks.

The data for Road Complexity Level indicated that overall workload ratings were significantly higher on average for curved segments (4.72) compared with straight segments (4.27). The interaction was not significant.

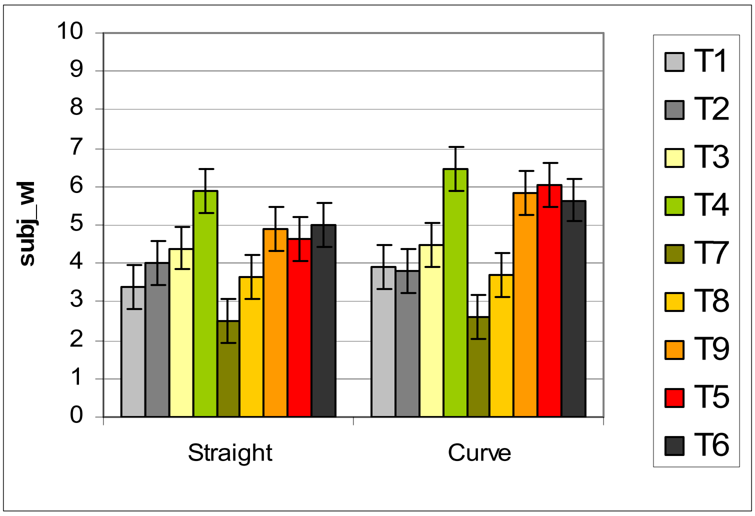


Figure 31 – Rated workload for the tasks performed with system A

5.7.1.2. Longitudinal control

A number of speed and headway-related measures were collected. The Mean Speed data are provided for tasks and road complexity in Figure 32. The Task effect was significant. The general pattern of the data reflects that drivers reduced their speed when engaged with the system tasks. The exceptions are for Tasks 3, 4 and 7 which did not differ from the baseline. The Mean Speed for Task 8 was marginally less than that for baseline. Neither the main effect for Road Complexity nor the interaction was significant.

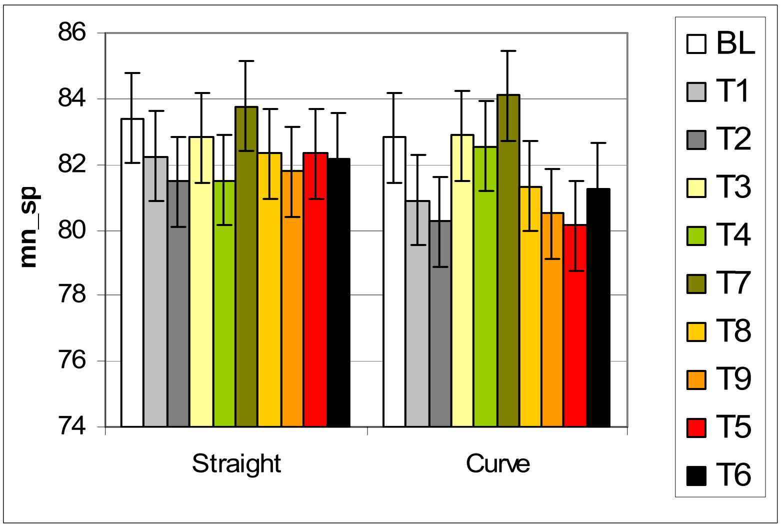


Figure 32 – Mean speed for the tasks performed with system A

With the exception of Task 6, the standard deviation of speed for all tasks was less than that observed for the baseline condition (see Figure 33). This was not an expected finding based on previous research and is suspected to be an artefact of the way in which the baseline data were calculated. Consequently, two other methodologies using a "moving window" technique were explored and are presented below. A description of the moving window technique is provided in Appendix 2.

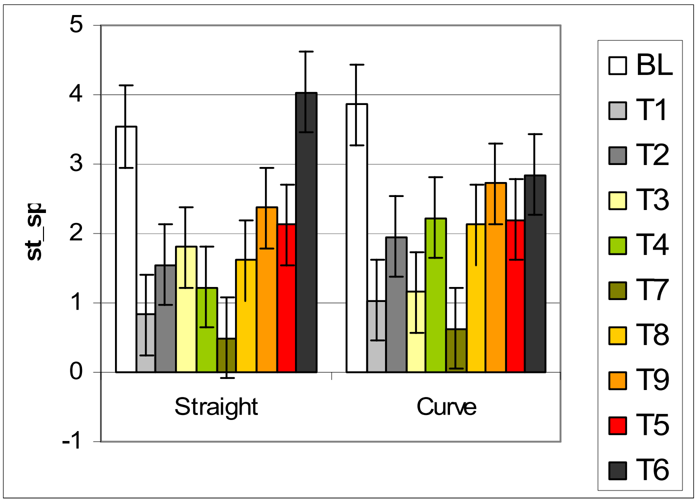


Figure 33 – Standard deviation of speed for tasks performed with system A

The procedure of using the 15s window for the analysis of the Standard Deviation of Speed data resulted in Tasks 1 and 7 being eliminated from the analysis as duration of each of these tasks was less than 15s. These data are presented in Figure 34. The analysis did result in

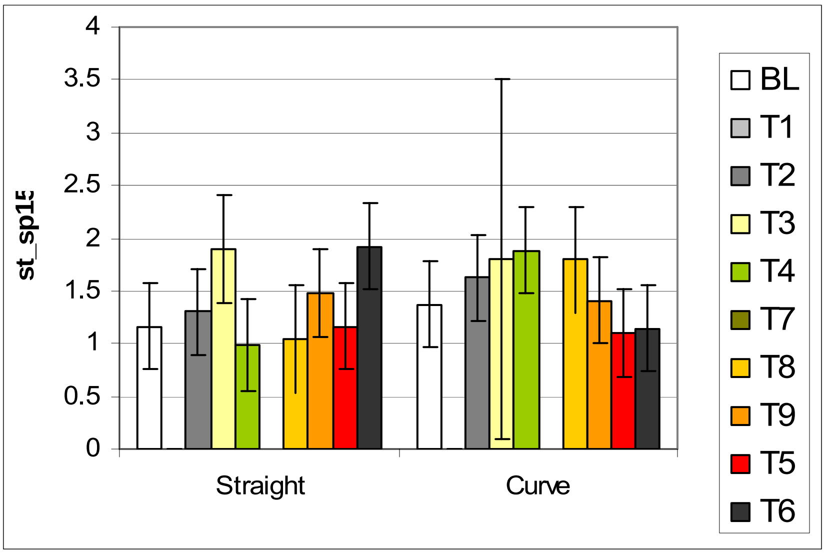


Figure 34 – Standard deviation of speed 15s window for the tasks performed with system A

baseline data that conformed more with previous work in that less speed variability was observed in the baseline condition. Neither of the main effects was significant, but there was a significant Task by Road Complexity interaction where Tasks 3 and 6 were greater than baseline but none of the measures for the curved segments differed from baseline.

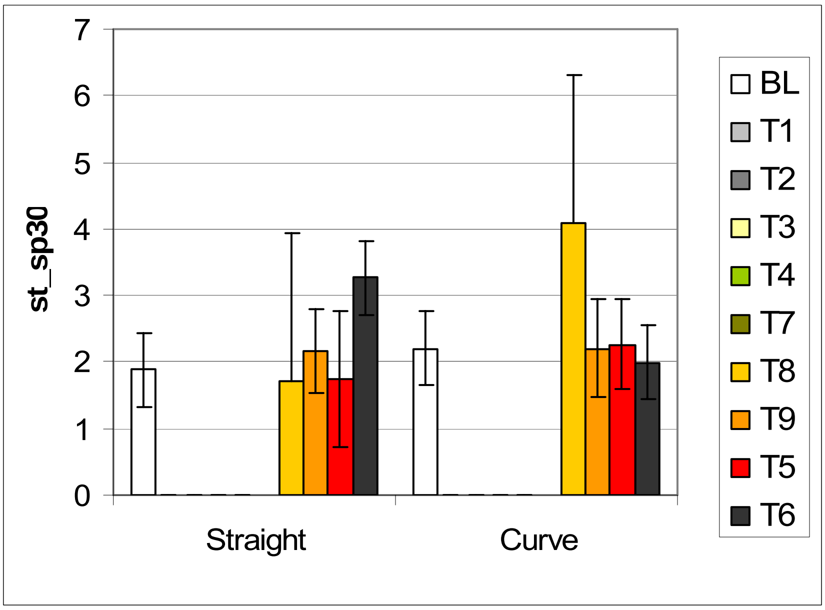


Figure 35 – Standard deviation of speed 30s window for the tasks performed with system A

Figure 35 displays the resulting data for Standard Deviation of Speed when a moving window procedure of 30s was used. In this analysis, only the baseline and Tasks 8, 9, 5 and 6 were of sufficient duration to be included in the analysis. Only the Task by Road Complexity interaction was significant. Further examination of the data indicated that only Task 6 on the straight segments differed from the baseline. None of the curved conditions differed from baseline.

The analysis of Minimum Speed data (see Figure 36*)* revealed a significant effect of Task. Neither the effects of Road Complexity nor the interaction were significant. Further examination of the Task effect revealed that the minimum speed for all tasks was significantly greater compared with the minimum speed for the baseline drive. This may be the result of the baseline calculation procedure.

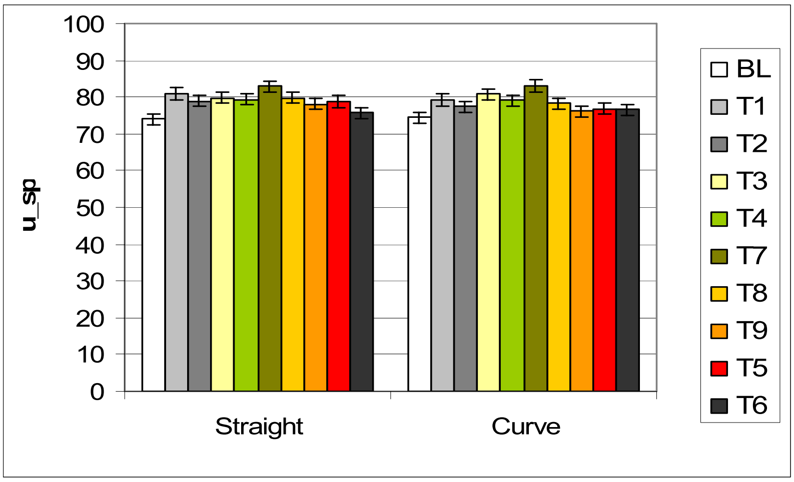


Figure 36 – Minimum speed for tasks performed with system A

There were no significant effects in the analysis for Speed Change, although the differences between Tasks 2 (p=.07) and 4 (p=.06) were marginally greater than baseline. There was a large amount of variability in the data.

The results for the Mean Headway Distance data (Figure 37) showed no significant effects but are included for comparison with the Standard Deviation of Headway analyses which follow. A large amount of variability was observed in the data. (The individual tests of means for Tasks 4, 6, and 9 were significant when compared to the baseline, indicating that when working on these tasks drivers maintained a greater headway distance.)

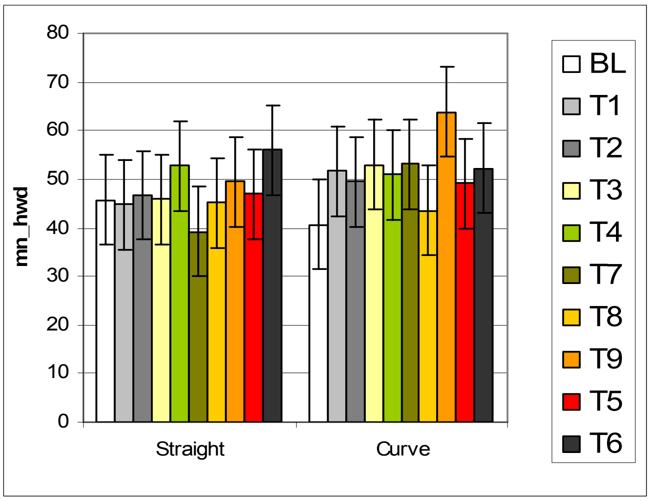


Figure 37 – Mean headway distance for tasks performed with system A

Only the effect for Task was significant in the ANOVA for Standard Deviation of Headway Distance. As can be seen in Figure 38, all task conditions showed less variability than the baseline condition which was unexpected and inconsistent with the comparisons of the tasks themselves which revealed increased variability with increased task difficulty. This is suspected to be an artefact of the way in which the baseline data were calculated. The "moving window" methodology was applied to this data in the two following analyses.

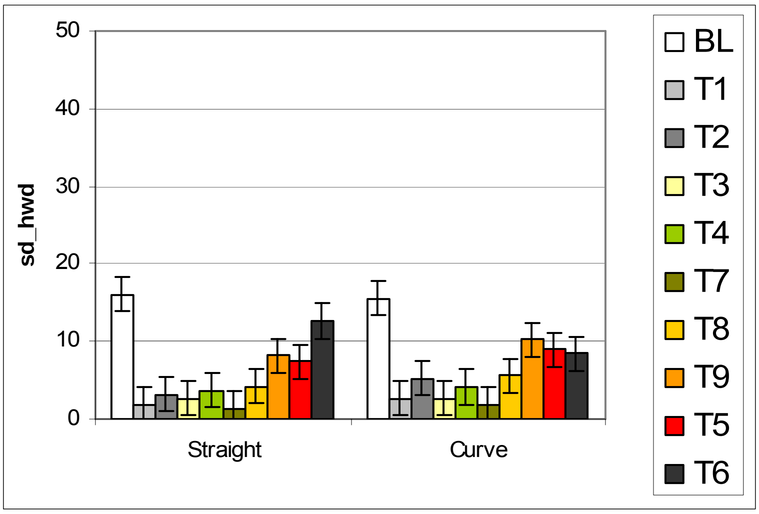


Figure 38 – Standard deviation headway distance for the tasks performed with system A

The Standard Deviation Headway Distance 15s Window data are presented in Figure 39. This procedure resulted in two tasks, Task 1 and 7, being eliminated from the analysis because they were less than 15s in duration. The moving window procedure resulted in less variable baseline as would be expected. None of the effects, however, were significant.

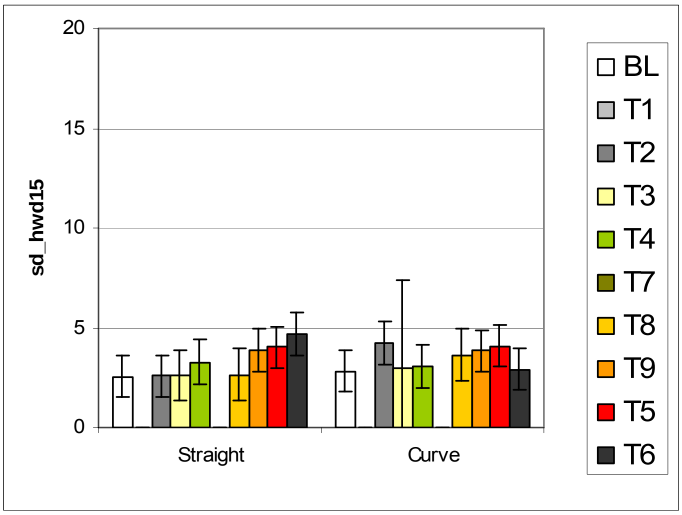


Figure 39 – Standard deviation headway distance 15s window for tasks performed with system A

A 30s window technique was applied to the Standard Deviation Headway Distance data (Figure 40) with the result that many of the tasks were eliminated due to their relatively short durations. It is also important to note that when the moving window method is used, the data from many subjects is also lost due to differences in task performance time. The ANOVA resulted in a significant interaction for Task by Road Complexity due to the significant effect for Task 6 on the straight segments and there was a marginally significant main effect for Task [p=.06].

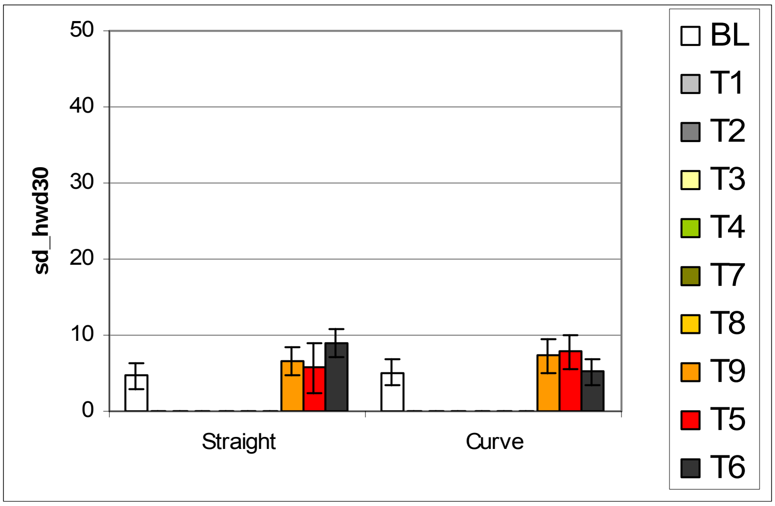


Figure 40 – Standard deviation headway distance 30s window for tasks performed with system A

The ANOVA for the Mean Headway Time analysis (see Figure 41*)* produced a marginal main effect for Task [p=.06]. Neither the main effect for Road Complexity nor the interaction approached significance. Further exploration of the Task differences from baseline indicated that for Tasks 4,6 and 9 drivers maintained longer Mean Headway Times (all p<.05).

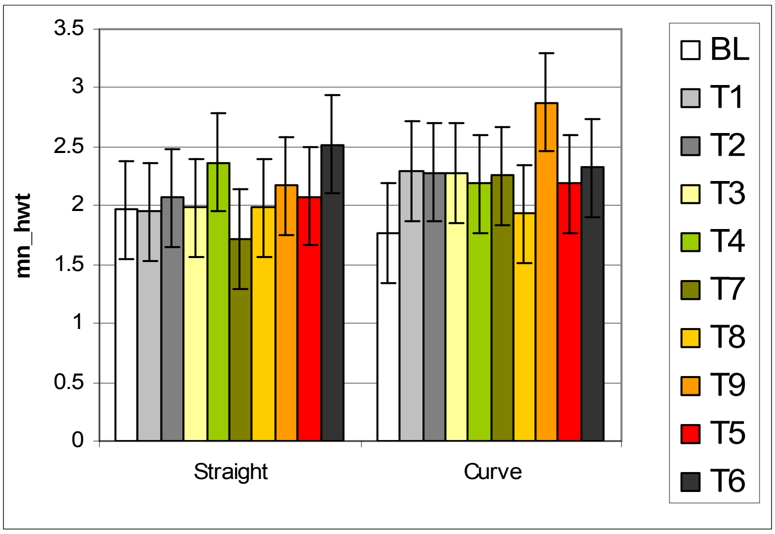


Figure 41 – Mean headway time for tasks performed with system A

The means for Standard Deviation Headway Time are presented in Figure 42**.** A significant main effect was found for Task. All task conditions differed significantly from the baseline but the difference was in the unexpected direction such that all Task conditions showed reduced variability compared to the baseline. Again, this is considered to be an artefact of the calculation of the baseline data and appropriate alternate methodologies are under consideration. There were a number of differences among the tasks themselves with a general pattern indicating greater variability in headway time with increased task difficulty.

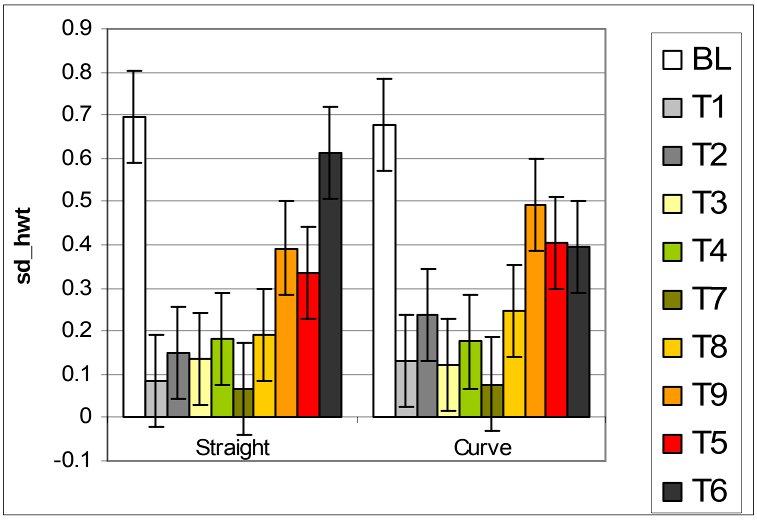


Figure 42 – Standard deviation headway time for tasks performed using system A

The results for the Standard Deviation Headway Time using the 15s Window technique are presented in Figure 43**.** The durations for both Tasks 1 and 7 were less than 15s, and consequently no data were available for these tasks when the 15s window technique was used with the Standard Deviation Headway Time data. The ANOVA resulted in a significant effect for Task, but no other significant effects. Tasks 5, 6 and 9 all showed significantly greater sd_hwt15 values than the baseline condition ($p<.05$).

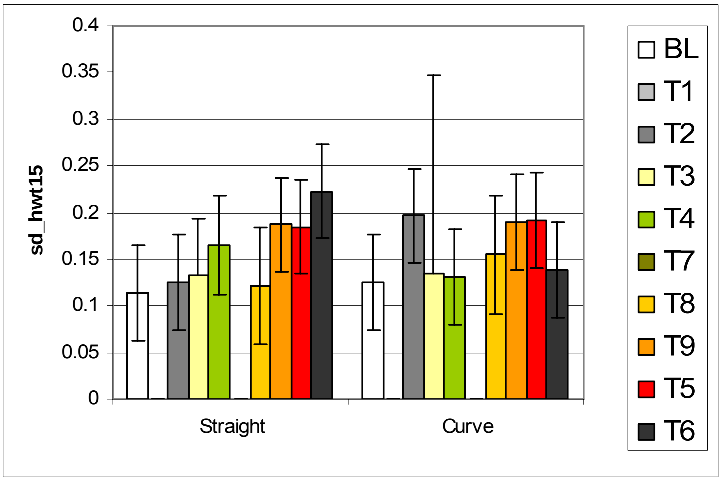


Figure 43 – Standard deviation headway time 15s window for tasks performed using system A

Figure 44 displays the Standard Deviation Headway Time for the 30s Window analysis. Due to the relatively short length (<30s) of many of the tasks, only data for baseline and Tasks 5, 6, and 9 were available for this analysis. This approach also resulted in differing numbers of subjects contributing to the cells for analysis. The ANOVA resulted in significant effects for Task and the Task by Road Complexity interaction.

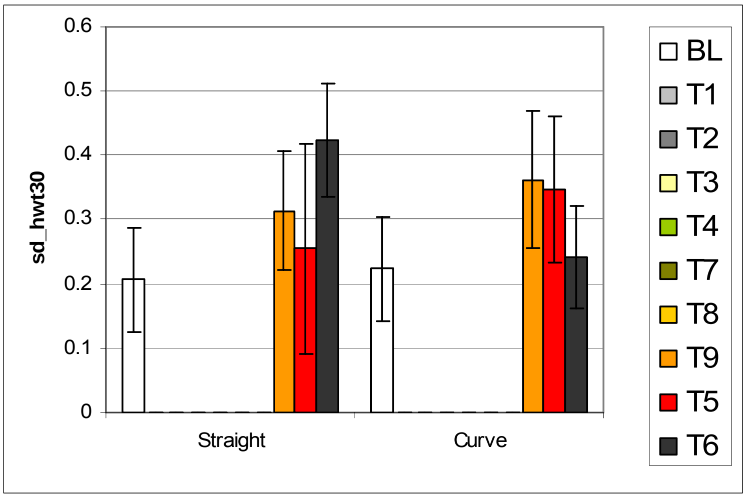


Figure 44 – Standard deviation headway time 30s window for tasks performed with system A

The means for Minimum Headway Time are presented in Figure 45. There was a significant main effect for Task in the ANOVA. The most striking finding is the very low minimum (under 1 s) associated with the baseline conditions compared to the task conditions where all task conditions have significantly longer minimum headway times (all p<.05) compared to the baseline condition.

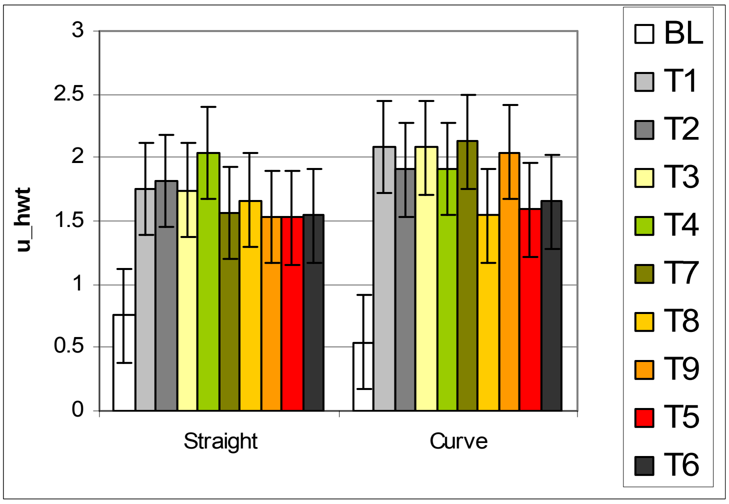


Figure 45 – Minimum headway time for tasks performed using system A

The data for Percent Time spent at various headways are presented in Figure 46**.** The ANOVAs conducted on the various bin sizes revealed effects for Road Complexity for only bin 0_1 and bin 2_3. For bin 0_1, a significantly greater percentage of time was spent in this headway range during driving on straights (17.08) compared with driving on curves (12.50). The analysis for bin 2_3 indicated that drivers spent a greater percentage of their time at this headway when driving on curves (26.73) compared to when they were driving on straight road segments (19.66). No other effects of Road Complexity were significant nor were any of the main effects of Task or interactions of the two variables.

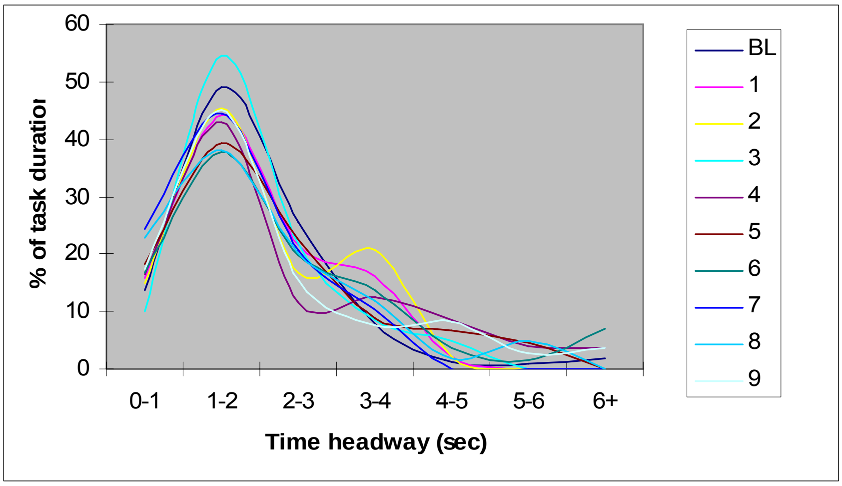


Figure 46 – Percent time headway: data combined for both straight & curved segments

The data for Percent Time Headway for Straight Segments and Curved segments are displayed in Figure 47 and Figure 48 respectively.

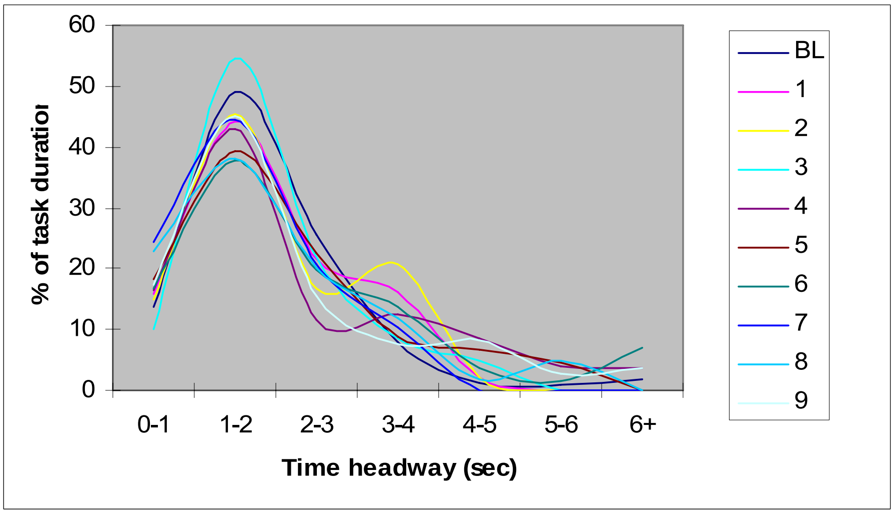


Figure 47 – Percent time headway: data for straight segments

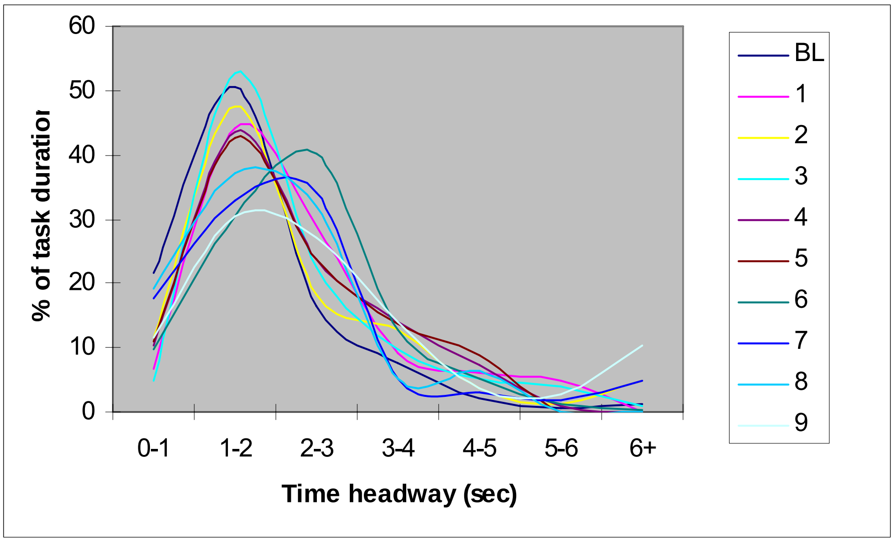


Figure 48 – Percent time headway: data for curved segments

5.7.1.3. Lateral control

There were no significant ANOVA effects for mean lane position.

The data for Standard Deviation of Lateral Position are presented in Figure 49. The ANOVA resulted in a significant interaction between Task and Road Complexity, as well as significant main effects for both Task and Road Complexity. In general, the variability increased with task difficulty and the effects were greater for the curved road segments. Again there is the finding of greater variability in the baseline data, which is inconsistent with previous findings and the data are subjected to windows analyses in the two following sections.

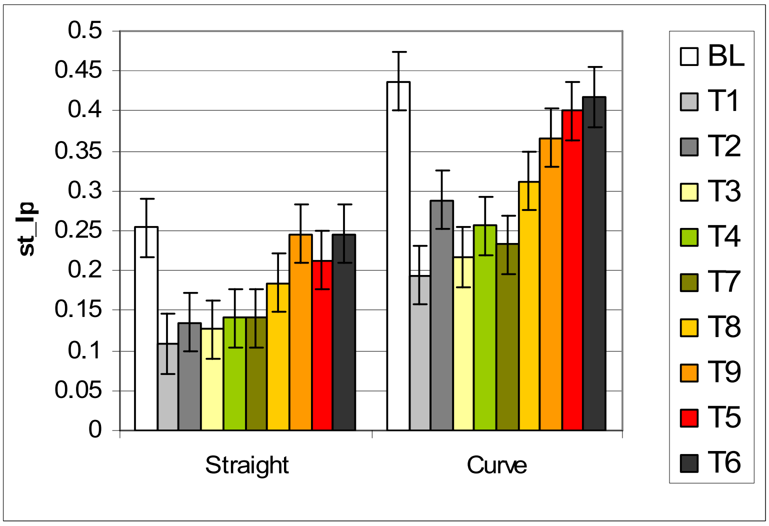


Figure 49 – Standard deviation of lane position for tasks performed with system A

The results of a moving window analysis (15s) for Standard Deviation of Lateral Position Data are displayed in Figure 50. Note that no data for Tasks1 and 7 are included in this analysis because these tasks were under 15s in duration. Both the main effect for Task and

Road Complexity were significant in the ANOVA. The interaction was not significant. Tasks 6 and 9 were significantly greater than the baseline ($p<.05$). Task 5 was marginally greater than baseline and Task 4 was marginally less than the baseline (both $p=.06$).

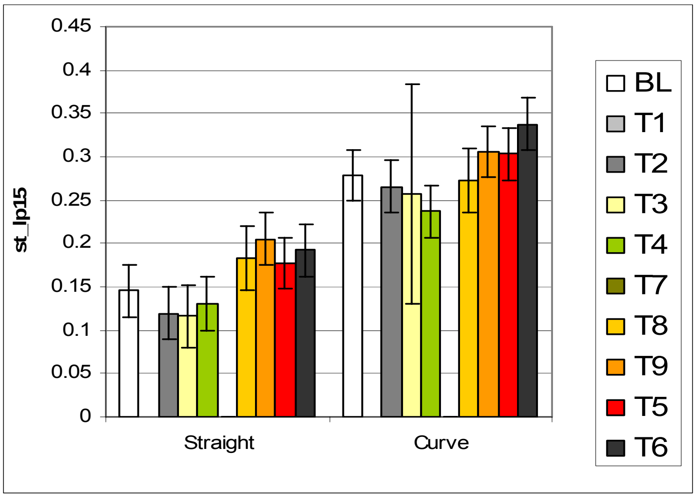


Figure 50 – Standard deviation of lateral position 15s window for tasks performed with system A

The results of the moving window analysis (30s) for the Standard Deviation of Lateral Position data are presented in Figure 51. Note that this procedure eliminated Tasks 1,2,3,4, and 7 from the analysis. In addition, a number of subjects were eliminated from the analysis. Significant main effects for Task and Road Complexity resulted from the ANOVA, but not the interaction. Tasks 5, 6 and 9 all resulted in greater lane variability than the baseline. More lane variability was observed during driving on the curves compared with driving on the straight segments.

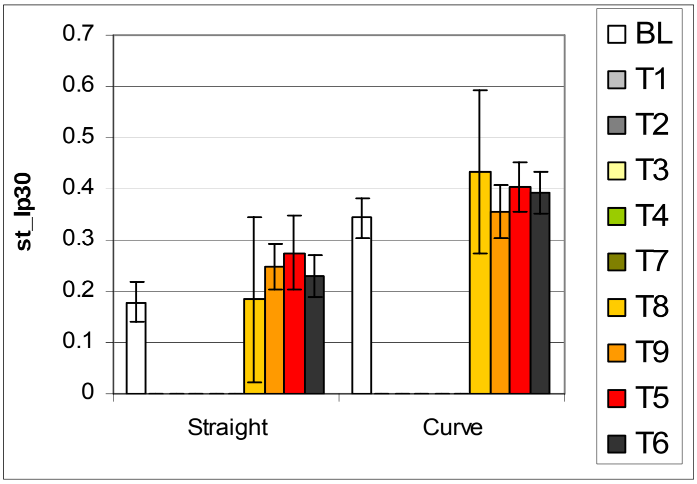


Figure 51 – Standard deviation of lane position 30s window for tasks performed with system A

Figure 52 displays the Steering Reversal Rate data (Amplitude of 1^{o}). The ANOVA resulted in a significant Task by Road Complexity interaction as well as significant main effects for Task and for Road Complexity. When the data for the Straight road segments are examined, Tasks 5, 6, and 9 all show values significantly greater than the baseline. When the data for the Curved road segments are examined, Tasks 1, 3 and 7 are all significantly less than the baseline value.

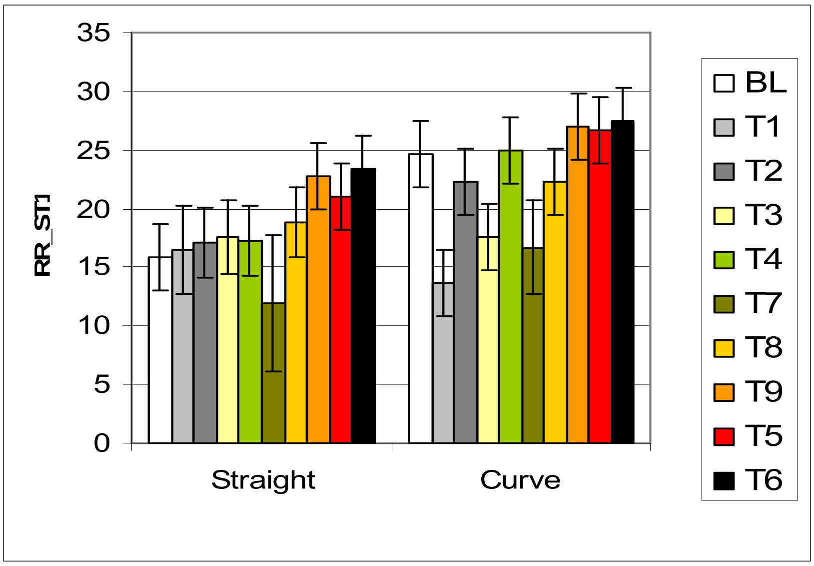


Figure 52 – Steering reversal rate: amplitude of 1° for tasks performed with system A

The Steering Reversal Rate data (Amplitude of 3°) are displayed in Figure 53. The ANOVA resulted in a significant Task by Road Complexity interaction as well as two significant main effects for Task and Road Complexity. When the data for the straight driving segments was examined, it was found that all tasks, except Tasks 3 and 4 had significantly greater values than baseline. For the curved driving segments, Tasks 5, 6, and 9 had significantly greater values than baseline.

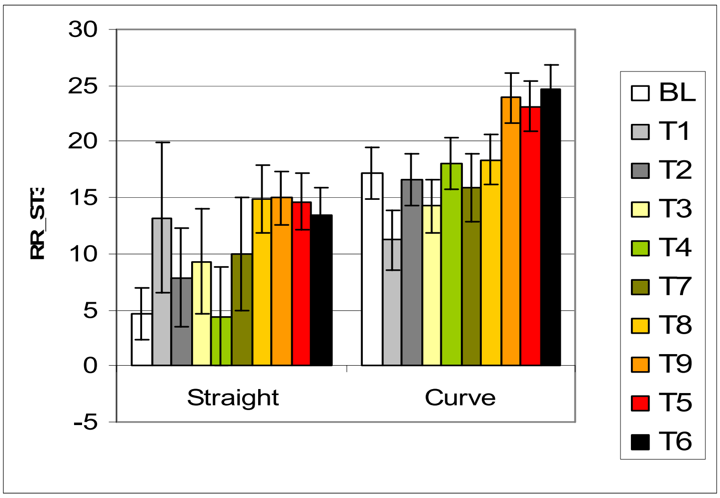


Figure 53 – Steering reversal rate: amplitude of 3° for tasks performed with system A

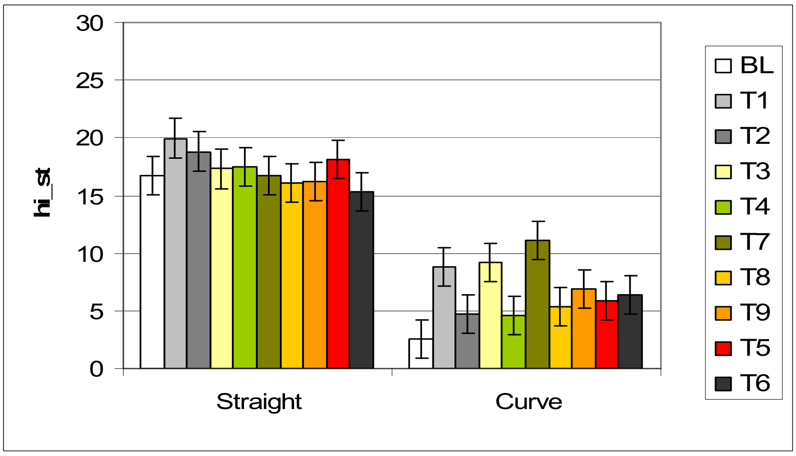


Figure 54 – High frequency component of steering wheel movements for tasks performed with system A

The High Frequency Component of Steering Wheel Movements data are displayed in Figure 54. There was a significant Task by Road Complexity interaction as well as two significant main effects for Task and Road Complexity. Exploration of the interaction revealed that for the straight segments of driving there were no significant effects of Task when compared with baseline. A different pattern of results was observed for the curved segment data, where all Tasks showed greater values for Hi-st, although Tasks 2 and 4 were only marginally greater (p=.07).

The Mean Time-to-line-crossing data are displayed in Figure 55. The ANOVA resulted in a significant Task by Road Complexity interaction as well as two significant main effects for Task and Road Complexity. The data for the straight driving segments revealed that time-to-line-crossing was significantly shorter for Tasks 5-9 (p=.06 for Tasks 6 and 9). All times to line crossing were smaller for the curved driving segments. In addition, time-to-line-crossing was significantly reduced compared to baseline for all tasks performed while driving on the curved segments (Tasks 2, 6 and 8 at p=.06, p=.07).

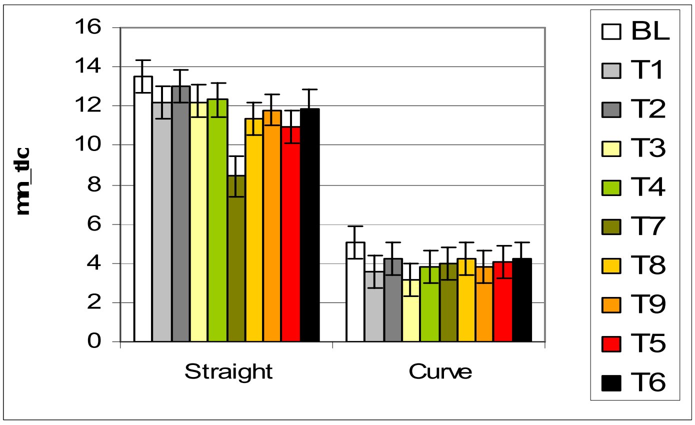


Figure 55 – Mean time-to-line-crossing for tasks performed with system A

5.7.1.4. PDT Performance

The Percentage Correct Detections or PDT Hit Rate is presented in Figure 56. The ANOVA resulted in significant main effects for both Task and Road Complexity. Drivers detected more PDT signals while they drove in the curved segments than the straight segments. Drivers showed comparable high performance in the baseline and Tasks 1 and 2, which were both auditory. When the Auditory Tasks became more complex (Tasks 3 and 4) and when there was a visual-manual requirement as in Tasks 5-9, the drivers did not perform so well.

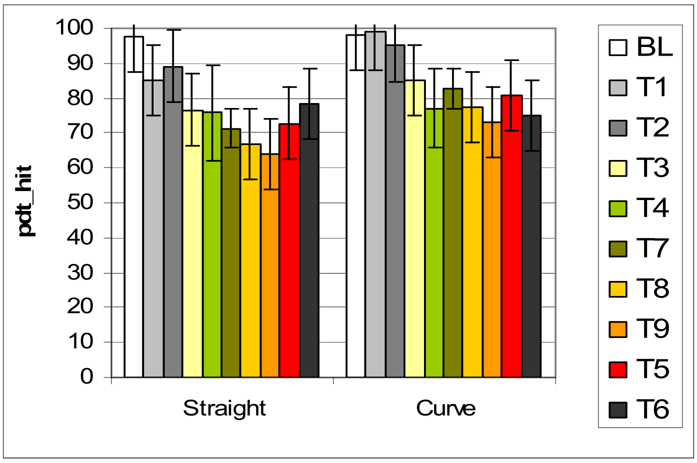


Figure 56 – Hit rate for PDT for the tasks performed with system A

The mean reaction times for correct PDT responses are displayed in Figure 57. The ANOVA for these data revealed a significant main effect for Task but no other significant effects. Responses were clearly fastest during the baseline drive and were significantly faster than every task condition. The auditory task RTs (for Tasks 1-4) did not differ from each other except that Task 2 RTs were marginally faster than Task 4 RTs. In general, the RTs for the auditory tasks were faster than those for the visual-manual tasks (Tasks 5-9).

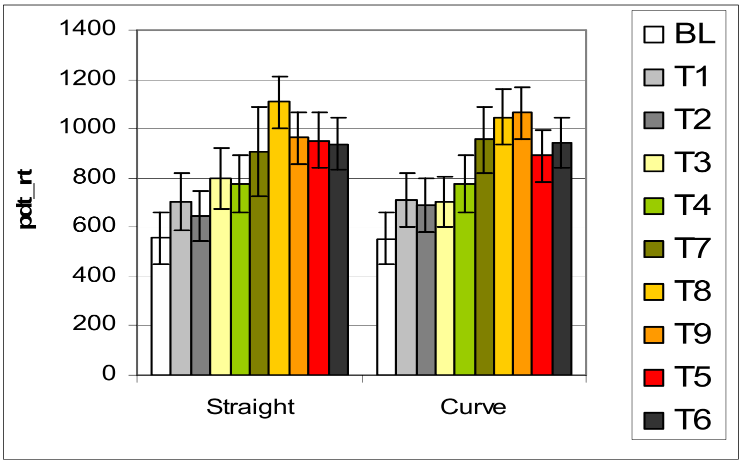


Figure 57 – PDT reaction times for tasks performed with system A

5.7.1.5. IVIS performance

The mean durations for Tasks are presented in Figure 58. There were considerable differences in Task Length across the various tasks. Although not indicated on the graph, the data collection for the baseline segments were based on the mean of 3 separate 60s samples during the drive. The ANOVA for Task Length produced a significant main effect for Task and a marginal main effect for Road Complexity (p=.07). The interaction of these two dependant variables was not significant.

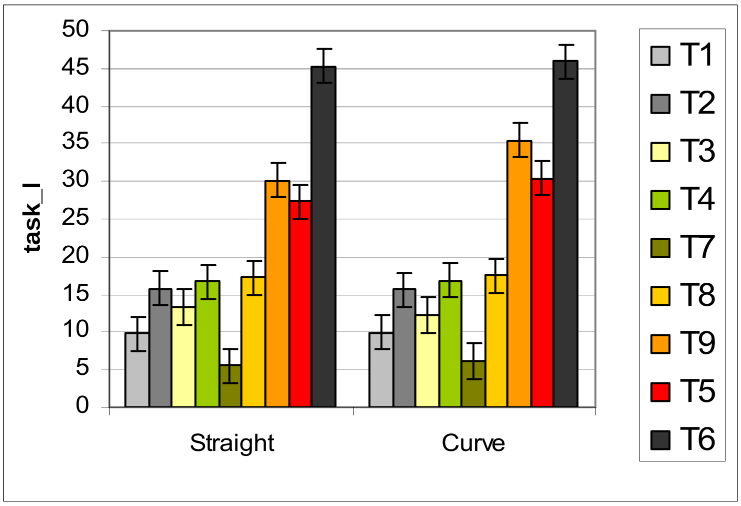


Figure 58 – Task length for the tasks performed with system A

Table 17 displays the numbers of subjects (out of a possible 20) who completed each task under the two Road Complexity Conditions.

Table 17 – The number of drivers who completed each task in each of the road complexity conditions

TASK →	1	2	3	4	5	6	7	8	9
Straight	20	17	7	20	20	20	20	20	20
Curve	20	18	8	20	20	20	20	20	20

5.7.2. Effects of System D

5.7.2.1. Drivers' reported Workload

Figure 59 presents the mean subjective workload ratings for drivers performing the tasks while using System D. The interaction and both main effects were significant. Means were higher when driving on curves.

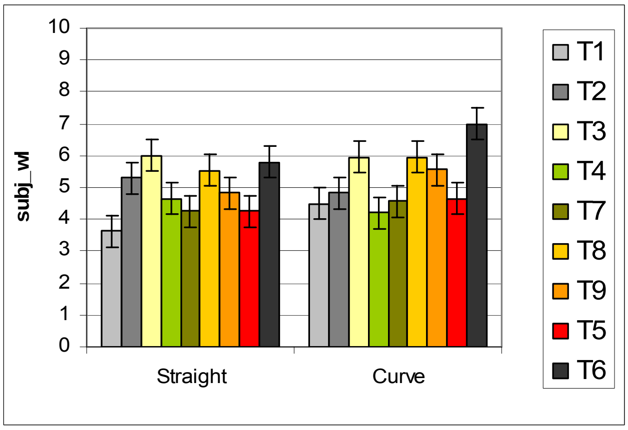


Figure 59 – Rated workload for the tasks performed with system D

1.7.2.2 Longitudinal Control

The mean speed data for the tasks and the two types of road complexity are presented in Figure 60. The ANOVA revealed significant effects for the interaction, task and road complexity effects. Mean speed was greater on straight segments. No differences were observed among tasks on the straight segments, but for the curved segments Tasks 5, 8 and 9 were all less than baseline.

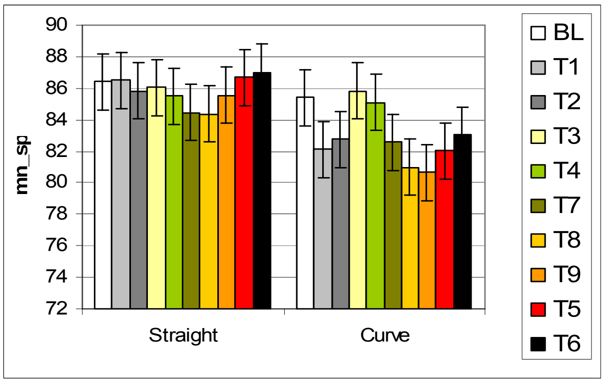


Figure 60 – Mean speed for the tasks performed with system D

Figure 61 presents the standard deviation of speed for the various tasks and the two levels of road complexity. There was a significant effect of Task in the ANOVA. However, tasks 1, 2, 4, 5, 7, 8 and 9 were all significantly *less* than the baseline, a result which calls the calculation of the baseline data into question.

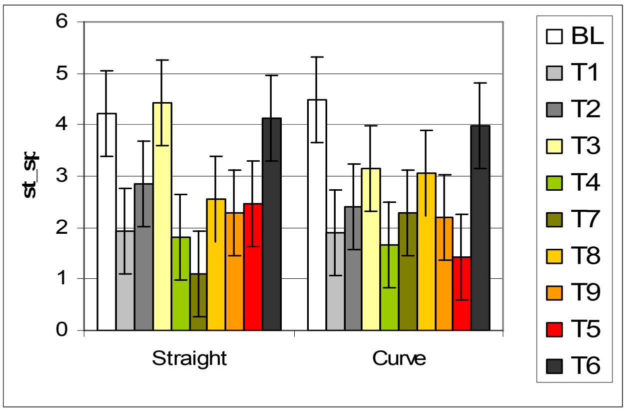


Figure 61 – Standard deviation of speed for the tasks performed with system D

There were no significant effects for standard deviation of speed with a 15s window.

The data for the analysis of the Standard Deviation of Speed using the 30s window technique are presented in Figure 62. Although significant ANOVA effects were found for Task, Road Complexity and their interaction, the number of subjects contributing to each mean varied widely from a low of 2 to a high of 20 due to the constraints of the procedure.

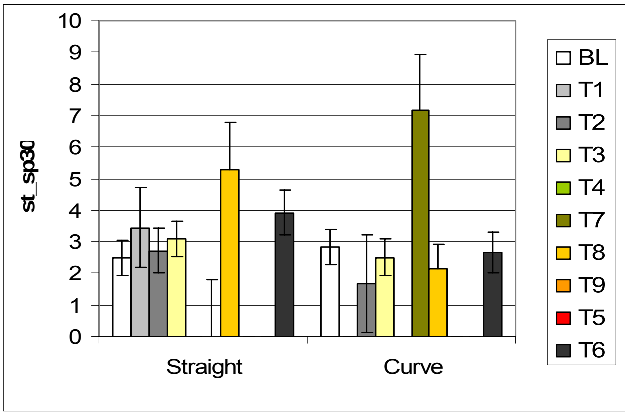


Figure 62 – Standard deviation of speed 30s window for system D

The Minimum Speed data are presented in Figure 63. The effects for both Task and Road Complexity were significant in the ANOVA. Minimum speed for all tasks was greater than baseline and the minimum speed was greater during straights than on curves.

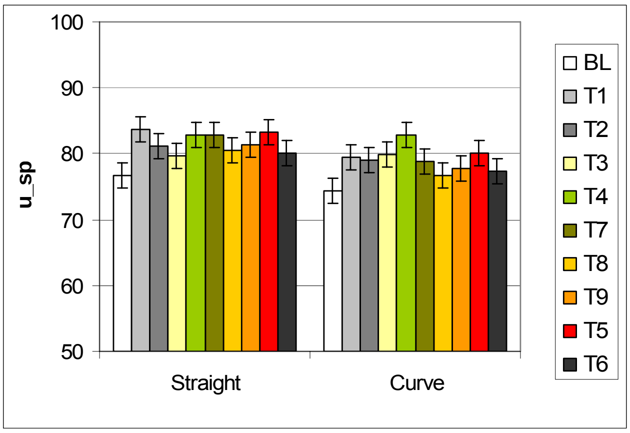


Figure 63 – Minimum speed for the tasks performed with system D

Speed Change (d_sp) was not calculated for System D. The previous analysis for System A indicated a large amount of variability in the data.

There were no significant effects for mean headway distance (mn_hwd).

Further exanimation of the significant Task main effect for Standard Deviation of Headway Distance (Figure 64) indicated that all tasks showed *less* of an effect than baseline. As found in several of the previous analyses, this is not the expected direction of the effect and is suspectd to be an artifact of the baseline data collection procedure.

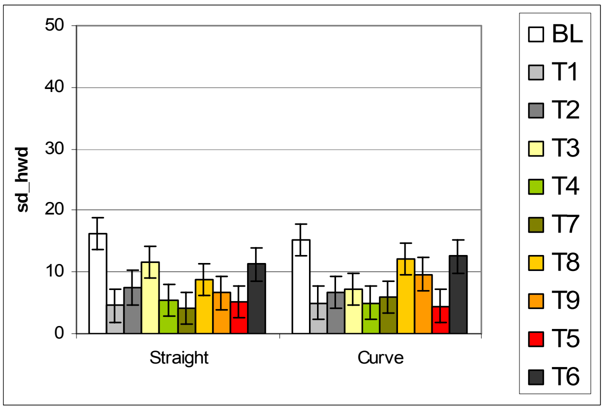


Figure 64 – Standard deviation headway distance for the tasks performed with system D

When the 15s Window technique was applied to the Standard Deviation of Headway Distance data (see Figure 65), the data took on the typical pattern. There was a significant main effect of Task and Tasks 5, 6, 8 and 9 all displayed greater values than baseline.

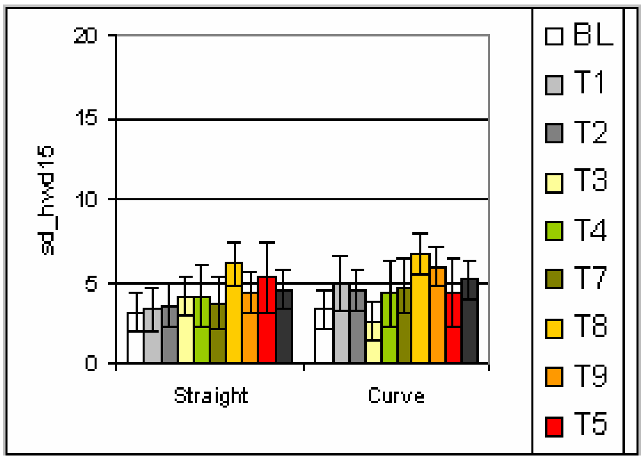


Figure 65 – Standard deviation headway distance 15s window for the tasks performed with system D

Figure 66 displays the Standard Deviation of Headway Distance data when the 30s window procedure is applied. Much of the data is lost as many of the tasks are less than 30s in duration. There is, however, a main effect of Tasks and tasks 6, 7 and 8 are all greater than baseline.

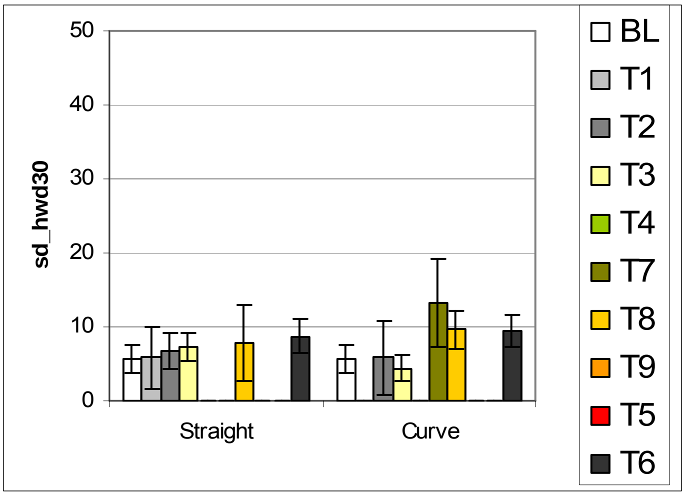


Figure 66 – Standard deviation headway distance 30s window for tasks performed with system D

None of the effects for Mean Headway Time were significant.

Standard Deviation of Headway Time showed a similar pattern to that for Headway distance. That is, all Tasks showed significantly smaller values than baseline.

As can be seen in Figure 67, the application of the 15s window bring the data in line with the expected pattern. There is a significant effect of Task where Tasks 5, 6, 8 and 9 all display greater variability than baseline.

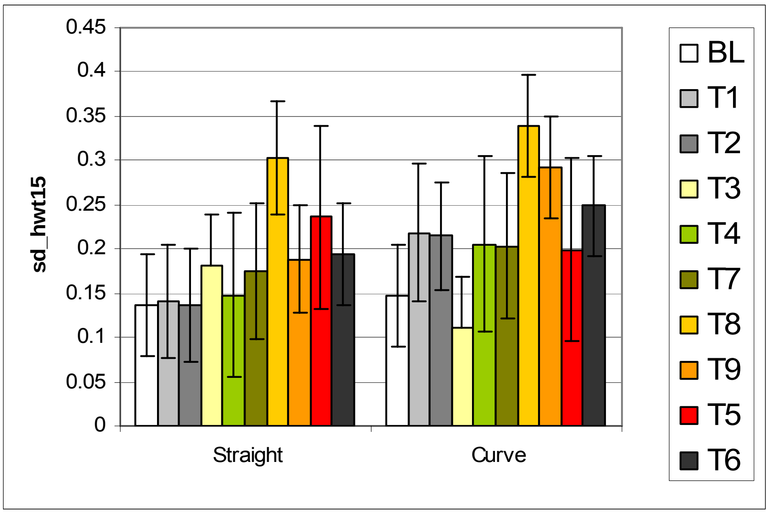


Figure 67 – Standard deviation of headway time 15s window for tasks performed with system D

Figure 68 displays the Standard Deviation of Headway Time data with a 30s window applied to the data. A number of tasks are lost from the analysis as their duration was too short. There was a significant effect of Task in that Tasks 6, 7 and 8 were all significantly greater than baseline.

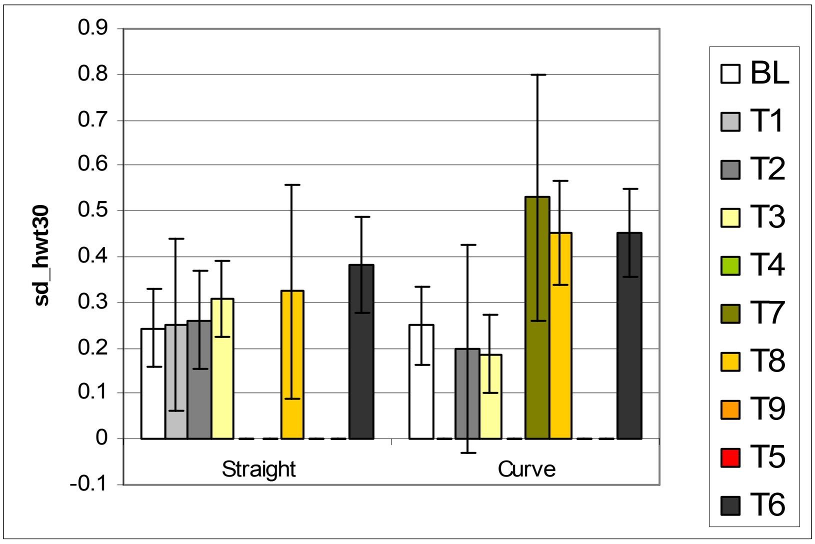


Figure 68 – Standard deviation of headway time 30s window for tasks performed with system D

The means for Minimum Time Headway are presented in Figure 69. The data show the same pattern as for System A in that the baseline condition displays the shortest value. Although there are some differences among the different tasks, their meaning is difficult to determine given the finding with the baseline.

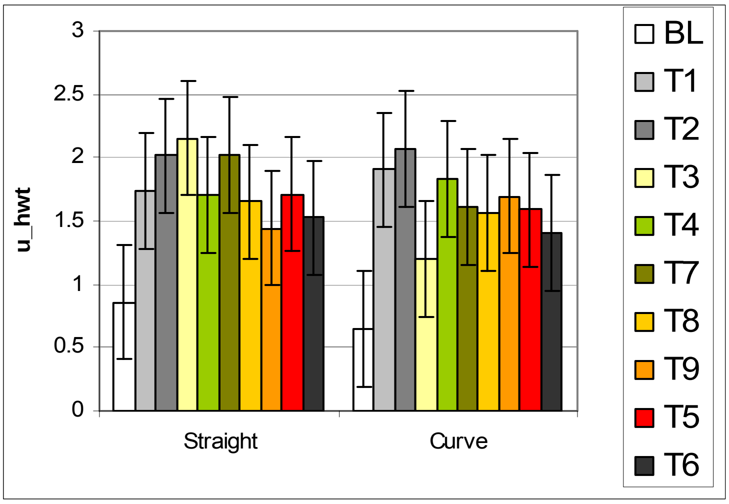


Figure 69 – Minimum headway time for tasks performed with system D

5.7.2.2. Lateral Control

Although there were significant effects of Task on Mean Lane position, it is difficult to interpret what they might mean. Tasks 1 and 7 show a reduced value of lane position, relative to the baseline. The Standard Deviation of Lateral Position is a more informative measure.

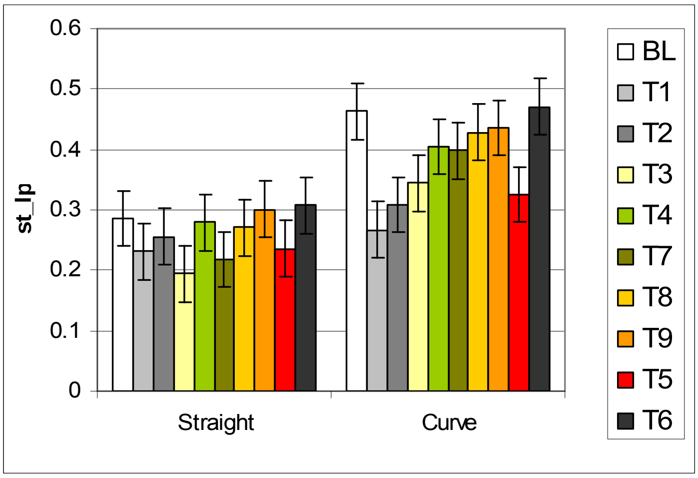


Figure 70 – Standard deviation of lane position for tasks performed with System D

The data for Standard Deviation of Lateral Position (Figure 70) suffer from the same problem as several of the other data sets in that the baseline data show greater variability than would be expected. This is likely due to the quantity of data sampled for that measure.

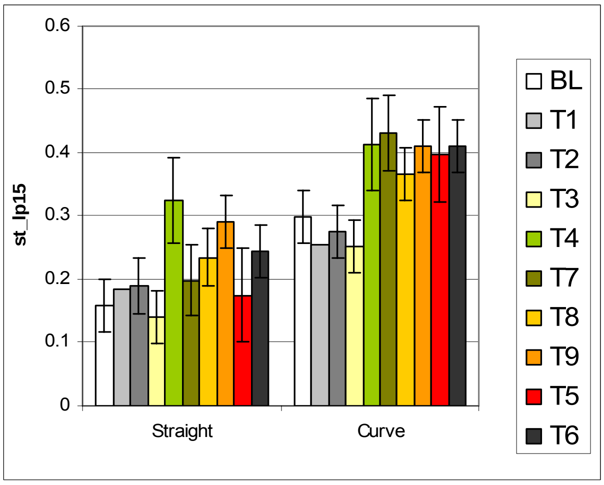


Figure 71 – Standard deviation of lane position 15s window for tasks performed using system D

The results for Standard Deviation of Lateral Position analysis using the 15s Window method are displayed in Figure 71. Both main effects are significant in the ANOVA. Tasks 4, 5, 6, 7, 8 and 9 all displayed greater values than the baseline measure. This measure also discriminated among the tasks themselves. Greater variability is observed on the curved portions of the route.

The results of the 30s moving window analysis for the Standard Deviation of Lateral Position data are presented in Figure 72. Several of the tasks are eliminated and some of the others suffer from loss of data. There is a significant interaction in the data where the more demanding tasks performed during curves produce the greatest effect. Both main effects were also significant. Most importantly, differences were observed among the tasks themselves on this measure. Tasks 2 and 3 were less than Tasks 6, 8 and 9. The data from the analysis suggest that this procedure works well when the samples are of sufficient duration.

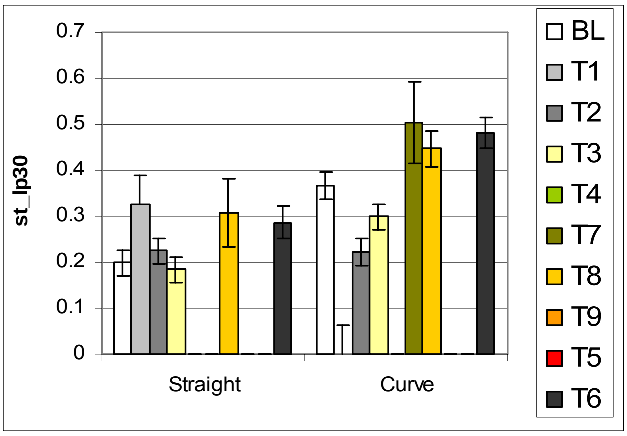


Figure 72 – Standard deviation of lane position 30s window for tasks performed using system D

Figure 73 displays the Steering Reversal Rate (Amplitude of 1) data. The ANOVA resulted in a significant Task and interaction effect. The effect for Road Complexity was marginal.

During the straight driving segments, reversal rates clearly increased when drivers performed any tasks compared to baseline. Interestingly, the two highest conditions during straight driving were for Task 1 and 2, which are simple visual reading tasks from the PDA. The results were less straightforward for the curved segments where the baseline rate was considerably higher.

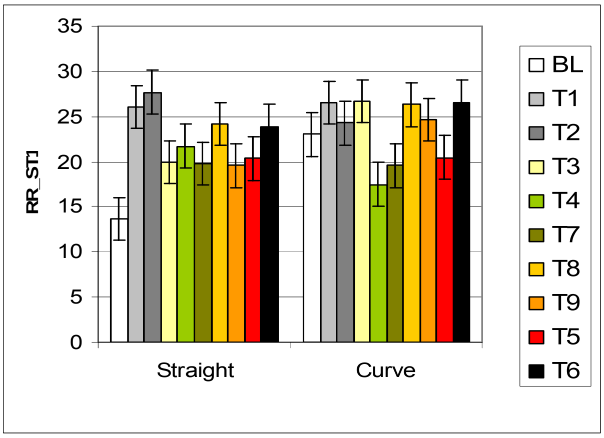


Figure 73 – Steering reversal rate: amplitude of 1° for tasks performed using system D

The Steering Reversal Rate (Amplitude of 3) data are presented in Figure 74. Both main effects and the interaction were significant in this analysis. The impact of performing tasks while driving was much more apparent for driving on the straight segments where the baseline value was very low.

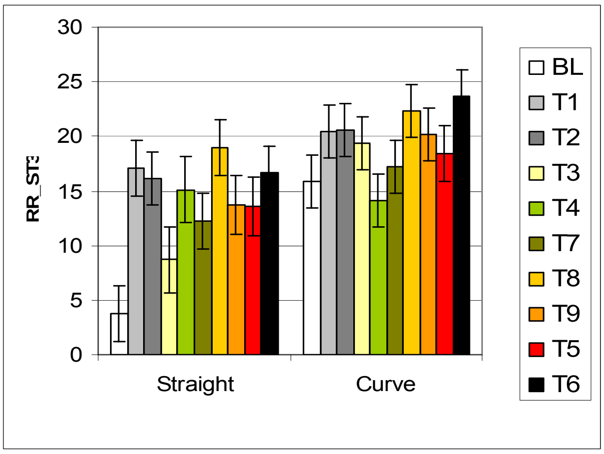


Figure 74 – Steering reversal rate: amplitude of 3° for tasks performed using system D

The High Frequency of Steering data are presented in Figure 75. Both main effects and the interaction were significant in the ANOVA. Values were clearly greater for driving on the straight segments. Differences among tasks were more apparent for the straight driving segments as well.

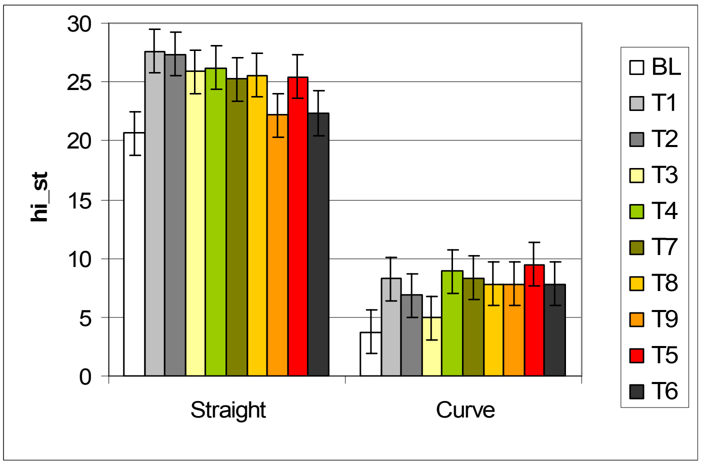


Figure 75 – High frequency component of steering for tasks performed using system D

Figure 76 displays the Mean Time-to-line-crossing results. Both main effects and the interaction were significant in the ANOVA. Clearly, times were shorter when drivers drove the curved segments. There was also greater variability among tasks during the straight section than during the curved driving sections.

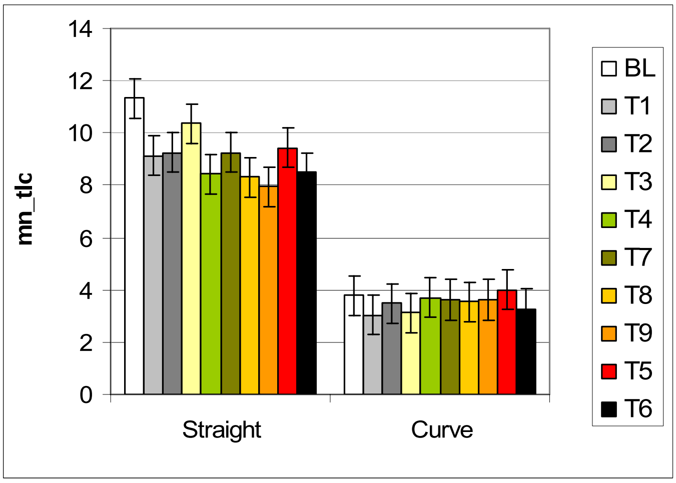


Figure 76 – Mean time-to-line-crossing for tasks performed using system D

5.7.2.3. PDT Performance

Figure 77 displays the PDT Hit Rate for the various tasks and the two road complexities. Performance while performing any task reduced the Hit Rate relative to baseline. Auditory-based Tasks (Task3) and easy Visual tasks (Tasks 1 and 2) showed moderate performance. Overall Tasks 4, 6, 7, 8 and 9 (ranked as the most difficult) showed the poorest performance. No other ANOVA effects were significant.

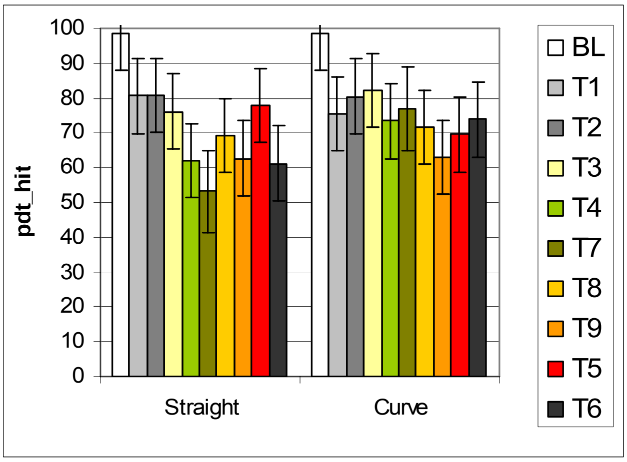


Figure 77 – Hit rate for PDT performance for tasks using system D

The mean Reaction Times for PDT responses are presented in Figure 78. There were significant effects of both Task and Road Complexity. Drivers responded significantly more quickly during baseline driving than when they were occupied with an in-vehicle task, by about 300 ms. The data followed a step-like pattern where the fastest RTs were observed during baseline, a mid-range of RTs during tasks that relied on Auditory (Task 3) or relatively easy Visual processing (Tasks1, 2). The slowest RTs were observed for the very demanding Visual/Manual Tasks. On average, drivers responded more quickly when driving on curves, although it is not clear why this would be the case.

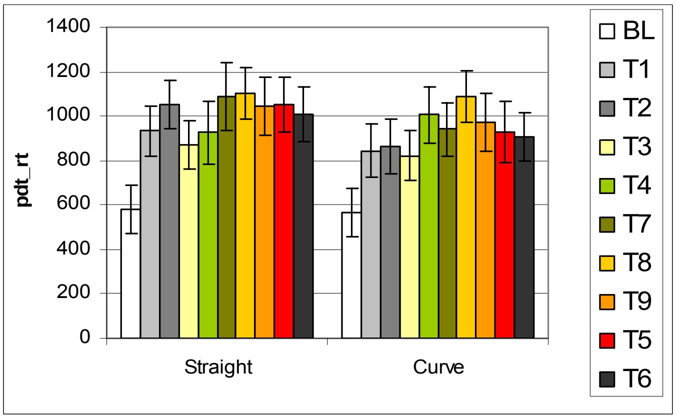


Figure 78 – PDT reaction time for tasks performed using system D

5.7.2.4. IVIS performance

There was considerable range in the length of the various tasks (14.40 to 51.57s; Figure 79).

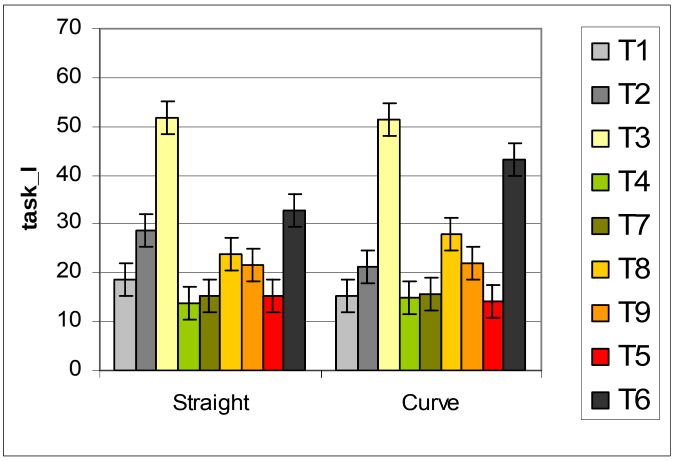


Figure 79 – Task length for tasks performed using system D

Table 18 displays the numbers of subjects (out of a possible 20) who completed each task using System D under the two Road Complexity Conditions. The number of drivers who completed task 3 was very low and this could be coupled to the length of task 3 (see graph above).

Table 18 – The number of drivers who completed each task using system D in each of the road complexity conditions

TASK→	1	2	3	4	5	6	7	8	9
Straight	20	17	7	20	20	20	20	20	20
Curve	20	18	8	20	20	19	20	20	20

5.7.3. Comparison between systems and groups of tasks

In general, subjective workload measures and task lengths were higher for System D. Although the tasks were not identical for both systems, they were designed to be similar. The interaction with System D was subjectively more demanding; the display was smaller and the use of the stylus was difficult for many participants. This may have contributed to the differences.

5.8. PDT as a measure of distraction

Transport Canada used the PDT task consistently in all conditions and both PDT detection and Reaction Time proved to be sensitive measures for Task Modality/Difficulty for both Systems A & D. Drivers performed the detection task quite well during baseline driving, as well as for auditory and visual manual tasks that were not very demanding. When the auditory task difficulty increased and during the more difficult visual-manual tasks, drivers' detection performance was much reduced. Interestingly, drivers detected more PDT targets while driving in curves. One possible explanation for this is that they tended to look out onto the scene more often while driving the more demanding curved sections.

Drivers' reaction times to the PDT displays also reflected the nature of the tasks they performed while driving. Reaction Time was clearly fastest when drivers were not occupied with any in-vehicle task. Reaction Times were in the mid-range during Auditory Tasks and the slowest reaction times were observed during the visual-manual tasks which required visual attention to the device and attention away from the scene outside the vehicle.

One of the appealing aspects of the PDT methodology is that it is sensitive to tasks with auditory (Harbluk & Lalande, 2005) and/or visual/manual (Olsson & Burns, 2000) demands. The present results confirm this. A similar pattern of performance was found for both Systems A and D.

5.9. Result summary and conclusions

A number of longitudinal control measures were included in these experiments. As found previously, Mean Speed was sensitive to task difficulty; drivers tended to slow down when they were involved with in-vehicle tasks. This effect was observed for both road complexity levels for System A, but only for the curved segments for System D. The Minimum Speed measure for both systems indicated that the minimum speed was greater for segments when drivers were using an in-vehicle system than when they were not. With System D, minimum speed was greater during straights than curves.

The longer segment of baseline data for the Standard Deviation of Speed resulted in an unlikely situation where drivers varied more during baseline than during task loading. Consequently, the standard deviation of speed measures (st_sp, st_sp15, st_sp30) were problematic. A moving window technique was applied to the data in an effort to remedy this situation. The 15s moving window resulted in data more in line with expectations, but the 30s moving window resulted in considerable loss of data due to tasks which were too short to be included in the analysis. It is possible that this technique would be more successful if the Tasks were pre-selected based on duration and then subjected to the analyses. More work is needed with the technique to determine when and how it should be applied. Other methods of determining the appropriate baseline measures should be explored.

The measures of Headway were also affected by the duration of the baseline segments. There were no significant effects for either system for Mean Headway Time and Mean Headway Distance. The measures for Standard Deviation of Headway Time and Distance both had baseline data that were more variable than the data for the experimental conditions. The moving window analyses were applied to both data sets. The window procedure was not successful for the Standard Deviation of Headway Distance data for System A. The 15s and 30s procedures resulted in values significantly greater than baseline for the most demanding tasks using System D.

Application of the moving window technique resulted in some of the conditions producing the expected pattern for Standard Deviation of Headway Time. The windows technique was more successful when applied to the data for the Standard Deviation of Headway Time for both Systems A and D where a greater variability in headway time was found for many of the more demanding tasks.

The data for Minimum Time Headway revealed very short minimum time headways for the baseline conditions, but considerably longer headways for all task conditions with both Systems. There was no discrimination by task difficulty.

In sum, the Headway measures were not consistently successful. Methodologies for exploring the treatment of the baseline data should be further explored. Comparisons with data from other sites may be of assistance in further understanding of these data.

A number of measures of Lane Position were calculated and analysed, but the window analyses for Standard Deviation of Lateral Position were the most meaningful. The window procedure was applied to these data sets because the basic analyses produced data where baseline performance showed high levels of variability relative to the task conditions. Again application of both the 15s and 30s windows techniques resulted in data being omitted from the analysis due to task durations.

The 15s window analysis for Standard Deviation of Lateral Position resulted in significant main effects for Task and Road Complexity for both Systems. Increased variability was associated with the curved road segments and the variability also reflected Task difficulty in general. The 30s window analysis reduced the number of tasks available for analysis. For system A, the most difficult tasks (5 and 6) showed significantly greater variability than baseline. There was more data available for System D analyses (due to the longer task durations in general) resulting in a pattern that was sensitive to task difficulty as well as task differences from baseline. This procedure looks promising as long as the tasks are sufficiently long.

The Steering Reversal data were inconsistent. For System A, the most difficult tasks (5, 6, and 9) resulted in increased rates for straight segments (but not in the curved segments) for the RR_ST1 analysis and increased rates for the same tasks in the curved but not the straight segments for the RR_ST3 analysis. For System D, the results were more clear-cut. For both measures of Steering Reversals, drivers showed greater values when engaged in a task than during the baseline driving. This held for the RR_ST1 straight condition and for many of the tasks in the RR_ST3 analyses. There did not appear to be any pattern in the data with respect to discrimination by task type.

Overall, the High Frequency Component of Steering Wheel Movements showed increased values when tasks were being performed with both systems compared to baseline. The exception was for System A during the straight segments when no differences were observed. There was some indication that “harder” tasks showed different values from “easier tasks”.

With system A, the Time-to-line-crossing results for the straight driving segments revealed that time-to-line-crossing was significantly shorter for the more demanding tasks (Tasks 5-9; p=.06 for Tasks 6 and 9). All time-to-line-crossing data were smaller for the curved driving segments. For system D, Time-to-line-crossing was reduced when a task was being performed on straight segments (except Task 3), but on curved segments only Tasks 1 and 3 showed a reduction in time. There did not appear to be an effect of task type.

Both PDT Hit Rate and Reaction Time measures proved to be sensitive to task modality and difficulty. This was supported by data from both systems. These measures are recommended for a testing protocol. The technique is simple to implement and appears robust.

5.10. Measures summary tables

5.10.1. System A

	Mean values										Effect
	Auditory tasks					Visual-Manual tasks					
measure	BL	T1	T2	T3	T4	T7	T8	T9	T5	T6	RLv (Int)
task_l	-	9.89	15.72	12.77	16.74	5.80	17.33	32.76	28.85	45.60	
subj_r	-	3.65	3.90	4.45	6.18	3.68	5.38	5.38	5.35	2.55	√
mn_sp	83.11	81.57	80.86	82.84	82.02	83.93	81.83	81.15	81.24	81.71	-
st_sp	3.70	.93	1.75	1.49	1.72	.56	1.87	2.55	2.17	3.44	(int)
st_sp15	1.27	-	1.46	1.63	1.41	-	1.48	1.44	1.13	1.53	(int)
st_sp30	2.04	-	-	-	-		2.57	2.41	1.98	2.63	(int)
u_sp	74.21	80.10	78.25	80.29	79.26	83.01	79.01	77.08	77.81	76.07	-
d_sp	-.17	-7.42	-8.55	-5.14	-8.84	-3.02	-4.03	-1.8	-.78	.62	-
mn_hwd	43.11	48.17	48.07	49.36	51.82	46.22	44.37	56.71	48.00	54.15	-
sd_hwd	15.83	2.27	4.15	2.63	3.91	1.61	4.92	9.19	8.22	10.56	-
sd_hwd15	2.69	-	3.42	2.48	3.02	-	3.51	3.86	4.04	3.79	-
sd_hwd30	4.87	-	-	-	-	-	7.80	7.39	6.85	6.87	(int)
mn_hwt	1.87	2.12	2.18	2.13	2.28	1.98	1.96	2.52	2.13	2.42	-
sd_hwt	.69	.11	.19	.13	.18	.07	.22	.44	.37	.50	-
sd_hwt15	.12	-	.16	.13	.14	-	.16	.19	.19	.18	-
sd_hwt30	.22	-	-	-	-	-	.31	.36	.30	.32	(int)
u_hwt	.65	1.92	1.86	1.91	1.97	1.85	1.60	1.79	1.56	1.60	-
mn_lp	.41	.38	.44	.43	.36	.40	.45	.36	.38	.39	-
st_lp	.35	.15	.21	.17	.20	.19	.25	.31	.31	.33	√ (int)
st_lp15	.21	-	.19	.17	.18	-	.23	.26	.24	.26	√
st_lp30	.26	-	-	-	-	-	.30	.30	.33	.31	√
rr_st1	20.25	15.87	19.96	17.47	21.24	15.58	20.83	24.89	23.87	25.40	√ (int)
rr_st3	10.91	12.22	12.25	11.75	11.21	12.94	17.41	19.51	18.89	19.06	√ (int)
hi_st	9.65	14.37	11.76	13.25	11.07	13.91	10.70	11.53	11.99	10.81	√ (int)
mn_tlc	9.29	7.81	8.61	7.68	8.05	6.21	7.80	7.81	7.51	8.03	√ (int)
pdt_hit	97.83	91.18	92.93	80.83	76.46	74.12	72.13	68.49	76.79	76.63	√
pdt_rt	555	713	676	727	774	959	1063	1013	919	940	-

(significant difference from BL indicated by grey background)

5.10.2. System D

	Mean values										Effect
		Visual-Manual tasks									
measure	BL	T1	T2	T3	T7	T9	T4	T8	T5	T6	RLv (Int)
task_l	-	16.95	25.00	51.57	15.48	21.87	14.40	25.81	14.78	37.91	
subj_r	-	4.06	5.06	5.98	4.41	5.19	4.43	5.74	4.45	6.40	√ (int)
mn_sp	85.92	84.32	84.28	85.93	83.51	83.11	85.30	82.66	84.36	85.03	√ (int)
st_sp	4.35	1.93	2.64	3.78	1.70	2.23	1.75	2.81	1.95	4.04	-
st_sp15	1.71	1.46	1.58	1.67	1.23	1.42	1.76	1.48	1.93	1.74	-
st_sp30	2.66	3.45	2.52	2.80	7.15	-	-	3.18	2.59	3.37	√ (int)
u_sp	75.49	81.58	80.12	79.75	80.86	79.55	82.76	78.55	81.69	78.72	√
d_sp	-	-	-	-	-	-	-	-	-	-	-
mn_hwd	45.05	51.73	60.92	56.02	51.44	50.01	51.45	55.61	47.29	55.87	-
sd_hwd	15.69	4.76	7.21	9.43	5.08	8.12	5.19	10.47	4.82	11.86	-
sd_hwd15	3.21	3.90	4.04	3.37	3.65	5.12	4.51	6.12	4.67	4.81	-
sd_hwd30	5.67	5.83	6.76	5.88	13.20	-	-	9.26	3.40	8.99	-
mn_hwt	1.90	2.22	2.63	2.40	2.23	2.19	2.18	2.47	2.03	2.38	-
sd_hwt	0.68	0.20	0.31	0.39	0.24	0.38	0.23	0.51	0.22	0.54	-
sd_hwt15	0.14	0.17	0.18	0.15	0.16	0.24	0.19	0.30	0.21	0.22	-
sd_hwt30	0.25	0.25	0.27	0.25	0.53	-	-	0.40	0.16	0.42	-
u_hwt	0.75	1.82	2.04	1.68	1.81	1.57	1.77	1.61	1.65	1.47	-
mn_lp	0.43	0.30	0.44	0.49	0.22	0.41	0.34	0.39	0.37	0.37	(int)
st_lp	0.37	0.25	0.28	0.27	0.31	0.37	0.34	0.35	0.28	0.39	√ (int)
st_lp15	0.23	0.22	0.23	0.20	0.33	0.35	0.35	0.30	0.29	0.33	√
st_lp30	0.28	0.33	0.26	0.24	0.51	-	-	0.38	0.25	0.37	√ (int)
rr_st1	18.34	26.32	25.98	23.29	19.68	22.11	19.61	25.23	20.40	25.26	(int)
rr_st3	9.97	18.83	18.35	13.76	15.04	17.02	15.50	20.73	16.31	20.16	√ (int)
hi_st	12.21	17.93	17.12	15.41	16.77	15.01	17.53	16.69	17.48	15.08	√ (int)
mn_tlc	7.56	6.09	6.36	6.75	6.43	5.79	6.06	5.94	6.71	5.88	√ (int)
pdt_hit	98.66	78.04	80.64	79.21	65.13	62.81	67.71	70.39	73.71	67.54	-
pdt_rt	574	896	964	846	994	997	916	1091	962	962	√

6. The VTI simulator experiment

6.1. Test site

The VTI driving simulator III was used in this study (see Figure 80). The simulator is a newly built high-fidelity, moving base dynamic driving simulator, based on the concepts of the two previous VTI moving base driving simulators. Several validation studies have been carried out with these previous simulators (Haakamies-Blomqvist, et. al 2000; Harms, 1996).

Figure 80 – The VTI driving simulator (left) and control panel (right)

6.2. Scenarios and participants

The HASTE standard rural road was implemented in this study and 20 average participants (10 females and 10 males) were included. The average age of the participants was 35 and ranged from 23 to 49 years. The average annual mileage was 18,700 km and ranged from 2,000 to 50,000 km.

6.3. IVIS included

Systems A and B were included and installed as prescribed in the operation manuals. All nine specified tasks of each system were used. For system B, Tasks 1 and 2 included a manual start, which was not supposed to be included for analysis. However, due to an error, the manual interaction was included, which lasted approximately five seconds. No corrected actions were taken with regard to this error and therefore the results for these data need to be interpreted somewhat different compared to the same tasks in other experiments.

6.4. Experimental design

All participants used both IVIS. Each participant also drove an additional run with one of the IVIS, whilst also using the Peripheral Detection Task (PDT). Each participant thus completed three experimental runs. The PDT was used by half of the participants for each IVIS. For each IVIS, the PDT was always used after driving without PDT in order to avoid the PDT affecting the driving behaviour data that was to be used for comparing between test sites. Also, VTI considered that there was a risk of information overload if the PDT and IVIS were introduced simultaneously to the participants. With this design, the participants were already familiar with the IVIS when the PDT was introduced. The VTI experimental data could not be used for investigating the effect of PDT on driving performance and workload due to the unbalanced order. The order of IVIS and IVIS tasks were counterbalanced across participants.

6.5. Procedure

Participants received written and spoken instructions, and practiced the tasks of the first IVIS to be used. Participants then practiced driving on the rural road for five minutes to familiarise themselves with the simulator. Then, still during driving, all participants rehearsed the IVIS tasks. Once all experimental components had been practiced, participants drove the experimental run with the first IVIS. If the PDT was to be used with the first IVIS, the PDT was practised and a second experimental session was run which included the PDT. The road environment and the participant's face were recorded on DVD. After driving, participants signed a document approving/not approving VTI to use the video recordings for scientific purposes.

During driving, the experimental leader instructed participants about the particular IVIS task, and its start time. Participants were instructed to say "klar" (*eng* "finished") when they had finished the task. If the task was not finished by the end of the 90 seconds window for task completion, participants were instructed to abandon the task. If participants had difficulties in completing the IVIS, the experimental leader provided guidance. Data from such situations was included in the analysis, as long as participants were considered to be attending to the task.

6.6. Measures and analysis method

Speed, lateral position and steering angle were measured. Several safety critical indicators were then were derived from these, such as lateral position variation, time-to-line-crossing and reversal rate. All measures were implemented according to the specifications described in this document (Appendix 2). The effects of the IVIS tasks and road complexity were analysed according to the common analysis specification. A note of the participants' attendance to the IVIS tasks was also made.

6.7. Results

There were very few PDT cheats, false and missed responses, and therefore these measures were not tested for significant effects. Since PDT hit rate and reaction time were based on the (high) proportion of correct responses, these measures were analysed. There were too few data points for proportion of TLC minima less than one second, and this data was therefore excluded from the analysis.

6.7.1. Effects of System A

6.7.1.1. Task length

The task lengths varied between 10 seconds (T7) and 61 seconds (T6), where the visual/manual tasks in general took much longer than the auditory tasks (Figure 81).

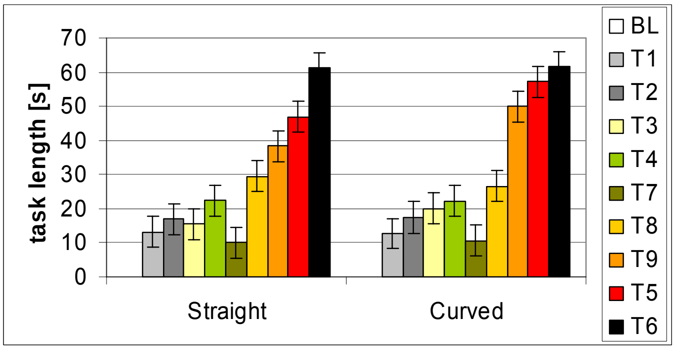


Figure 81 – Average task length for System A

6.7.1.2. Self-reported driving performance

Drivers judged their driving performance to be reduced significantly by all tasks. The effects were however larger for the visual/manual tasks (average 2.5 decrease from baseline) than for the auditory tasks (average 0.5 decrease from baseline). There was a clear resemblance between inverted task lengths and subjective ratings, indicating that the task length strongly influenced the self-reported driving performance. Of course, it could also be that the longer tasks were also more complex (Figure 82).

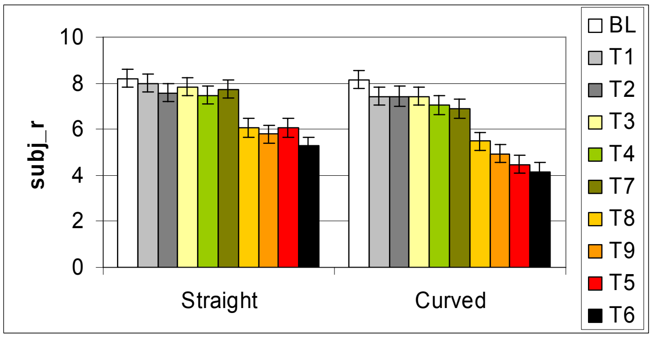


Figure 82 – Self-reported driving performance

6.7.1.3. Longitudinal control

Mean and minimum speed (mn_sp and u_sp) decreased significantly when attending to system A. The overall mean headway increased significantly, but there was a different effect between the cognitive and the visual/manual tasks, with the cognitive tasks resulting in *decreased* headway. Such indications were also found in HASTE WP2, and might be explained by the driver being distracted in the car following task. However, since average headway was still very large, there is no safety impact of this result. The largest effects in speed and headway were found for the tasks that took the longest time (T6) (see Figure 83). The visual/manual tasks caused a larger decrease in speed than the auditory tasks. This effect was found in mean speed, but not in speed change (d_sp). This may be because the speed profile did not show a linear change – which is required for d_sp to reflect speed change. An example of such a profile is a fast speed drop followed by a slow speed increase.

The reduction in speed may be explained by the drivers compensating for the increased visual and cognitive demands.

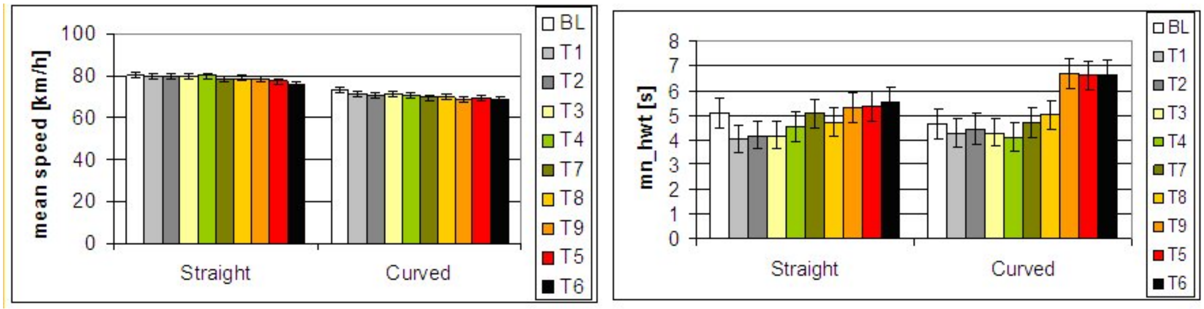


Figure 83 – Mean speed (left) and mean time headway (right)

The standardised time headway variation measures (sd_hwt15, sd_hwt30, sd_hwd15, sd_hwd30) indicated that the longer tasks resulted in higher variations, caused by these tasks being affected by the car following performance (see Figure 84).

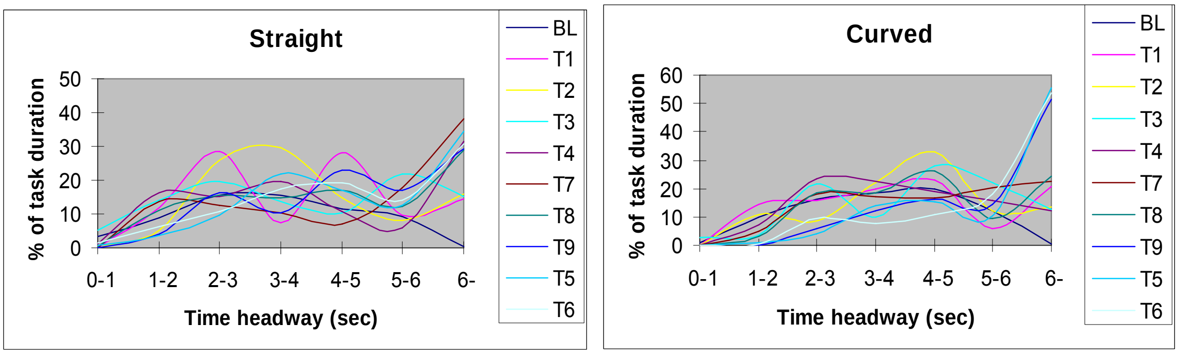


Figure 84 – Time headway histograms

Effects were also found in the non-standardised time and distance headway variations, although these measures are highly sensitive to task length, as can be seen in Figure 85. In general, the effects on longitudinal control were larger in curves than in straight road sections.

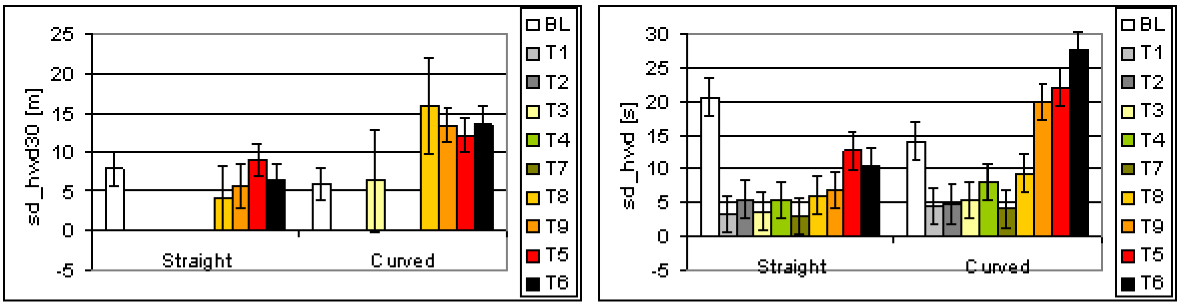


Figure 85 – Standardised distance headway variation (left) and conventional distance headway variation (right)

6.7.1.4. Lateral control

The lateral position was clearly shifted towards the right line as the IVIS task increased in difficulty and length (see Figure 86). This can be interpreted as an increase in distance to the road edge and thus an increase in the swerving margin. An increase in lateral position variation and decrease in TLC was also seen (see Figure 87).

The steering activity increased significantly for most tasks, and in particular those tasks taking a long time and requiring complicated visual/manual interactions (see Figure 87). This effect was found in reversal rate (one degree) and high frequency steering. In spite of the increased steering activity, lateral position variation increased in eight of nine tasks and the time-to-line-crossing decreased in five of nine tasks. This indicates that the increased steering effort could not compensate for the visual/manual distraction. Of course there was an effect of road curvature in lateral control, resulting in higher lateral position variation and lower TLC values.

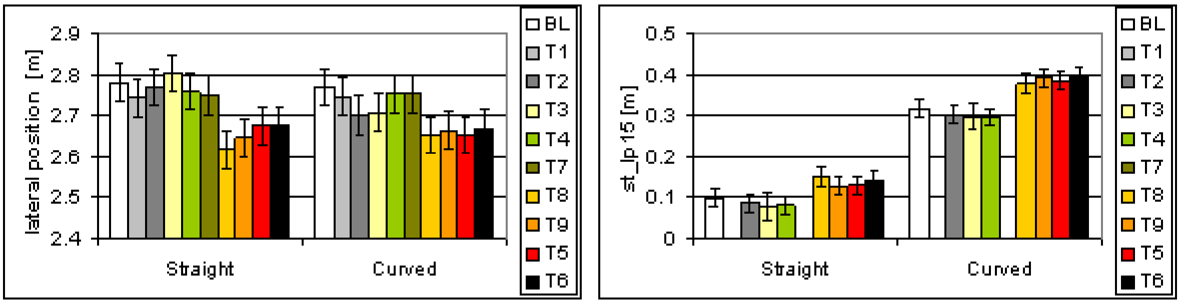


Figure 86 – Lateral position (left) and lateral position variation (right)

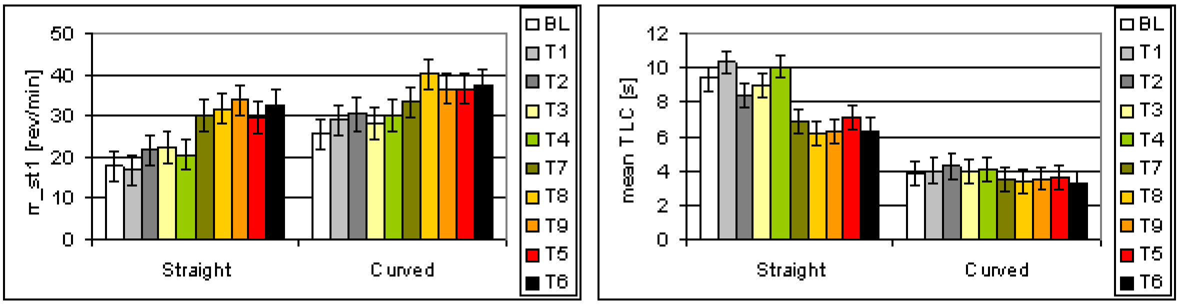


Figure 87 – one-degree reversal rate (left) and mean time-to-line crossing (right)

6.7.1.5. PDT results

The PDT reaction time was significantly increased for seven of the nine tasks. Also, hit rate was reduced, but this measure was less sensitive than reaction time (see Figure 88). The results show that event detection was most affected by the more complex visual/manual tasks. For T6 (destination entry), the reaction time was doubled compared to the baseline, which would suggest that this task would for instance have a severe impact on stopping distance in a critical event requiring immediate brake reaction.

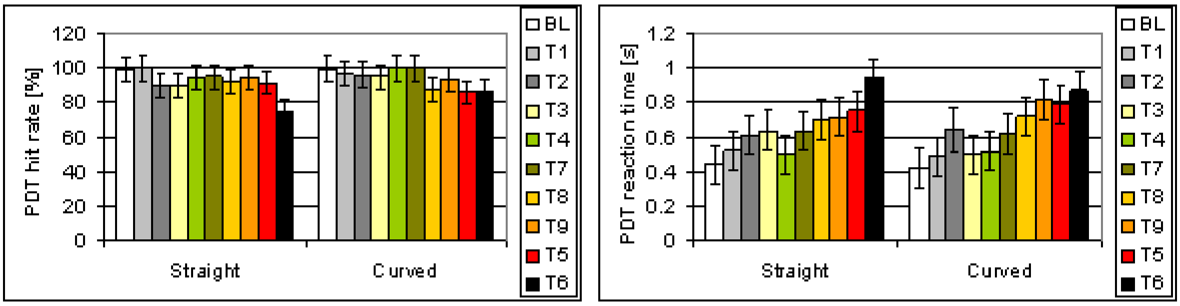


Figure 88 – PDT hit rate (left) and reaction time (right)

6.7.1.6. IVIS performance

The IVIS tasks were conducted, as instructed. There were a very few occasions where the participants could not perform the tasks without some guidance.

6.7.2. Effects of System B

6.7.2.1. Task length

The task lengths varied between 17 seconds (T9, add favourite) and 52 seconds (T6, destination entry). As for system A, it was found that the destination entry task was the most time consuming task. It was also found that the scale change task (T7) took rather a long time (35 seconds), although it was quite a simple task. As a comparison, the more difficult T8 (deeper into the menu hierarchy) took less time: 27 seconds (see Figure 89).

6.7.2.2. Self-reported driving performance

The most complex task (T6, destination entry) produced the highest value for drivers' reported driving performance (see Figure 89). The relationship between task length and effect on subj_r is not as clear as for system A, indicating that subj_r was not simply affected by task length, but also by task complexity/difficulty.

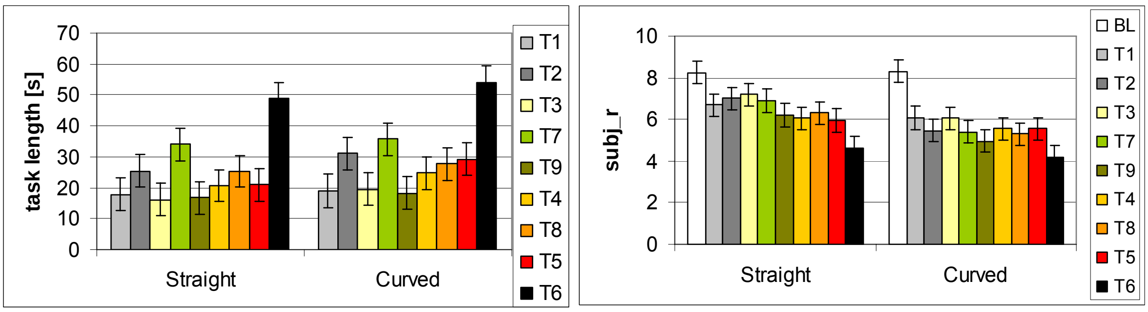


Figure 89 – Task length (left) and subjective rating of driving performance (right)

6.7.2.3. Longitudinal control

All tasks resulted in a significant speed reduction, primarily reflected in mean speed (mn_sp) (see Figure 90). Also, the headway increased, which was reflected in mean headway values and the headway histograms (Figure 91). This effect may be as a result of a compensatory speed reduction, rather than of trying to get farther away from the lead vehicle, since the lead vehicle was already at quite a far distance (>100m). The effect on speed was on average 5 km/h across tasks. This effect was not found in the speed change measure (d_sp), probably due to the speed change profiles not being linear, which is required for the least-square fitting method to be appropriate. Also, no effect was found in min speed (u_sp).

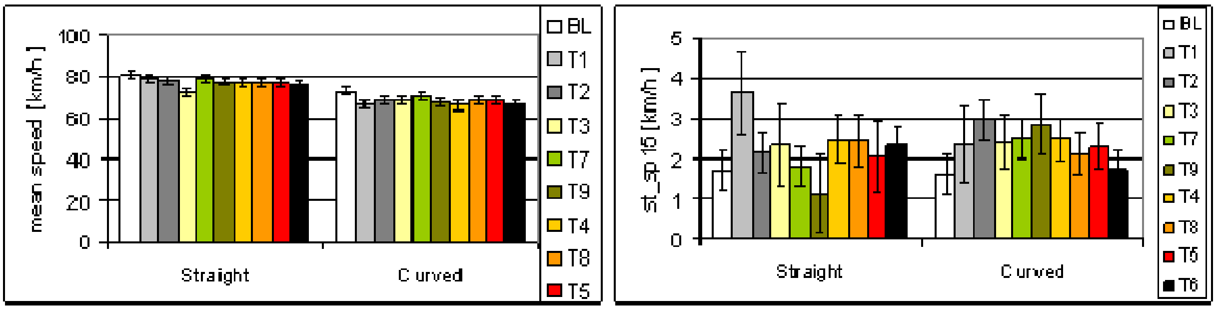


Figure 90 – Mean speed (left) and speed variation (right)

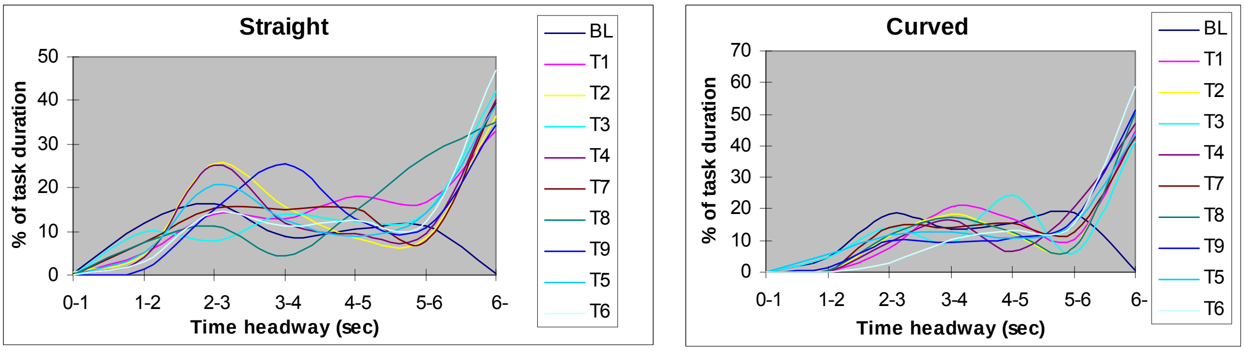


Figure 91– time headway histograms for straight (left) and curved (right) road sections

On average, the speed and headway variation increased, not only for the non-standardized measures, but also for the sliding window normalised measures. This indicates that the tasks differed in terms of effects on driving performance, independent of task length (see Figure 92).

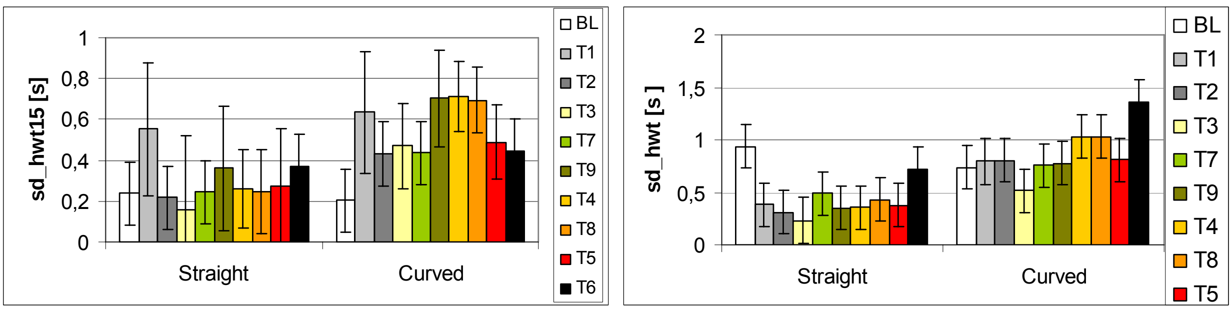


Figure 92 – Standardised time headway variation (left) and conventional time headway variation (right)

6.7.2.4. Lateral control

The lateral position was shifted towards the centre line as an effect of the tasks, probably to compensate for the increased secondary task demands (see Figure 93). The steering activity increased (Figure 94), the lateral position variation increased (Figure 93) and the Time-To-Line crossing decreased (Figure 94) as an effect of most tasks. The strongest effects were found in one-degree reversal rate (50% increase) and high frequency steering (11% increase). The three-degree reversal rate decreased, indicating that the number of larger steering reversals decreased as a result of IVIS performance.

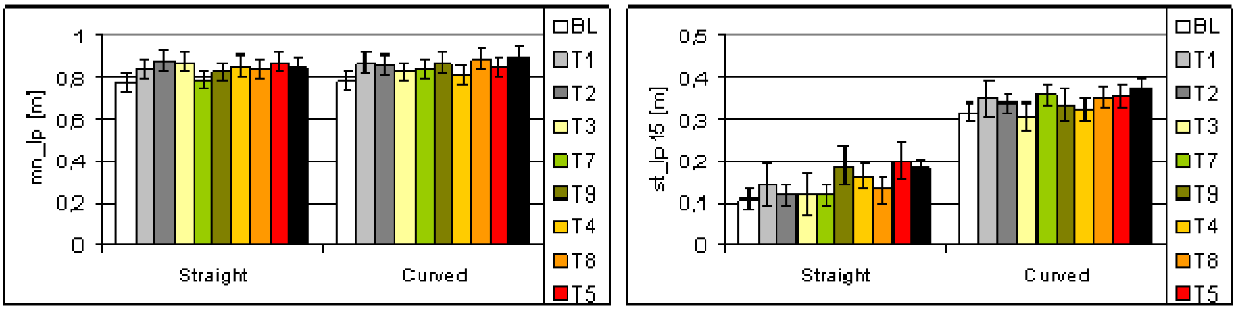


Figure 93 – Lateral position (right) lateral position variation (right)

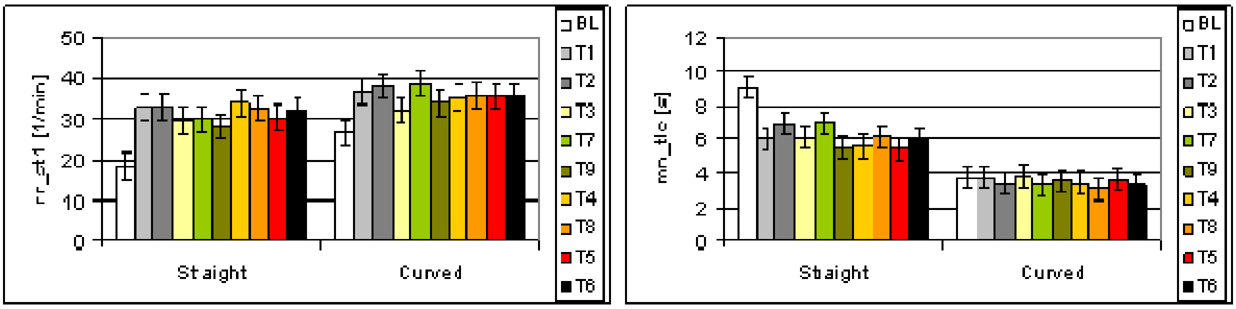


Figure 94 – One degree reversal rate (left) mean time-to-line crossing (right)

6.7.2.5. PDT results

The PDT reaction time increased as an effect of all tasks, but no differences were found between the tasks (Figure 95). On average, the increase in reaction time was close to 70%. The effect of the IVIS on mental workload was thus very high. The hit rate was on average 91% when attending to the IVIS, and statistically unaffected by the IVIS.

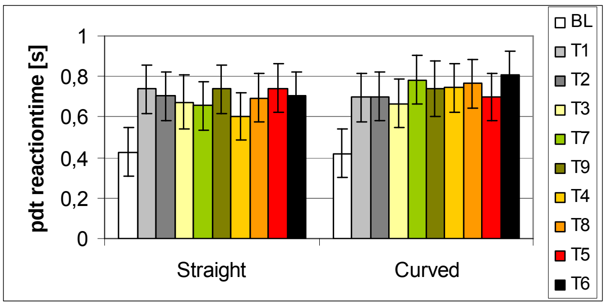


Figure 95 – PDT reaction time

6.7.2.6. IVIS performance

As for System A, participants needed guidance in completing the tasks on only a handful of times.

6.7.3. Comparison between Systems

System A had a larger variation in task durations, which was reflected in several of the included measures, however not in the subjective driving performance. In general, the effect of system A tasks on performance were larger, although there were exceptions. For system A, 25 measures were significantly affected by IVIS, and for System B, the corresponding number was 23.

6.8. Result summary and conclusions

Clearly, both systems and all tasks resulted in increased mental workload (e.g. as shown by PDT results) and behavioural changes (e.g. reduced speed, increased steering activity) perhaps because subjects tried to compensate for the increased visual/manual distraction. However, the vehicle control was somewhat deteriorated despite the attempted behavioural adaptation; the headway and speed variation increased, not only due to speed reduction, while the lateral position variation and TLC results indicated a less stable lateral control.

The tasks differed in completion time. Especially for System A, a similarity between inverted task length and several other measures could be found. Some measures were biased by task length (sd_hwd, sd_hwt, st_lp and st_sp). However, when an attempt was made to compensate for this bias with the "sliding window technique" (using sd_hwd15, sd_hwd30, st_lp15), differences were found between the tasks, indicating that the tasks differed in terms of instantaneous effect on driving performance. This was primarily found in System A.

The destination entry tasks required a long time, demanded high visual attention and also required some tricky manual control. These tasks affected driving performance severely: Subjective driving performance decreased 25-30%, speed variation increased 15-35% and lateral position variation increased 15-20%. But most importantly, reaction time to visual stimuli (PDT) was delayed by 60-70%, which corresponded to nearly 0.3 seconds or 6.3 metres at 75km/h. This difference can be the difference between a near crash or actual crash, especially at short headways.

6.9. Measures summary tables

6.9.1. System A

		Auditory			Visual/Manual						Effects		
Measure	**BL**	**T1**	**T2**	**T3**	**T4**	**T7**	**T8**	**T9**	**T5**	**T6**	**SLv**	**RLv**	**SLv*RLv**
taskl_l		12,91	17,16	17,78	22,28	10,36	28,02	44,15	52,09	61,36	✓	✓	✓
subj_r	8,18	7,71	7,50	7,63	7,26	7,32	5,76	5,37	5,26	4,71	✓	✓	✓
mn_sp	76,80	75,49	75,40	75,68	75,49	74,13	74,39	73,67	73,36	72,46	✓	✓	✗
u_sp	70,16	72,77	72,16	72,18	71,58	72,12	70,05	68,79	67,71	66,41	✓	✓	✗
st_sp	3,52	1,85	2,04	2,15	2,37	1,36	2,57	2,98	3,20	3,39	✓	✓	✗
st_sp15	1,75	.	1,92	2,26	1,98	.	2,13	1,88	1,94	2,00	✗	✗	✗
st_sp30	2,39			3,22			2,42	3,16	2,72	2,75	✗	✓	✗
d_sp	-0,13	-6,16	-3,12	-0,63	-1,28	7,37	-1,99	5,61	4,25	1,83	✗	✗	✗
mn_hwd	102,63	86,15	89,98	88,70	90,59	100,74	98,74	118,84	118,50	119,19	✓	✗	✓
sd_hwd	17,36	3,88	5,15	4,56	6,75	3,55	7,69	13,35	17,40	19,00	✓	✓	✓
sd_hwd15	3,73	.	4,55	4,01	4,77	.	4,86	4,98	5,26	5,19	✗	✗	✗
sd_hwd30	6,87			6,38			10,00	9,55	10,56	9,95	✓	✓	✓
u_hwt	3,46	3,78	3,89	3,80	3,69	4,50	4,13	4,62	4,21	4,10	✓	✗	✗
mn_hwt	4,88	4,15	4,32	4,24	4,32	4,89	4,87	6,00	5,98	6,09	✓	✗	✓
hwt_0_1	2,08	0,00	0,03	3,93	0,49	0,00	0,27	0,00	0,38	0,72	✗	✗	✗
hwt_1_2	9,42	13,30	8,14	8,79	11,60	9,48	7,10	2,17	2,49	3,46	✓	✓	✗
hwt_2_3	15,83	22,18	17,17	20,68	19,43	15,21	16,87	10,92	6,78	9,99	✗	✗	✗
hwt_3_4	17,13	13,67	26,01	11,65	21,12	13,92	16,82	11,13	17,95	12,47	✗	✗	✗
hwt_4_5	15,55	25,59	23,73	19,14	14,91	11,8	21,55	19,63	16,18	15,03	✗	✗	✗
hwt_5_6	10,58	7,67	10,14	22,00	10,72	19,15	10,76	15,89	11,30	16,26	✗	✗	✗
hwt_6	0,30	17,59	14,74	13,80	21,73	30,43	26,63	40,25	44,92	42,08	✓	✗	✓
sd_hwt	0,84	0,23	0,28	0,27	0,41	0,22	0,45	0,73	0,96	1,05	✓	✓	✓
sd_hwt15	0,21		0,24	0,22	0,30		0,31	0,32	0,33	0,34	✓	✓	✗
sd_hwt30	0,367			0,435			0,583	0,572	0,601	0,599	✓	✓	✓
mn_lp	0,77	0,80	0,81	0,79	0,80	0,80	0,92	0,89	0,89	0,88	✓	✗	✗
st_lp	0,24	0,19	0,20	0,20	0,19	0,19	0,26	0,27	0,28	0,28	✓	✓	✗
st_lp15	0,21		0,20	0,19	0,19		0,26	0,26	0,26	0,27	✓	✓	✗
st_lp30	0,22			0,24			0,30	0,27	0,27	0,29	✓	✓	✗
rr_st1	21,77	22,89	26,06	25,28	25,35	31,75	35,92	35,23	33,18	35,13	✓	✓	✗
mn_tlc	6,59	7,16	6,33	6,49	7,06	5,20	4,79	4,94	5,38	4,83	✓	✓	✓
hi_st	38,83	41,51	39,15	41,71	42,02	44,99	48,20	45,17	45,80	48,17	✓	✓	✓
pdt_hit	99,56	98,34	92,95	92,50	97,09	97,50	89,67	93,66	88,61	80,62	✓	✗	✗
pdt_rt	0,43	0,50	0,63	0,57	0,51	0,63	0,71	0,77	0,77	0,90	✓	✗	✗

6.9.2. System B

		Visual		Visual/Manual							Effects		
Measure	BL	T1	T2	T3	T7	T9	T4	T8	T5	T6	SLv	RLv	SLv*RLv
task_l		18,41	28,22	17,83	34,95	17,51	22,74	26,52	25,12	51,48	✓	✓	✗
subj_r	8,28	6,37	6,23	6,63	6,15	5,58	5,80	5,80	5,75	4,43	✓	✓	✗
mn_sp	76,75	72,85	73,45	70,16	74,28	72,53	71,27	72,39	72,59	71,47	✓	✓	✓
u_sp	70,24	68,18	67,93	66,86	68,60	68,74	66,88	67,88	68,71	65,57	✗	✓	✗
st_sp	3,33	2,98	3,41	2,74	3,30	2,43	2,77	2,91	2,41	3,34	✗	✗	✗
st_sp15	1,65	2,99	2,57	2,36	2,17	2,00	2,48	2,30	2,18	2,03	✓	✗	✓
st_sp30	2,24	6,66	3,47		3,53		3,17	4,20	2,74	2,97	✓	✓	✗
d_sp	0,26	-4,15	10,28	4,44	4,91	-3,96	-1,24	-6,38	-6,54	1,62	✓	✗	✗
mn_hwd	108,40	113,70	112,38	110,66	116,92	113,38	117,02	113,46	112,07	130,12	✓	✗	✗
sd_hwd	17,25	8,34	9,76	6,86	10,72	8,99	10,52	11,00	9,65	17,22	✓	✓	✓
sd_hwd15	3,75	7,62	5,00	5,21	5,29	8,07	7,05	6,41	5,51	5,41	✓	✓	✓
sd_hwd30	6,90	34,79	7,54		8,71		10,99	10,22	10,83	10,04	✓	✗	✓
mn_hwt	5,17	5,89	5,70	5,67	5,85	5,86	6,14	5,89	5,73	6,90	✓	✓	✗
u_hwt	3,74	4,94	4,59	4,87	4,77	4,90	4,97	4,72	4,70	4,82	✓	✗	✗
hwt_0_1	0,21	0,00	0,32	0,00	0,00	0,00	0,00	0,00	0,00	0,00	✗	✗	✗
hwt_1_2	8,25	2,48	2,33	7,30	2,02	3,98	3,89	1,49	5,42	1,54	✓	✗	✗
hwt_2_3	17,50	10,70	18,41	10,87	17,31	14,62	11,54	12,30	16,18	8,46	✓	✗	✗
hwt_3_4	11,21	16,97	16,96	11,64	14,07	14,46	10,88	17,52	12,60	10,71	✗	✗	✗
hwt_4_5	13,02	17,51	10,38	18,22	7,895	15,44	13,71	11,88	9,85	12,72	✗	✗	✗
hwt_5_6	15,11	13,40	8,16	10,15	15,15	10,37	17,47	14,03	14,71	13,77	✗	✗	✗
hwt_6	0,34	38,86	43,44	41,83	43,57	41,14	42,52	42,80	41,23	52,79	✓	✗	✗
sd_hwt	0,84	0,59	0,56	0,37	0,63	0,57	0,69	0,73	0,59	1,04	✓	✓	✓
sd_hwt15	0,22	0,59	0,32	0,32	0,34	0,53	0,49	0,47	0,38	0,41	✓	✓	✗
sd_hwt30	0,371	1,628	0,33	0,565	0,532		0,64	0,628	0,612	0,647	✗	✗	✗
mn_lp	0,78	0,85	0,87	0,85	0,81	0,85	0,83	0,86	0,86	0,88	✓	✗	✗
st_lp	0,24	0,24	0,24	0,21	0,25	0,26	0,25	0,26	0,26	0,30	✓	✓	✗
st_lp15	0,21	0,25	0,23	0,21	0,24	0,26	0,24	0,24	0,28	0,28	✓	✓	✗
st_lp30	0,23	0,26	0,25		0,24		0,26	0,25	0,26	0,30	✓	✓	✗
rr_st1	22,41	34,86	35,50	30,93	34,36	30,99	34,53	34,19	32,94	33,75	✓	✓	✗
rr_st3	6,38	4,86	5,16	4,94	5,15	4,55	4,51	4,63	4,52	4,64	✓	✓	✓
hi_st	38,50	44,05	47,11	39,28	49,29	38,68	41,39	44,56	40,15	45,21	✓	✓	✓
pdt_hit	98,77	97,22	94,54	87,97	92,30	88,06	90,08	89,26	87,78	88,60	✗	✗	✗
pdt_rt	0,42	0,72	0,70	0,67	0,72	0,74	0,67	0,73	0,72	0,75	✓	✓	✗

7. The Leeds simulator experiment: testing the reliability of a low cost driving simulator for the evaluation of IVIS

7.1. Test site

This experiment utilised two facilities: the Leeds Driving Simulator and the Leeds LabSim. The Driving Simulator is based on a complete Rover 216GTi with all of its basic controls and dashboard instrumentation still fully operational (Figure 96). On a 2.5m radius, cylindrical screen in front of the driver is projected a real-time, fully textured and anti-aliased, 3-D graphical scene of the virtual world. Realistic sounds of engine and other noises are generated by a sound sampler and two speakers mounted close to each forward road wheel. Although the simulator is fixed-base, feedback is given by steering torques and speeds at the steering wheel. Data were collected at 60Hz and include information of the behaviour of the driver (i.e. driver controls), that of the car (position, speed, accelerations etc.) as well as information on other autonomous vehicles in the scene (e.g. identity, position and speed).

LabSim is a low cost alternative to the full-scale driving simulator (Figure 96). LabSim runs the same software as the main simulator, but its driver controls and image generation are less immersive. The driver sits at a desk accommodating a Logitech Momo force-feedback steering wheel and pedals (accelerator and brake only). A real-time, fully textured and anti-aliased, 3-D graphical scene of the virtual world is displayed on an Acer 17" flat-panel display in front of the driver. The display is a single 1280x1024 channel with a horizontal field of view of 50° and a vertical field of view of 39°. It is generated by a dual processor, 3.2GHz PC hosting an NVidia FX3000G graphics card. This PC is connected via Ethernet to a Pentium 3 (CPU speed 700MHz) running the vehicle dynamics model at around 1.5MHz. To minimise latency, the driver controls are connected to this machine. The visual display update and data collection rates are 60Hz.

Figure 96 – The Leeds driving simulator (left) and the Leeds LabSim (right)

Both these environments were used to evaluate the effects of driving whilst interacting with an IVIS system.

The Seeing Machines' 'faceLAB 3.0 system' was used for collecting eye movement data. The analysis was made by using in-house developed analyses tool.

7.2. Scenarios and participants

A group of 18 average drivers (25-50 years old) were recruited for the study, see Table 19. They drove the standard rural road, in each of the driving simulators as described above.

Table 19 – Description of participants used in Leeds simulator study

	Males n=9	Females n=9
Mean Driver Age (SD)	32	30
Years held driving licence	14	16

7.3. IVIS included

This study involved the use of System B. Drivers used the system in both of the driving simulator environments.

7.4. Experimental design

The design of the experiment followed the general principles as described in Section 2.4. However, as the main aim of this experiment was to evaluate one testing environment against another, a within subjects experimental design was employed. This allowed us to minimise between subject effects. Nine drivers used System B in the LabSim first, followed by the Driving Simulator. The remaining nine completed the experiment in the reverse order. All drivers completed all tasks for System B, as described in Section 2.2.2 Task 9 however, was excluded as it relied on the availability of a GPS signal, which was not possible in the simulator laboratory.

7.5. Procedure

All participants were provided with written instructions for the experiment, and completed a consent form. Participants then underwent an intensive training period with System B to ensure they were fully familiarised with the tasks. Participants were then calibrated with the eye movement cameras. They were then allowed a familiarisation period in the simulator (or LabSim) and given the opportunity to practice all the tasks whilst driving. Once they felt confident they could remember and perform the tasks, they completed the experimental drive. Following this, a rest period was allowed before practising in the other environment. Once this practice was completed, they undertook the remaining experimental drive. Each participant was paid £30 at the end of the study.

Group1 N=9	Briefing	Static task practice	Dynamic task practice (Driving Sim)	Experimental trial (Driving Sim)	Dynamic task practice (LabSim)	Experimental trial (LabSim)	Debrief
Group2 N=9	Briefing	Static task practice	Dynamic task practice (LabSim)	Experimental trial (LabSim)	Dynamic task practice (Driving Sim)	Experimental trial (Driving Sim)	Debrief
Time	*5 mins*	*20 mins*	*20 mins*	*40 mins*	*20 mins*	*40 mins*	*5 mins*

7.6. Measures and analysis method

The mandatory measures as described in Section 2.7 were used. In addition, the optional eye movement measures were recorded, although, due to equipment limitations, these were only achieved in the simulator. The analysis was undertaken as previously described in Section 2.7. In addition to the fixed factors of IVIS task and Road Level, the factor of Simulator Type was also included (Driving Simulator or LabSim).

The data were pooled and overall main effects were examined; these are presented in the figures to follow. Where a main effect for the Simulator Type was found, additional graphs are presented to highlight the differences.

7.7. Results (System B)

7.7.1. Self-reported driving performance

There was found to be a significant effect of the type of task on subjective ratings. In general, subjective ratings decreased as the complexity of the tasks increased (Figure 97). Ratings for all the tasks were significantly different from baseline driving.

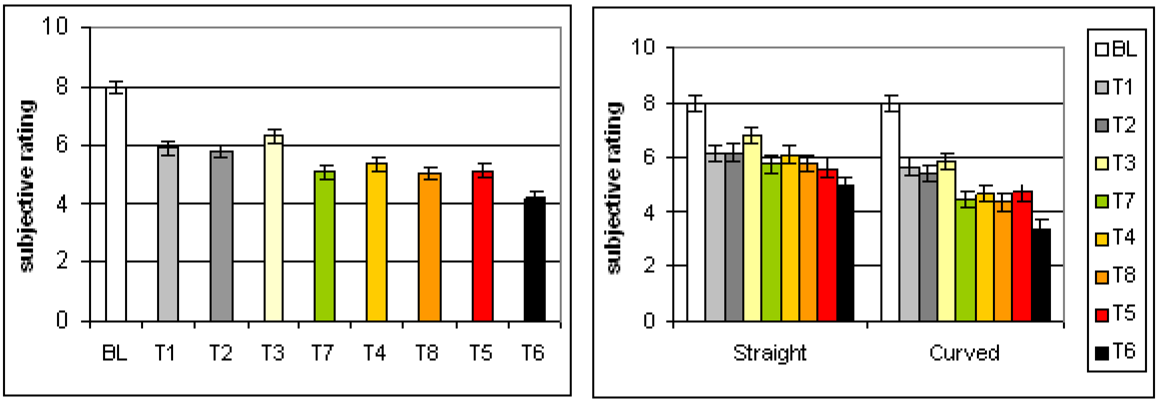


Figure 97 – Mean subjective ratings, main effect and by road type

There was a significant main effect also of road type, with average ratings being 6.1 on the straights and 5.1 on the curved sections of road. Accompanied by a significant interaction with task type, this meant that drivers reported a worsening of their driving performance on curves, compared to when driving on straights, for some particular tasks. This was particularly so for tasks 4 and 6, see Figure 97. There was a main effect of testing environment, whereby drivers reported that their performance was significantly better overall in the simulator compared to in the LabSim, see Figure 98.

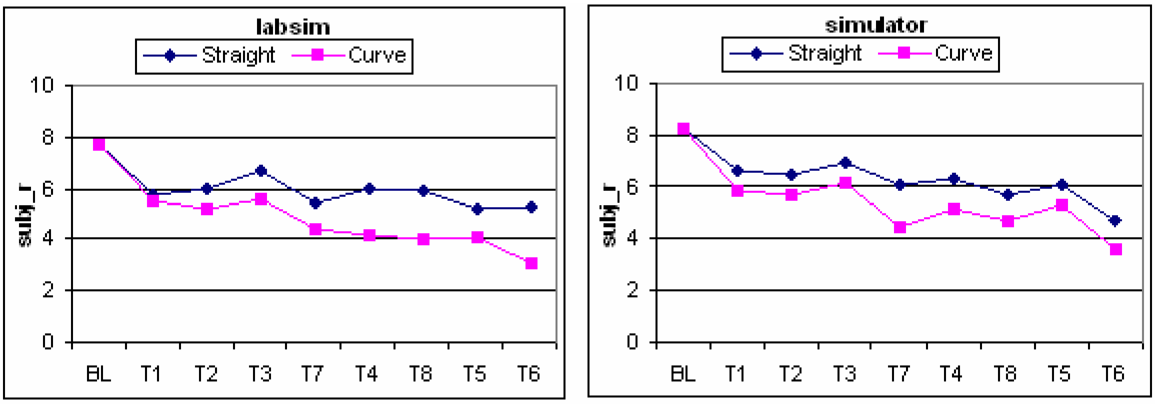


Figure 98 – Mean subjective ratings in the LabSim (left) and Driving Simulator (right)

7.7.2. Task length

The time taken to complete each of the tasks varied widely between tasks and between road types. The easier visual task (T1) took some drivers just one second to complete, whilst the most difficult tasks took up to 50 seconds to complete, for other drivers. Main effects were noted for both these factors (Task type and Road type), see Figure 99. The most time consuming tasks were T2 (the visual task), T7 (scale change task) and T6 (destination entry). The absence of an interaction between these two factors indicates that the tasks took similar amounts of time to complete, regardless of the road environment.

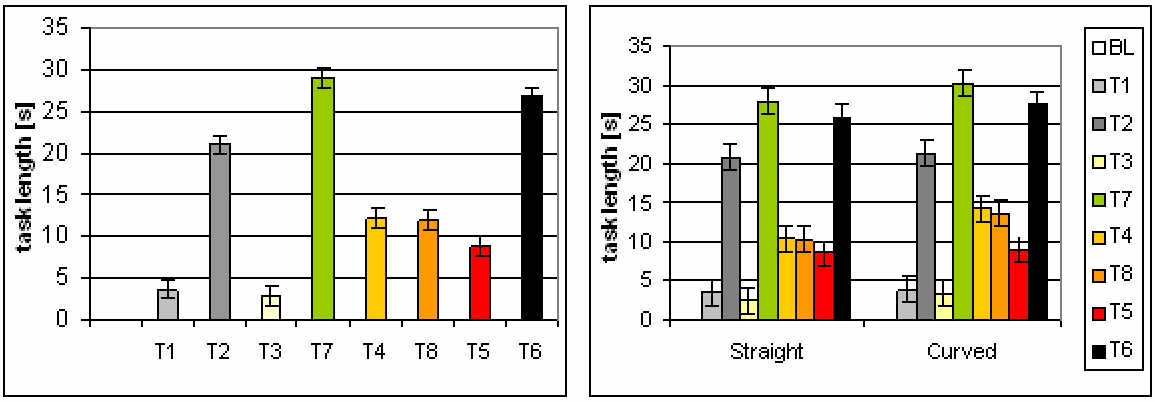


Figure 99 – Mean task lengths, overall (left) and by road type (right)

A significant negative correlation exists between subjective ratings and task length, such that as task length increased, drivers rated their performance as worsening. There was no effect of testing environment, suggesting that drivers took comparable amounts of time to complete the task, regardless of whether they were using the Driving Simulator or the LabSim.

7.7.3. Task completion

Of the 576 tasks presented across all drivers, only 5 were not completed within the time allowance. It was mostly Task 7 (scale change task) that the drivers were unable to complete.

7.7.4. Longitudinal control

The completion of the tasks (regardless of which one) resulted in a significant speed reduction, compared to baseline driving. Mean speed reduced by approximately 3km/h for the easier tasks (T1) and 6 km/h for the harder visual/manual task (T8), see Figure 100. A main effect of road environment indicated that, overall, drivers travelled slower on the curves (77 km/h) compared to the straights (79km/h). This is unsurprising given the nature of the road layout, but the absence of a significant interaction between the factors suggests that drivers reduced their speed in a similar fashion when performing the tasks, regardless of the road layout. There was no effect of the type of simulator used, indicating that drivers' speed choice was similar in Driving Simulator and the LabSim. A similar trend was seen in the results for minimum speed.

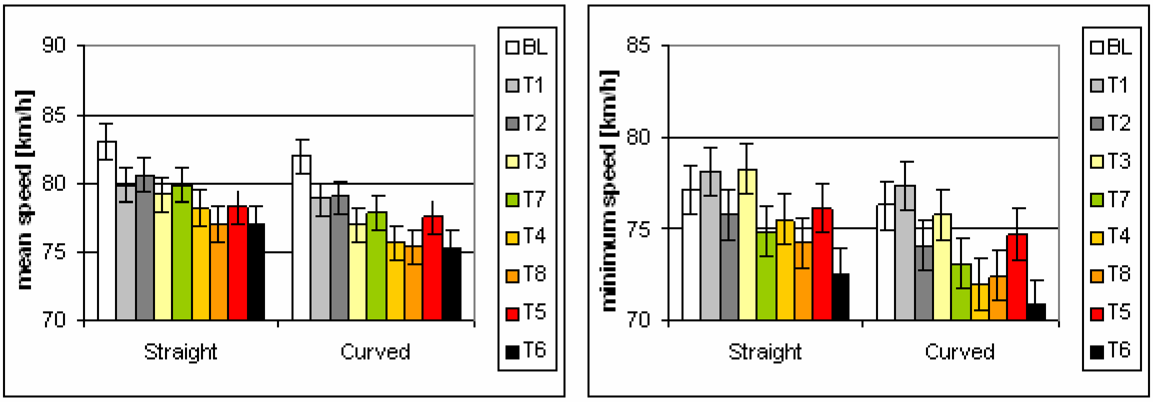


Figure 100 – Mean speed (left) and minimum speed (right)

However, not all drivers reduced their speed whilst performing a task, and some even increased their speed. Figure 101 shows how, across all tasks and baseline periods, the majority of drivers increased or decreased their speed only slightly (less than 5%).

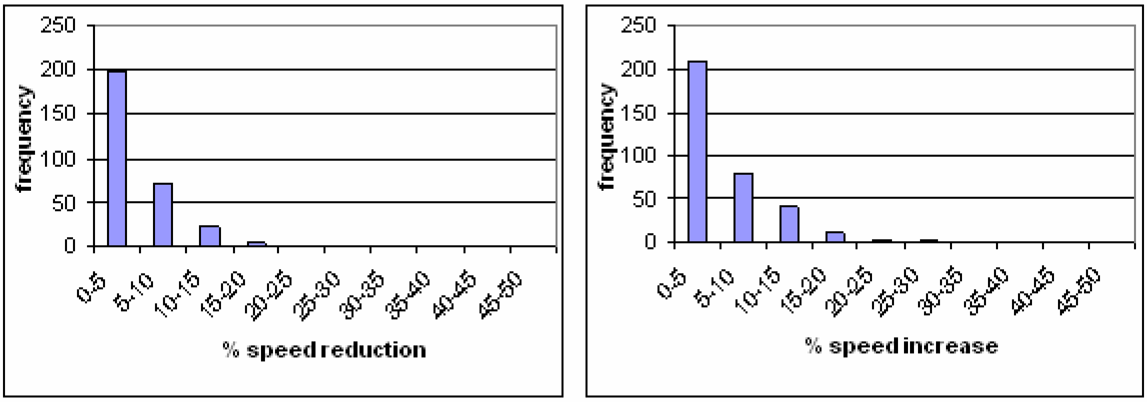


Figure 101 – Frequency of speed increase and decreases

This split between increases and decreases was deemed interesting, so the data were analysed separately, by task, Table 20 (a 5% threshold in each direction was used). Task 6 caused one third of drivers to decrease their speed by more than 5%, whilst Task 7 caused the same proportion drivers to increase their speed. These speed increases and reductions were stable between the two driving environments (Driving Simulator and LabSim).

Table 20 – Percentage of drivers who increased and decreased their speed during task completion

	% drivers who decreased speed	% drivers who increased speed
Baseline	12.5%	16.7%
Task 1	9.7%	25.0%
Task 2	12.5%	23.6%
Task 3	8.3%	13.9%
Task 7	12.5%	36.1%
Task 4	23.6%	30.6%
Task 8	15.3%	8.3%
Task 5	13.9%	20.8%
Task 6	33.3%	18.1%

A reduction in speed, of course, was coupled with a significant increase in mean time headway to the vehicle in front (Figure 102). Thus, whilst drivers were completing the tasks, their headway increased on both straight and curved sections of the road. There was no effect of the type of simulator used, as would be expected given the speed results reported above.

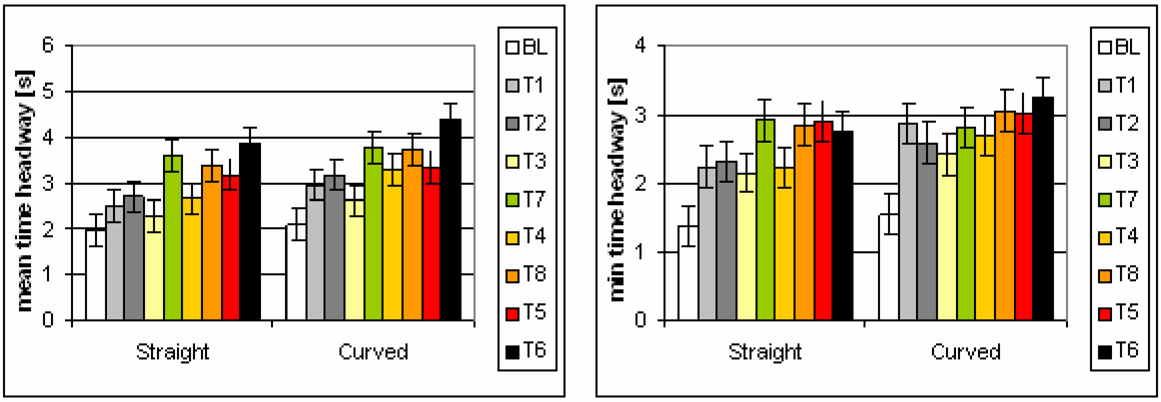


Figure 102 – Mean time headway (left) and minimum time headway (right)

On average, in the baseline sections, drivers maintained two second headways. This increased gradually as task difficulty increased – to four seconds for the destination entry task (T6).

The headway distributions are shown in Figure 103. As already noted, there was a general increase in headways when tasks were completed. Drivers did not spend more time at shorter headways when completing the tasks, compared to baseline, even for tasks with relatively long completion times.

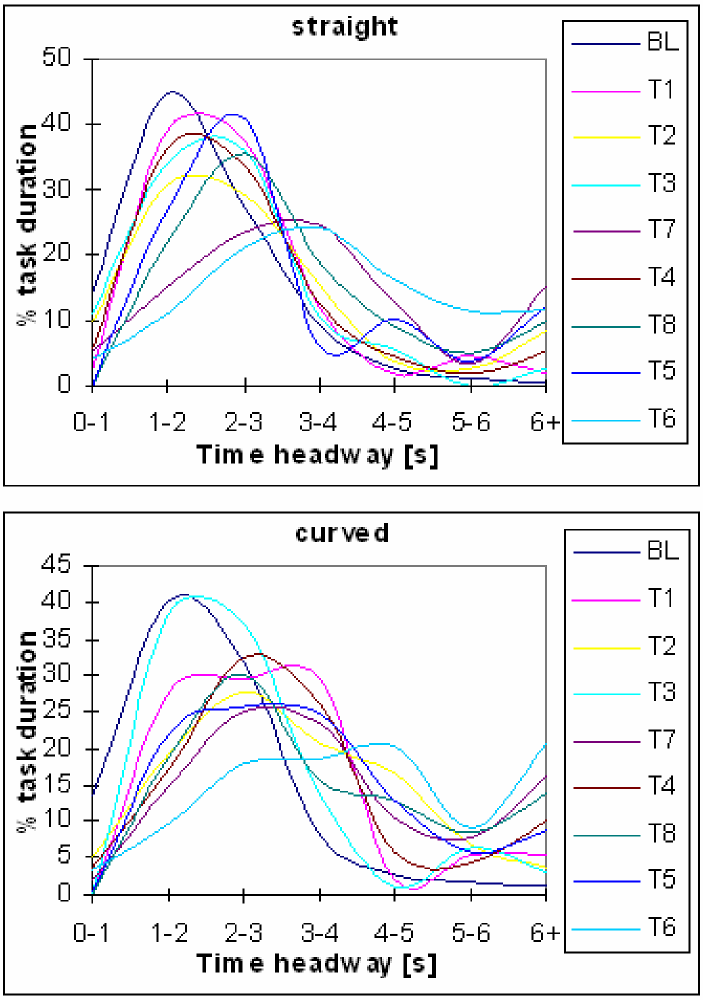


Figure 103 – Time headway distributions during task completion

There was a difference, however, between the amounts of time drivers spent at short headways in the two simulators. Drivers were more inclined to spend longer periods of time at shorter headways (0-2 seconds) in the LabSim. Figure 104 shows the amount of time drivers spent at these short headways whilst they completed the task. Even in the baseline, it appears that drivers were more inclined to travel close to the vehicle in front.

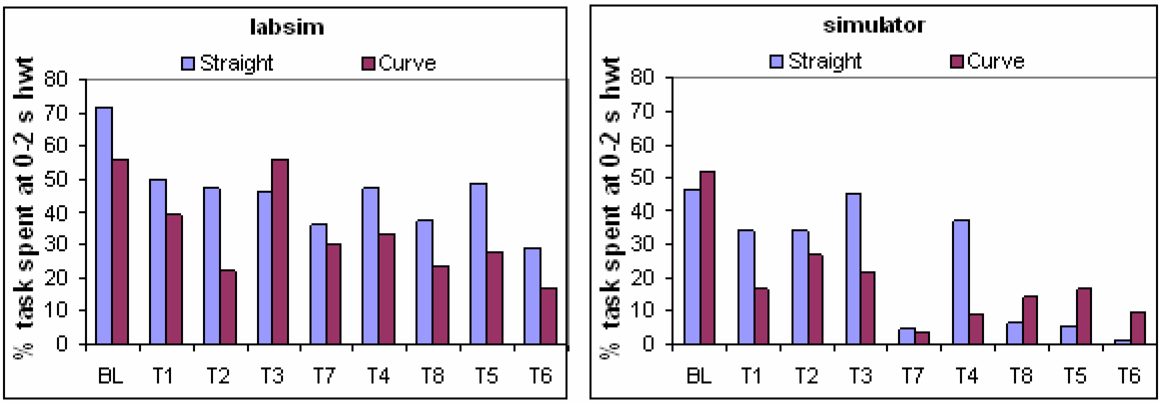


Figure 104 – Critical short headways in LabSim (left) and Driving Simulator (right)

7.7.5. Lateral control

There was little change in mean lateral position, apart from when drivers completed Tasks 8 and 5 where drivers travelled closer to the edge line, Figure 105. The lateral position variation results are more difficult to interpret, as the sliding window technique gave different results depending on the size of window used. Not using a sliding clearly gives results biased by task length (Figure 105).

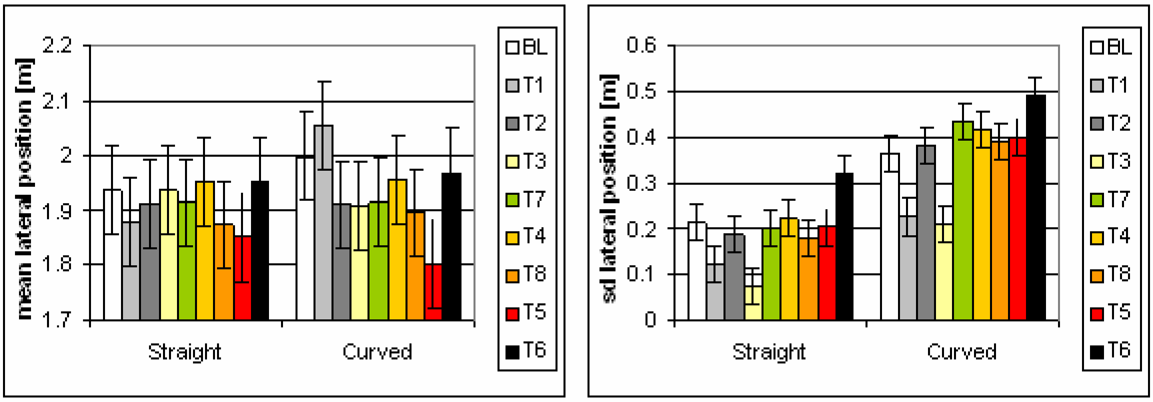


Figure 105 – Mean lateral position (left) and lateral position variation (right)

However, Figure 106 shows the 15 second sliding window results: they indicate that except for Tasks 2, 6 and 7, lateral position variation decreased during task completion. In contrast, where a longer sliding window is used (30 seconds) the data look different. Due to the loss of much data (due to task lengths being shorter than 30 seconds) it is probably not so reliable.

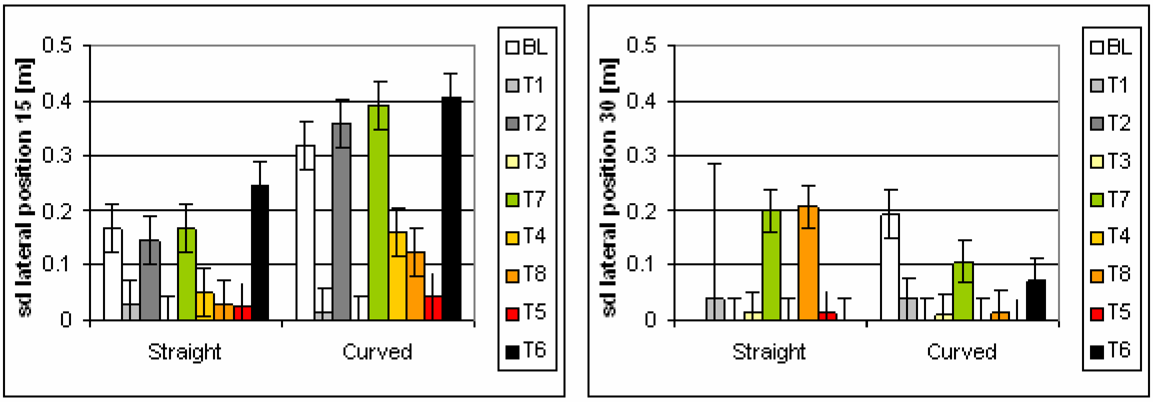


Figure 106 – Lateral position variation using sliding windows (15 secs left, 30 secs right)

There was a main effect of the type of simulator used on the amount of variation in lateral position. Using the normal technique and the 15 second sliding window technique, there was more deviation in the LabSim than in the Driving Simulator. Figure 107 shows that this reduction in lateral position variation was particularly apparent on the straight sections of road.

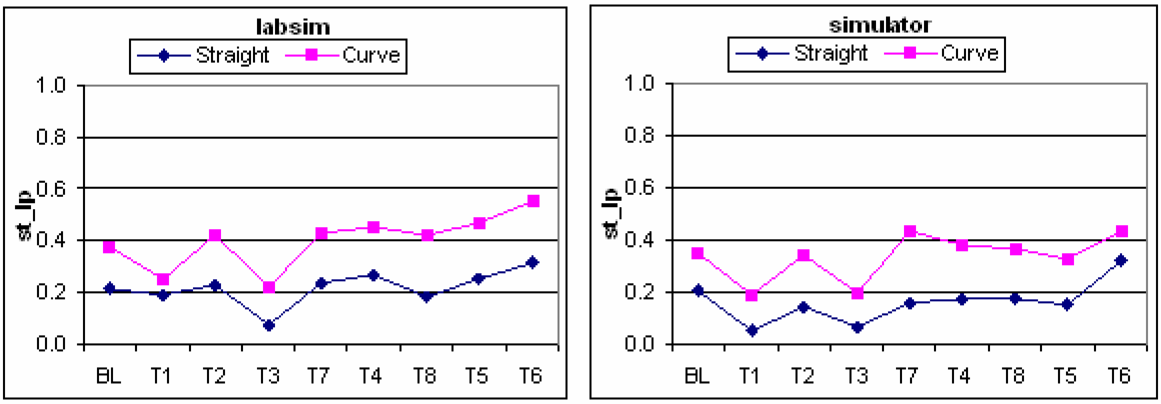


Figure 107 – Variation in lateral position in labsim (left) and full-scale simulator (right)

This increase in lateral variation in the LabSim maybe reflected in the subjective ratings - drivers reported that their performance was significantly poorer in the LabSim, (Figure 98).

The steering measures indicate an increase in steering activity, as seen in Figure 108. Nearly all tasks induced more rapid steering movements up to a threshold of 20 degrees. In general, there were no effect of road type (straight or curve) or between the two testing environments.

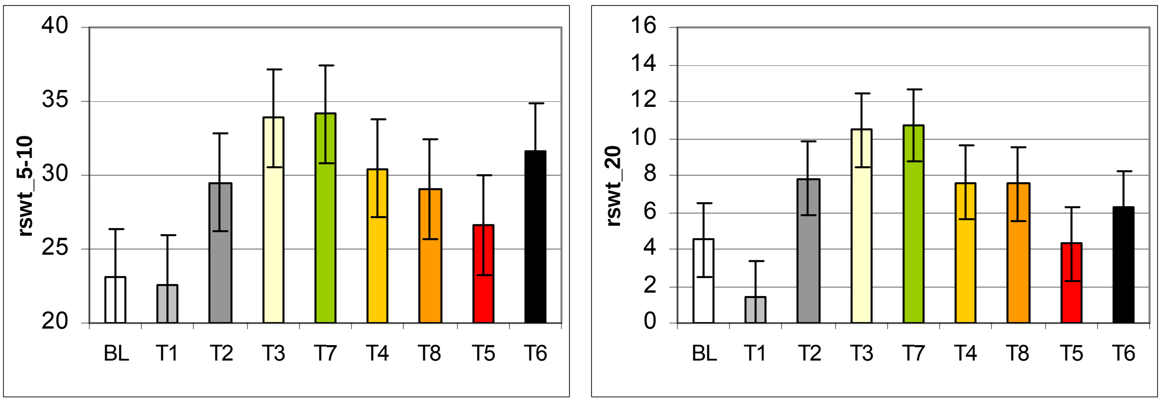


Figure 108 – Rapid steering movements with thresholds of 10° (left) and 20° (right)

The steering wheel reversal rate data is more difficult to interpret. It appears that for some tasks the reversal rate increases, whilst for others it decreases. Figure 109 indicates that the easier tasks (T1-T7) did not increase the reversal rate (and in most cases it decreases). Those tasks that involved more manual components were unsurprisingly associated with more steering reversals, compared to baseline. These results should be treated with caution as reversal rates are dependent on task length.

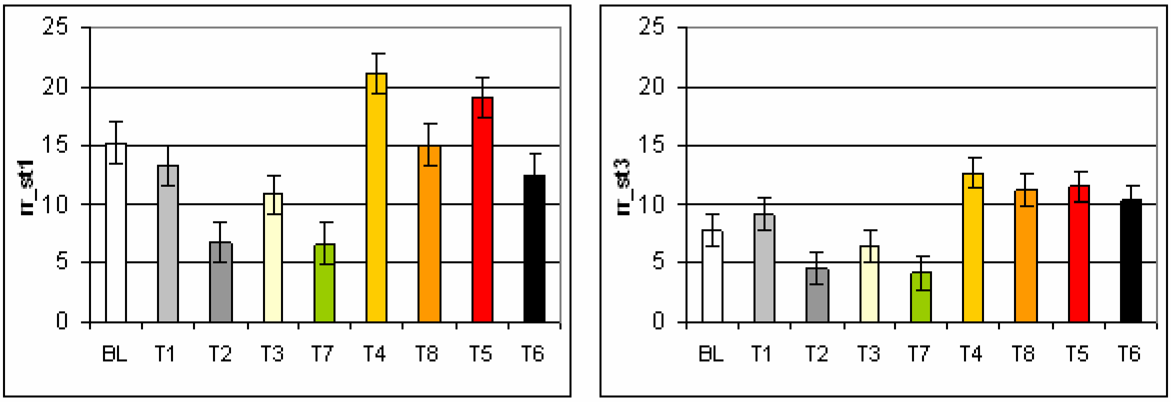


Figure 109 – Steering wheel reversal rate with thresholds of 1° (left) and 3° (right)

There appears to be a difference in reversal rate data, depending on the type of simulator. A main effect was found for both the one degree and three degree thresholds. Overall, the reversal rate was higher in the Driving Simulator than in the LabSim, see Figure 110. It can be seen that if we consider only the baseline situation, it appears that in the LabSim, there are very few reversals on curves. This is likely to be an artefact of the data.

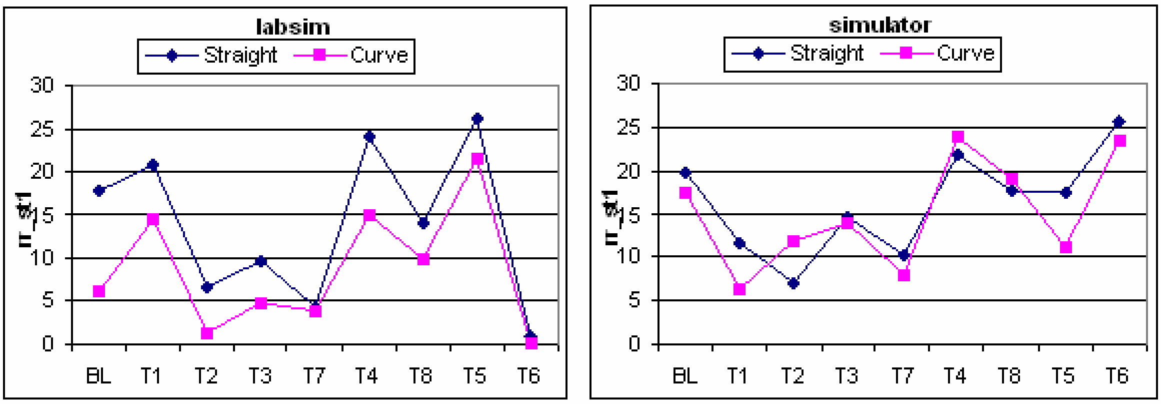


Figure 110 – Steering wheel reversals (1°) in LabSim (left) and Driving Simulator (right)

Data were also collected relating to the amount of time drivers spent out of lane. This amount of time varied widely, making this data difficult to interpret. Therefore a calculation of the proportion of drivers who exceed the lane boundary was made, Figure 111. In general, more drivers exceeded the lane boundaries during both baseline driving and during task completion when in the LabSim. As would be expected, drivers exceeded the lane boundaries more when driving on curves, compared to straights.

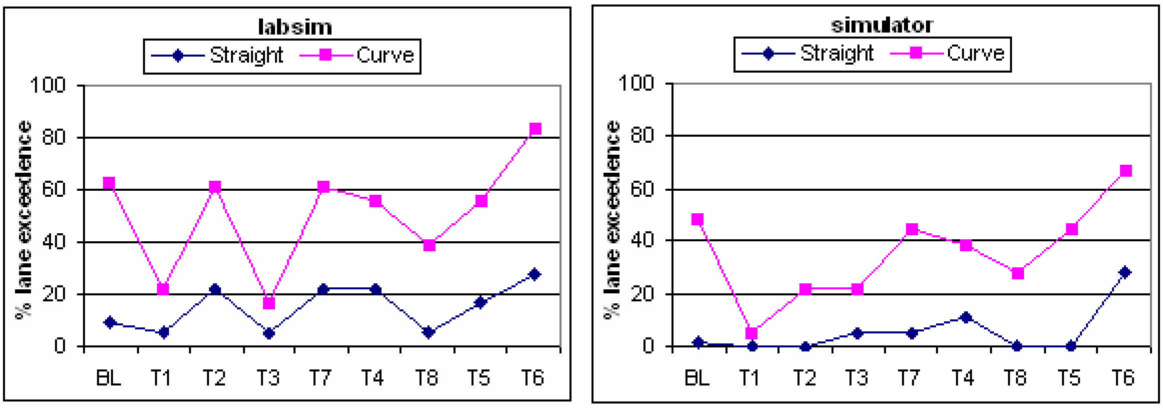


Figure 111 – Percentage of drivers who exceeded the lane boundary during task completion

Average minimum time-to-line crossing values were calculated (with an upper threshold of 20 seconds). A main effect of task type was found, with Tasks 5 and 8 producing longer minimum time-to-line crossing values (i.e. safer) compared to baseline driving (Figure 112).

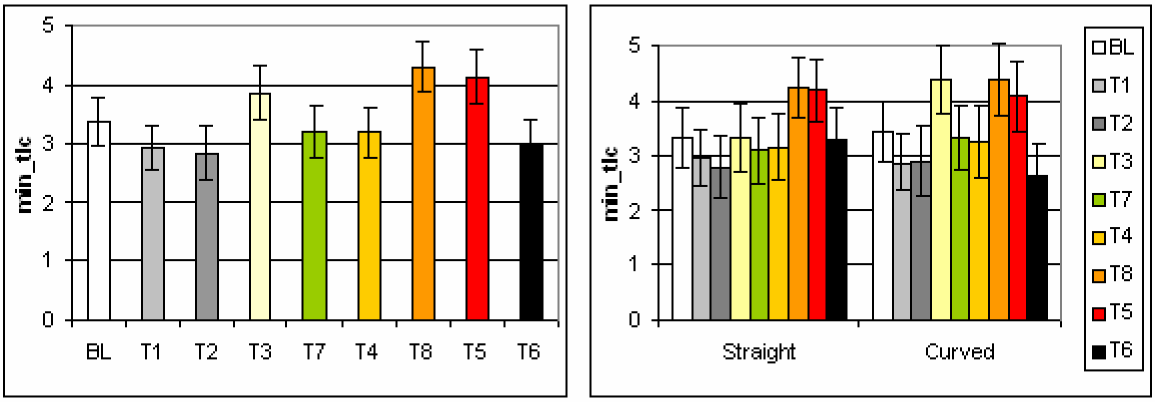


Figure 112 – Minimum time-to-line crossings, overall (left) and by road type (right)

A main effect of simulator type was also found; whereby minimum time-to-line crossings were shorter in the LabSim, compared to the Driving Simulator, by approximately 1 second, see Figure 113.

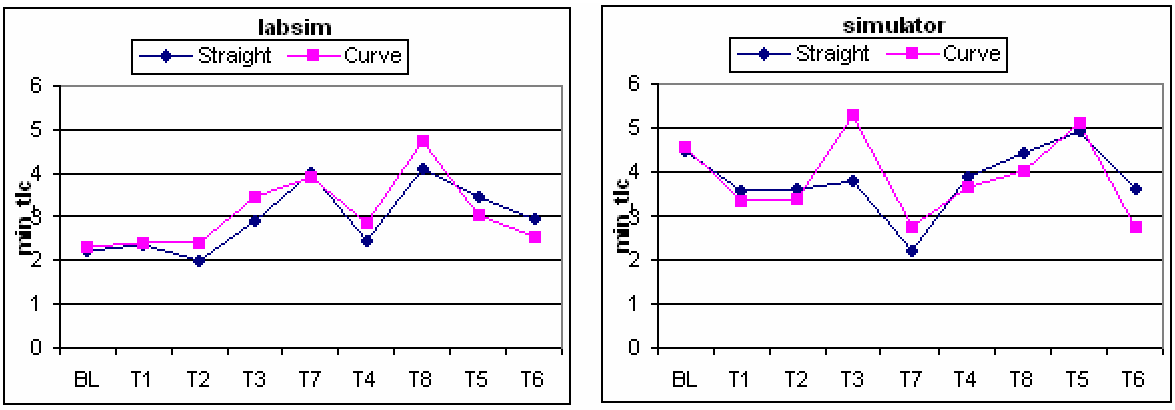


Figure 113 – Min time-to-line crossing, in LabSim (left) and Driving Simulator (right)

7.7.6. Eye Movements

Percentage of glances to the road centre and glance-based measures to the system were computed for all participants whilst they drove the simulator.

Mean duration of single glances to the system was found to increase with task difficulty (p= .001) (**Figure 114**). There was a significant difference in mean glance duration between the two road types, with longer durations for the straight sections. This is presumably because driving the curved sections required more of the drivers' visual attention towards the road.

The number of glances towards the system was also shown to be affected by task type (p < .001), although analysis failed to show a significant difference between the two road levels. The particularly high number of glances seen for tasks 2, 6 and 7 is explained as follows: (i) *Task 2* contained a large number of written instructions, and drivers therefore were forced to shift their gaze between the system and the road a considerable number of times (ii) a very short delay was experienced by the system following each input from the driver for *Task 6,* prompting drivers to look towards the system more often to check whether this input had been received (iii) of all the tasks, *Task 7* comprised of the largest number of visual/manual interactions with the system, requiring a larger number of glances between system and road.

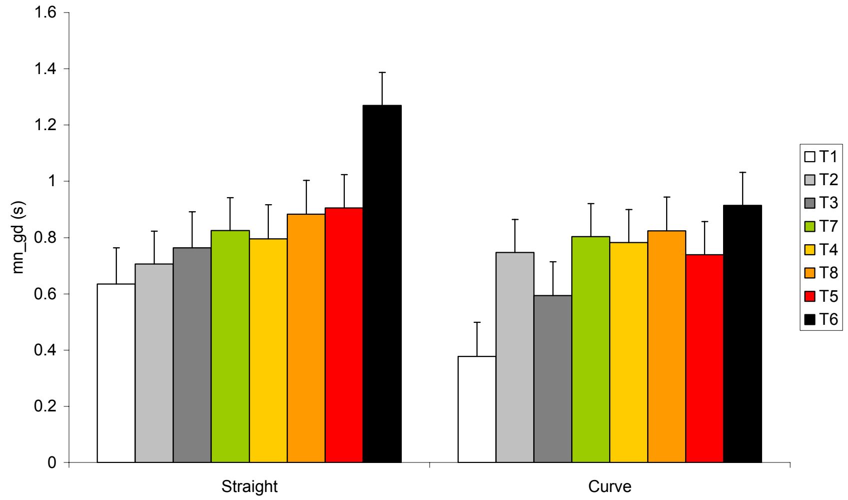


Figure 114 – Mean duration of single glances to the system for each task type

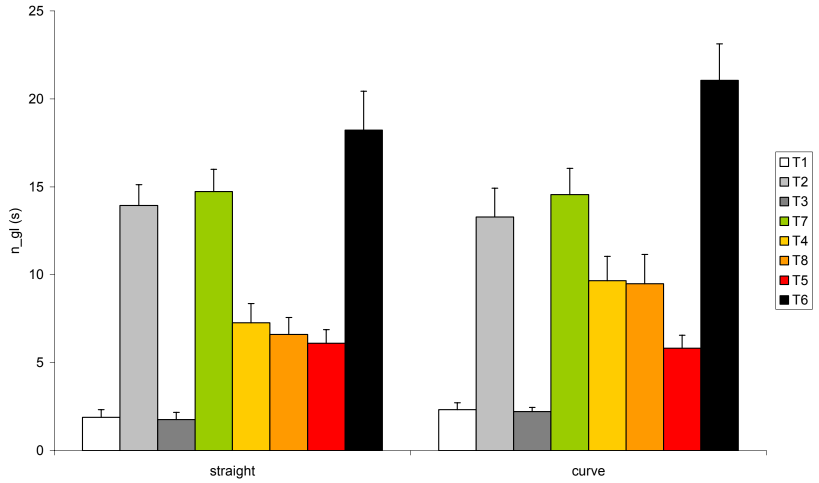


Figure 115 – Frequency of glances to the system for each task type

The pattern of eye movements towards the road centre was examined using Percent Road Centre (Victor and Johansson, 2005). Figure 116 shows the mean percentage of driver fixations (within one minute) that fell within a specified area of the road centre. As shown in the figure, when compared to baseline, the percentage of fixations to this area fell during performance of each of the eight tasks, a one-way analysis of variance showed this difference to be significant between all tasks and baseline ($p < .01$). Task difficulty did not have an effect on the percentage of fixations on road centre and this value was also not found to be different for the two road sections.

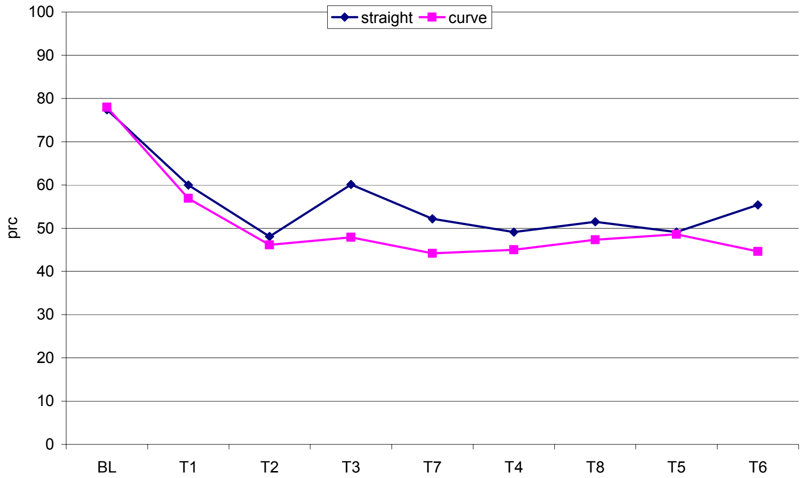


Figure 116 - Percentage road centre by road type

The standard deviation of gaze angle was found to increase from baseline when participants were engaged in the tasks. Analysis of variance showed no difference in this value across the tasks types, but a reliable increase from baseline ($p < .01$).

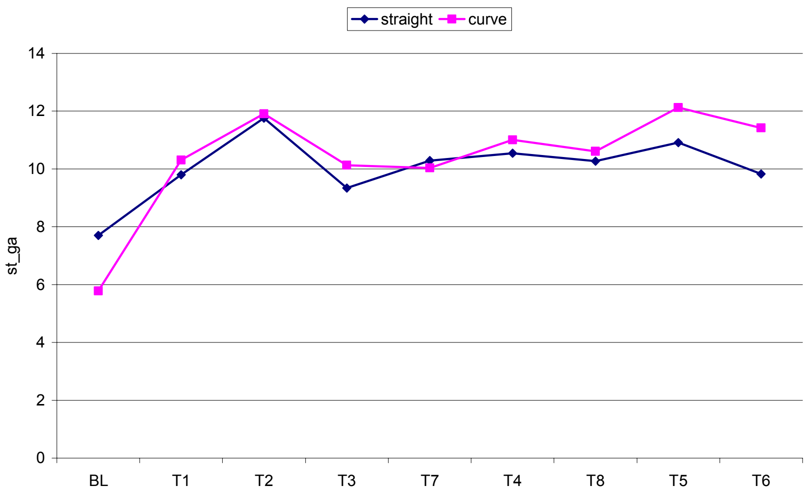


Figure 117 – Standard deviation of gaze angle by road type

7.8. Summary and conclusions

The main objective of the study was to be able to make comparisons between data derived from a low-cost simulator and that from a full-scale one. We were interested to see if the low-cost simulator was as effective as the full-scale one in inducing changes in behavioural parameters associated with distraction. This was achieved by using the same group of drivers, carrying out identical tasks in both environments, on identical pieces of road.

Overall, the completion of the IVIS tasks was associated with changes in driving behaviour. Drivers were able to perceive these changes in behaviour in their self-reporting of reduced performance, although we cannot be completely sure what particular aspect of their driving was poorer. The tasks were varied in the amount of time they took drivers to complete and thus provided an indication of difficulty – although this was also influenced by the complexity of manual operations involved in the task. All drivers were able to complete the tasks (with only a few exceptions) within the 90 second window, and only a few drivers required help to complete the tasks. Drivers reported that they found driving on the curves more difficult overall than the straights, and that their performance was particularly poor on curves for some specific tasks.

With regards to a priori task groupings, there was a clear trend with the subjective ratings. Those tasks that were considered a priori as more difficult were indeed rated so by the drivers. Overall, drivers rated their performance as worse when they drove the LabSim, suggesting that drivers believed the more immersive environment of the full-scale simulator aided their driving performance. Drivers were, however able to complete the tasks in comparable times in the two driving environments.

The completion of the tasks induced an overall reduction in speed, compared to baseline driving. This reduction in speed was similar on the straights and curved sections of the road. Further analysis showed that drivers did not uniformly reduce their speed, and that indeed some drivers increased their speed during task completion. Tasks 4 (Change settings) and 6 (Data entry) caused the greatest number of drivers to decelerate. The completion of Task 4 also caused a significant proportion of drivers to increase their speed, along with Task 7 (Scale change task). These increases and decreases in speed were similar across the two driving environments, suggesting that drivers reacted in similar ways (in terms of speed performance) to changes in their workload.

These changes in speed meant that drivers altered their headway to the lead vehicle. In general, this translated to an increase in headway. There were, however, differences between the two driving environments when the headway distributions were studied. When using the LabSim, drivers were more inclined to spend longer periods of time at shorter headways. This may be an indication of reduced realism or perceived risk presented in the LabSim and could have contributed to the decrease in subjective ratings that were observed in the LabSim.

With regards to lateral control, most tasks induced a decrease in lateral position variation, probably as a direct result of the speed reductions noted above. The most reliable results appear to be provided by using a 15 second sliding window – larger windows cause a loss of a significant proportion of the data and thus become unstable. However, new problems arise with the sliding window technique (see discussion on this matter in section the appendix on the technique). The a priori groupings were not robust here as there was no pattern or trend across the tasks. In contrast to the speed results, there were some differences in lateral control observed between the Driving Simulator and LabSim. Drivers exhibited poorer control in the LabSim – and this was attributed directly to their performance on straight sections of road. Drivers using the LabSim were not able to control “the vehicle” as well as in the Driving Simulator when driving on straight sections of road, during task completion (values were comparable during baseline driving). This poorer vehicle control could have affected the subjective ratings, particularly as it appears that drivers actually exceeded the lane boundaries more when driving in the LabSim. This is also supported by the minimum time-to-line crossing values.

In summary, as far as the tasks were concerned, their completion caused (varying) effects on longitudinal control. The a priori groupings of tasks, in terms of difficulty, appeared to be reflected in the mean speed results – the more difficult tasks produced lower mean speeds. With regards to the ability of the LabSim to replicate the results from the full scale driving simulator, similar changes in mean speed were observed. On closer inspection of the distribution however (in headway data), it appears that drivers spent longer periods of time travelling at shorter headways (in both the baseline and task completion sections). For lateral control, the results are more difficult to interpret and the a priori groupings were not predictive of the performance results. When completing some of the tasks, drivers wandered in their lane more, but conversely due to a speed reduction (and thus a reduced workload) their lateral performance improved for some tasks. In contrast to the longitudinal results, LabSim did not reproduce the same results as the full-scale driving simulator. Drivers have more difficulty in controlling the vehicle, perhaps as a direct result of the lack of field of view in the LabSim and reduced feedback from the vehicle controls.

7.9. Measures summary tables

7.9.1. System B – specific for Simulator

		Visual		Visual/Manual						Effects		
Measure	BL	T1	T2	T3	T7	T4	T8	T5	T6	IVt	RLv	IVt*RLv
task_l	76.29	3.19	21.39	3.08	28.03	12.71	12.41	8.42	29.08			
subj_r	8.25	6.22	6.08	6.56	5.31	5.72	5.19	5.67	4.17	✓	✓	×
mn_sp(km/h)	81.74	79.25	80.02	78.76	78.27	76.89	76.67	77.76	76.44	✓	×	×
u_sp(km/h)	76.15	77.83	74.72	77.68	73.84	73.41	73.49	75.12	71.43	✓	×	×
st_sp(km/h)	2.70	0.77	2.92	0.73	2.61	2.29	1.97	1.57	2.79	✓	×	×
st_sp15(km/h)	2.35		2.66		1.92	2.00	3.66		2.38	×	×	×
st_sp30(km/h)	2.51		3.32		2.13	6.35	5.16		2.63	✓	×	×
d_sp(km/h)	0.00	0.01	0.05	-0.01	0.03	0.01	0.01	0.03	-0.02	✓	×	×
mn_hwd(m)	47.64	63.56	65.55	52.19	83.75	65.07	73.78	73.57	88.37	✓	✓	✓
u_hwd(m)	36.46	61.44	56.58	50.11	68.43	55.71	63.75	68.00	36.46	✓	✓	✓
sd_hwd(m)	6.55	1.18	5.14	1.26	8.18	5.76	6.09	2.98	11.70	✓	×	✓
sd_hwd15(m)	2.63	0.00	3.97	0.00	5.26	1.48	1.22	0.00	6.98	✓	×	×
sd_hwd30(m)	4.11	0.00	0.21	0.00	3.19	0.31	1.04	0.00	6.03	✓	×	×
mn_hwt(s)	2.10	2.95	3.00	2.41	3.93	3.14	3.53	3.48	4.24	✓	✓	✓
u_hwt(s)	1.59	2.86	2.58	2.29	3.18	2.65	2.98	3.23	3.16	✓	✓	✓
sd_hwt(s)	0.30	0.05	0.25	0.07	0.42	0.30	0.33	0.14	0.59	✓	×	×
sd_hwt15	0.13	0.00	0.20	0.00	0.27	0.07	0.06	0.00	0.37	✓	×	×
sd_hwt30(s)	0.19	0.00	0.01	0.00	0.15	0.02	0.05	0.00	0.30	✓	×	×
mn_lp(m)	1.51	1.57	1.45	1.63	1.48	1.48	1.52	1.48	1.51	×	✓	×
st_lp(m)	0.28	0.12	0.24	0.13	0.30	0.28	0.27	0.24	0.38	✓	✓	✓
st_lp15	0.22	0.00	0.23	0.00	0.29	0.08	0.06	0.00	0.32	✓	✓	✓
st_lp30	0.28	0.00	0.01	0.00	0.19	0.00	0.00	0.00	0.11	✓	✓	✓
rr_st1(1/minute)	18.53	9.10	9.43	14.37	9.11	22.75	18.23	14.35	24.54	✓	×	✓
rr_st3(1/minute)	13.83	4.29	7.61	7.48	5.78	18.16	15.56	5.24	20.05	✓	×	✓
mn_tlc(s)	4.59	3.34	3.67	3.99	2.27	3.94	3.88	4.78	3.29	✓	×	×

		Visual		Visual/Manual						Effects		
Measure	BL	T1	T2	T3	T7	T4	T8	T5	T6	IVt	RLv	IVt*RLv
hi_st	0.11	0.20	0.12	0.20	0.23	0.13	0.13	0.21	0.12	✓	✓	✓
hi_st2	0.42	3.16	0.23	3.48	2.54	1.67	1.51	3.85	-0.48	✓	✗	✓
hwt_0_1	7.66	2.79	3.62	2.78	0.17	1.49	0.00	0.00	0.55	✓	✗	✗
hwt_1_2	41.66	22.44	26.57	30.27	3.88	21.65	10.59	11.16	4.60	✓	✗	✗
hwt_2_3	37.19	32.69	25.12	48.18	28.87	33.46	38.30	42.61	21.38	✗	✗	✗
hwt_3_4	10.74	27.78	22.60	13.21	36.29	22.95	20.62	16.94	26.82	✓	✗	✗
hwt_4_5	1.73	1.97	13.03	0.00	11.63	8.22	11.95	12.03	21.82	✓	✗	✗
hwt_5_6	0.72	9.55	5.94	5.56	2.53	5.01	7.11	6.15	7.36	✗	✗	✗
hwt_6	0.29	2.78	3.12	0.00	16.62	7.23	11.42	11.11	17.46	✓	✗	✗
lnx(%)	15.99	23.54			6.77		11.05			✗	✗	✗
rswt_5-10(1/minute)	35.98	17.34	33.79	16.69	18.75	38.17	37.53	14.26	38.18	✓	✓	✗
rswt_20(1/minute)	8.81	1.01	12.66	1.99	2.08	13.11	12.54	1.34	8.75	✓	✗	✗
rswt_40(1/minute)	0.80	1.72	0.30	0.77	0.59	0.45	0.40	1.00	0.75	✗	✓	✗
rswt_70(1/minute)	0.00	0.24	0.00	0.20	0.00	0.13	0.05	0.11	0.00	✗	✗	✗
tot_gd_t	-	0.93	9.22	1.40	6.08	4.44	17.84	11.73	6.30	✓	✗	✗
mn_gd	-	0.24	0.36	0.32	0.39	0.41	0.55	0.41	0.41	✓	✓	✗
n_gl	-	2.11	13.61	2.00	8.47	5.97	19.64	14.64	8.06	✓	✗	✗
tot_gl	-	1.22	11.94	1.79	7.90	5.80	23.16	15.25	8.19	✓	✗	✗
PRC	0.78	0.58	0.47	0.54	0.48	0.47	0.49	0.49	0.50	✓	✓	✗

7.9.2. System B – specific for Laboratory

		Visual		Visual/Manual						Effects		
Measure	BL	T1	T2	T3	T7	T4	T8	T5	T6	IVt	RLv	IVt*RLv
task_l	74.93	4.16	20.80	2.72	30.17	11.73	11.28	9.04	24.50			
subj_r	7.68	5.61	5.56	6.11	4.92	5.06	4.94	4.64	4.19	✓	✓	✓
mn_sp(km/h)	83.29	79.38	79.50	77.25	79.32	76.83	75.58	77.91	75.71	✓	✓	×
u_sp(km/h)	77.21	77.65	75.07	76.33	74.05	74.03	73.10	75.66	71.93	✓	✓	×
st_sp(km/h)	2.95	1.05	2.74	0.61	3.21	1.70	1.51	1.33	2.32	✓	×	×
st_sp15(km/h)	2.01	4.74	0.00	2.47	1.90	2.24	1.59	1.96	1.90	×	×	×
st_sp30(km/h)	2.45		2.52		3.36				3.09	×	×	×
d_sp(km/h)	-0.01	-0.01	0.03	0.00	0.02	-0.01	0.00	-0.01	-0.03	✓	×	×
mn_hwd(m)	44.39	53.13	61.90	50.23	73.14	56.30	70.09	63.24	79.76	✓	×	×
u_hwd(m)	30.55	49.71	49.70	47.07	54.59	45.94	58.23	56.76	57.80	✓	×	×
sd_hwd(m)	8.61	2.14	6.99	1.86	11.12	6.03	6.93	4.20	13.53	✓	×	×
sd_hwd15(m)	3.88	0.66	4.41	0.00	5.42	2.04	1.91	0.83	7.14	✓	×	×
sd_hwd30(m)	5.85	0.00	0.11	0.00	5.30	0.00	0.00	0.00	3.35	✓	×	×
mn_hwt(s)	1.94	2.51	2.90	2.44	3.47	2.84	3.59	3.02	4.02	✓	✓	×
u_hwt(s)	1.32	2.26	2.33	2.27	2.56	2.27	2.93	2.69	2.82	✓	✓	×
sd_hwt(s)	0.39	0.18	0.35	0.10	0.59	0.32	0.39	0.23	0.73	✓	✓	×
sd_hwt15	0.18	0.06	0.24	0.00	0.27	0.12	0.11	0.05	0.40	✓	×	×
sd_hwt30(s)	0.26	0.00	0.01	0.00	0.31	0.00	0.00	0.00	0.18	✓	×	×
mn_lp(m)	1.51	1.57	1.45	1.63	1.48	1.48	1.52	1.48	1.51	✓	✓	✓
st_lp(m)	0.30	0.22	0.33	0.15	0.34	0.36	0.30	0.36	0.43	✓	✓	×
st_lp15	0.27	0.04	0.28	0.00	0.27	0.13	0.10	0.07	0.33	✓	✓	✓
st_lp30	0.28	0.00	0.01	0.00	0.19	0.00	0.00	0.00	0.11	✓	✓	✓
rr_st1(1/minute)	12.02	17.63	4.04	7.25	4.21	19.54	11.98	23.77	0.62	✓	✓	✓
rr_st3(1/minute)	1.82	13.96	1.53	5.37	2.52	7.29	6.91	17.70	0.48	✓	✓	×
mn_tlc(s)	2.27	2.40	2.35	3.34	4.24	3.02	4.41	2.84	2.93	✓	×	×

		Visual		Visual/Manual						Effects		
Measure	BL	T1	T2	T3	T7	T4	T8	T5	T6	IVt	RLv	IVt*RLv
hi_st	0.22	0.07	0.27	0.17	0.21	0.24	0.22	0.10	0.26	✓	✓	✗
hi_st2	0.42	3.16	0.23	3.48	2.54	1.67	1.51	3.85	-0.48	✓	✓	✓
hwt_0_1	20.50	0.00	11.37	8.33	7.08	8.09	0.13	0.34	7.12	✓	✗	✗
hwt_1_2	43.22	44.44	23.14	42.46	26.00	31.88	30.28	37.80	15.97	✓	✓	✓
hwt_2_3	21.92	34.34	31.80	25.06	19.78	33.00	27.21	23.90	17.88	✓	✓	✓
hwt_3_4	7.00	13.90	13.50	10.88	11.83	15.64	13.99	13.85	15.98	✗	✓	✓
hwt_4_5	3.66	2.19	7.55	6.63	11.86	2.15	9.83	10.98	14.92	✗	✗	✗
hwt_5_6	2.19	0.38	3.56	0.96	8.70	1.05	6.28	3.43	13.24	✓	✗	✗
hwt_6	1.50	4.74	9.07	5.68	14.76	8.19	12.28	9.71	14.90	✗	✗	✗
lnx(%)		4.64	36.54		36.20	9.53		9.87		✓	✗	✗
rswt_5-10(1/minute)	10.15	27.91	25.19	51.02	49.60	22.75	20.58	38.99	24.94	✓	✓	✗
rswt_20(1/minute)	0.19	1.73	2.97	18.92	19.34	2.08	2.56	7.20	3.69	✓	✗	✗
rswt_40(1/minute)	1.78	1.47	1.44	0.15	1.45	0.75	0.79	0.57	1.32	✗	✗	✗
rswt_70(1/minute)	0.04	0.36	0.00	0.00	0.00	0.06	0.04	0.00	0.20	✗	✗	✗

8. The Leeds field experiment

8.1. Test site

Participants drove the Leeds Instrumented vehicle, a Ford Mondeo 2.0 litre petrol hatchback. The car is equipped with front and rear facing rangefinder sensors (radar), and provides lateral position information (operational when vehicle above 30mph), using SafeTRAC. Speed related information as well as clutch and throttle position were collected from the vehicle's CAN (Computer Aided Network), while head and eye movement position were logged using Seeing Machines FaceLAB version 3.0 (Figure 118).

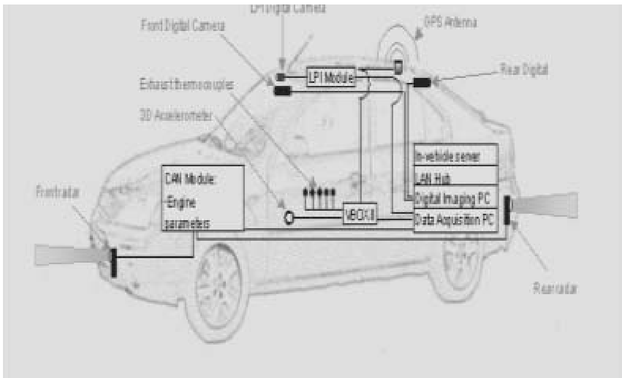


Figure 118 – The institute for transport studies' instrumented car

8.2. Scenarios and participants

The same 18 drivers used in the Leeds simulator and LabSim experiments were recruited for this experiment.

8.3. Road category

Data collection took place in one section of motorway and one section of rural road.

8.4. IVIS included

Each participant in the experiment performed 8 tasks from System B. One of the tasks used in the simulator and LabSim (task 6) was considered too time-consuming and therefore eliminated from the field experiments.

8.5. Experimental design and Procedure

Two experimenters accompanied each participant throughout the drive. One experimenter sat in the front passenger seat and provided driving directions and task instructions to the driver, whilst the second experimenter recorded observational data from the back seat, using the Wiener Fahrprobe protocol (this information is not reported here).

Upon arriving at the test site, all participants were briefed on the experimental procedure and agreed to sign a consent form. After undergoing a calibration procedure for FaceLAB, all participants were given sufficient opportunity to practice all tasks on system B, after which data collection for the static version of the task took place. Participants were then given sufficient opportunity to drive the instrumented vehicle and practice completing the tasks whilst driving, before the commencement of data collection. Baseline data collection always started in the rural road followed by the motorway. Driving and task performance on system B was then completed on the motorway section followed by the rural road. The entire experiment took approximately 2 hours per participant.

8.6. Measures and analysis method

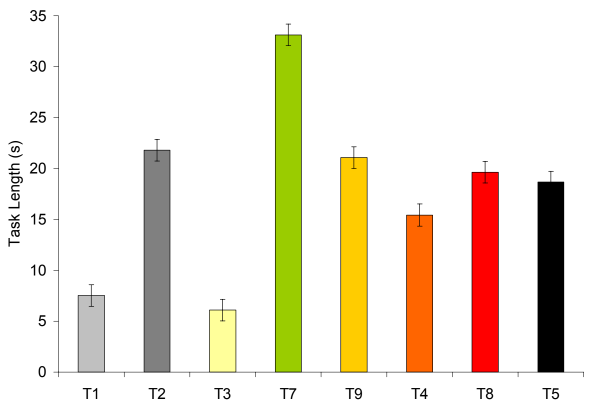
The speed and steering related mandatory measures for field were computed for this experiment. Lateral position data were also recorded using SafeTRAC, although due to uncertainties about its accuracy, this data is omitted from this report. The optional eye movement related measures were also recorded and are outlined below.

8.7. Results

The order in which all task related results are presented in the graphs is made according to an a-priori assumption, where BL is considered to be easiest and Task 5 the most difficult.

8.7.1. Task length

Task length varied between 6.08 and 33.14 seconds (means). Task completion time was shortest for Task 3 (open and close navigation programme) and longest for Task 7 (scale change). Univariate analyses of variance showed main effects of task type and road level on task completion time, with longer completion times seen during the motorway drive (mean task completion time: 18.5 seconds, versus 16.3 seconds during the rural drive). No interaction was found between road and task level, suggesting that the particular road type affected task completion time in the same way (see Figure 119). As can be seen from the graphs, the a priori categorisation method did not apply to task length; because for instance participants took longer to complete tasks 2 and 7 than first anticipated.



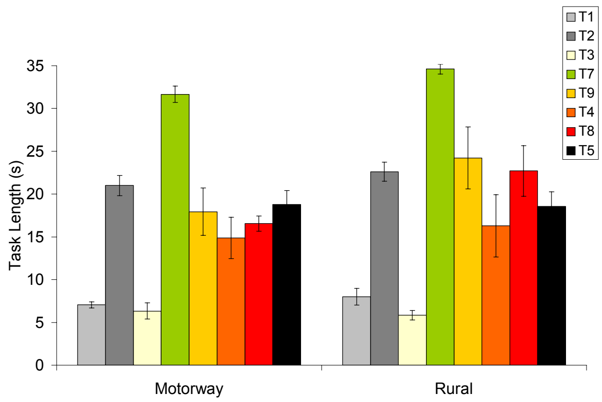


Figure 119 – Task length overall (left) and for each road category

8.7.2. Self-reported driving performance

Univariate analyses showed a main effect of task type and road level on subjective rating of driving performance. Participants thought their baseline driving performance was better than when driving was done with any of the system B tasks. Significant differences for subjective rating were found between many of the tasks, with driving performance was rated best during Task 3 (open and close navigation system), probably related to the fact that this task had the shortest duration. Participants felt that task completion had a more deleterious effect on their rural road driving, compared to the motorway, and an absence of an interaction meant that the effect of each task on driving performance was thought equal for both road categories (see Figure 120).

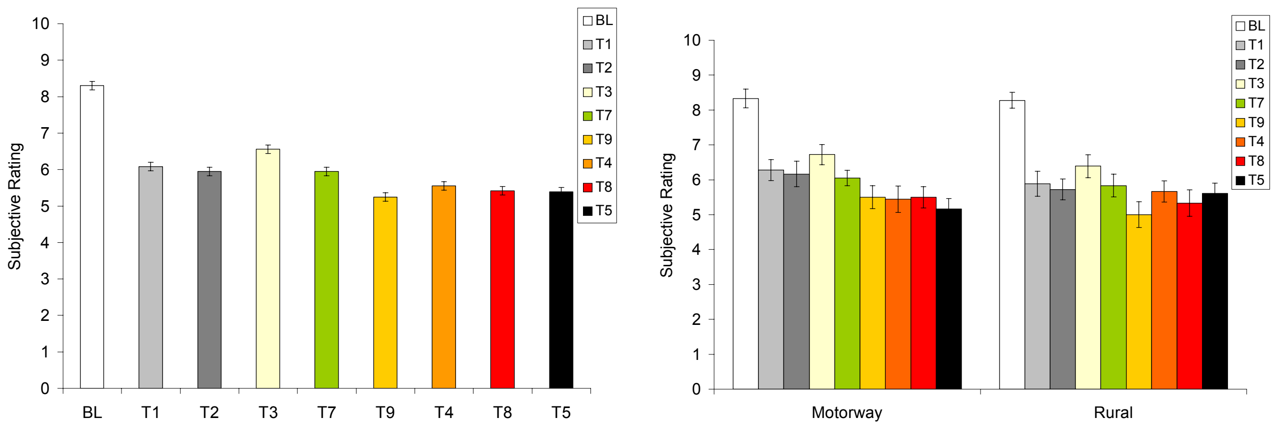


Figure 120 – Subjective rating of driving performance, overall (left) and by road category (right).

8.7.3. Longitudinal control

As expected, mean speed in the motorway was significantly higher than in the rural road, but an absence of an interaction between task type and road level suggests that the various tasks did not have differential effects on mean speed across the two road environments. Mean speed did not necessarily drop from baseline with the onset of a task, and was indeed seen to increase in the motorway during completion of task 5 (set destination). In the rural road, only task 9 (create a waypoint) was found to cause a significant reduction in mean speed compared to baseline. Overall, minimum speed was found to fall with the introduction of tasks, although the exception to this was again the effect of task 5 in the motorway (Figure121).

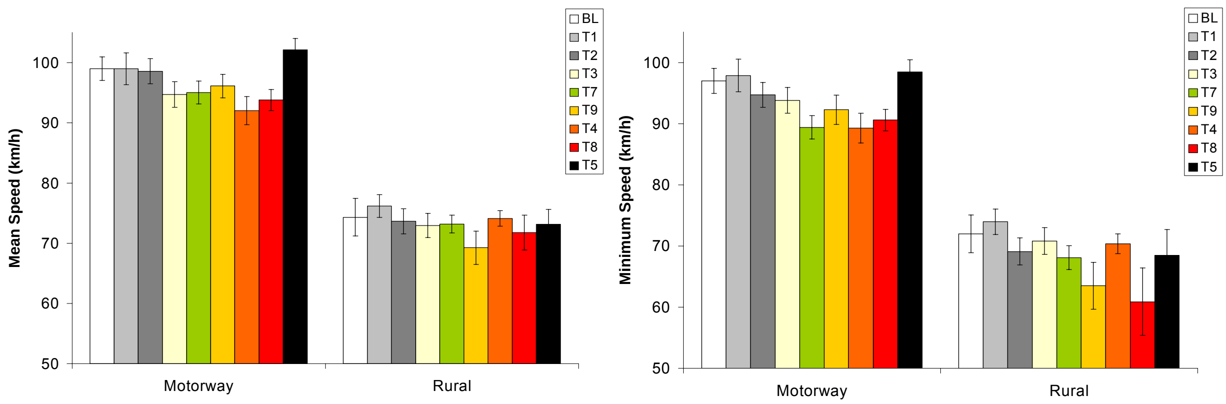


Figure121 – Mean (left) and minimum (right) speed of travel for each road category

The effect of tasks on speed change was examined further and results showed that on 63% of occasions, introduction of a task resulted in a minor (less than 5%) speed change. When speed change was greater than 5%, drivers were more likely to increase their speed with the introduction of tasks than decrease their speed (21% versus 15% of the time).

Unfortunately, due to problems with the radar of the instrumented vehicle, measures of time and distance headway were found to be incomplete, with only 54% of the data available for analysis. Mean and minimum time headway to the vehicle in front were always shown to increase from baseline with the introduction of a task. Average time headway for the baseline drive was just under 2s for the motorway and just over 2.5 seconds for rural road driving. These baseline headways were found to increase to an average of 3 seconds for the motorway and up to 5 seconds for the ‘change settings’ task (T8) in the rural road (Figure 122). Overall, participants maintained longer time headways on the rural road than on the motorway, even during the baseline drives. This was also the case when the more ‘extreme’ data points for tasks 1 and 4 were eliminated from analyses. Such results are to be expected, considering difference in speed between the two road levels.

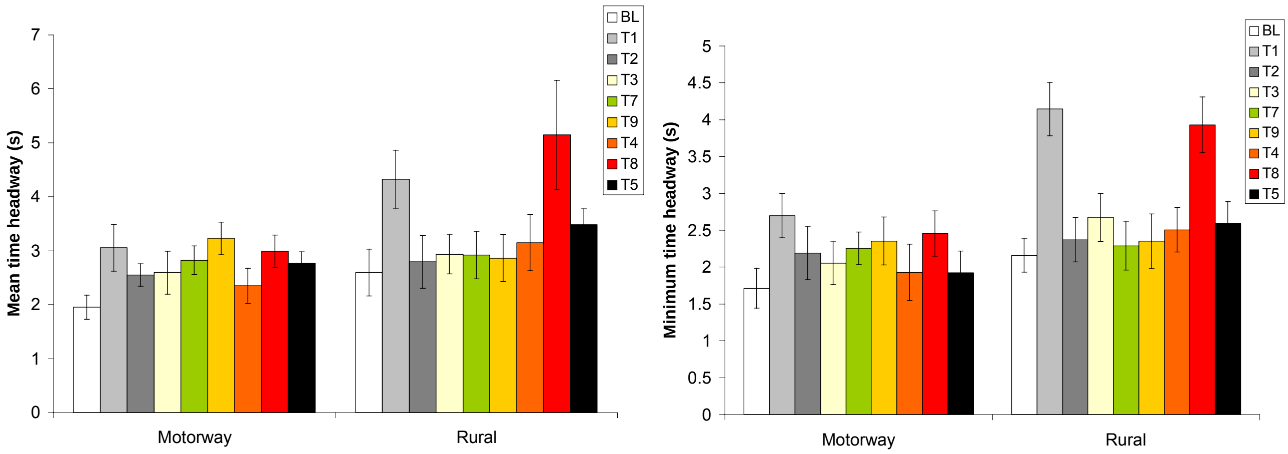


Figure 122 – Mean and minimum time headway by task type, for motorway and rural road

8.7.4. Lateral control

As outlined above, due to uncertainties with the accuracy of data gathered from SafeTRAC, lateral position data will not be reported. The steering wheel reversal rate showed a general increase from baseline for both roads, when a task was introduced during driving. However, caution must be taken when interpreting this data, since reversal rate is related to task length. For instance, reversal rate for T3 is quite low, as this task only lasted around 5-6 seconds Figure 123). The longer lasting tasks 2 and 7 both resulted in the highest number of reversal rates (1^o) in the motorway and rural road. The number of reversal rates per minute was higher in the rural road than in the motorway, which is expected considering the nature of driving in the rural road.

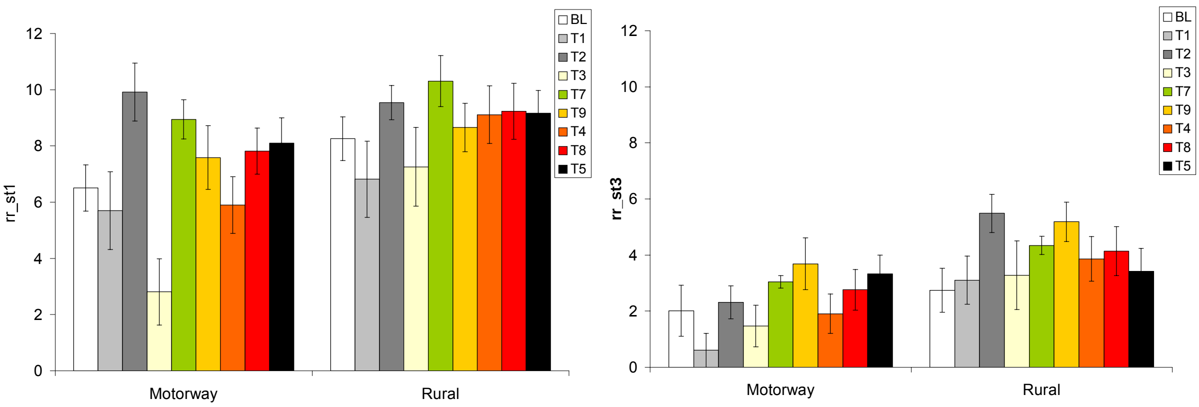


Figure 123 – Steering wheel reversal rate with thresholds of 1^o (left) and 3^o (right)

8.7.5. Eye movements

Eye movement measures were only collected successfully for 16 of the 18 subjects in the field. Furthermore, conditions in the field meant that a large amount of eye movement data was lost during data collection. Measures of percent road centre and number of glances to the system must therefore be treated with caution.

During baseline motorway driving, participants were seen to direct their visual attention to the IVIS system for a large proportion of the drive, as shown by number of glances, which consequently resulted in a drop in glances towards the centre of the road (Figure 124). Analyses showed an effect of road on percent road centre; with many more glances towards the centre of the road during rural driving. Number of glances and percent road centre were not found to be significantly different across the different task types for either road level.

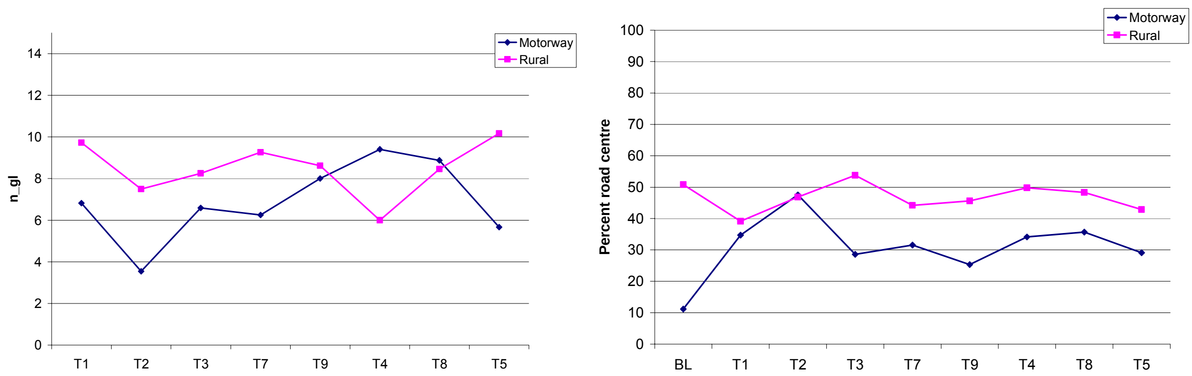


Figure 124 – Number of glances to system B (left), percent road centre (right).

8.8. Summary and Conclusions

The a priori grouping of tasks by difficulty level was found to be mostly accurate upon when compared to the subjective ratings of driving performance.

Task completion time and the a priori grouping of tasks were not found to be well correlated. In particular, the completion of T3 (ranked 3rd) was found to take the least time, whilst T7 (ranked 4th) was found to take the longest to complete. The a priori classification of tasks was mainly based on the ease at which tasks were completed, and not necessarily the method by which subjects attempted the tasks. Observation of task completion during the experiments demonstrated that subjects had a tendency to 'chunk' their visual attention and action between the system and the road, which therefore resulted in longer completion times for some of the tasks e.g. task 7.

Although longitudinal measures such as mean speed and time headway changed from baseline with the introduction of tasks, there was no clear relationship between these measures and subjective rating of driving, or the a priori ranking technique. The effect of task type on steering related measures was found to be related to task length, with longer tasks causing more reversals. Due to the short length of tasks, the sliding window technique did not produce conclusive results.

Eye movement data were able to provide information on the presence of tasks, but showed no differences by task or road type.

8.9. Measures summary tables

8.9.1. System B

Significant differences from baseline (BL) are indicated by grey background.

	BL	T1	T2	T3	T7	T9	T4	T8	T5	IVT	RLV	IVT*RLV
subj_r	8.31	6.08	5.94	6.56	5.94	5.25	5.56	5.42	5.39	✓	✓	×
task_l	13.25	7.53	21.81	6.08	33.14	21.08	15.41	19.64	18.67	✓	✓	×
mn_sp	86.65	87.58	86.11	83.82	84.11	82.69	83.07	82.77	87.62	✓	✓	×
u_sp	84.48	85.91	81.89	82.30	78.74	77.88	79.80	75.72	83.46	✓	✓	×
st_sp	1.07	0.99	2.59	0.93	2.99	2.86	2.01	3.48	2.26	✓	✓	×
st_sp15	0.07	0.02	0.35	0.00	0.55	0.23	0.05	0.30	0.27	✓	×	×
st_sp30	0.00	0.00	0.02	0.00	0.17	0.06	0.01	0.06	0.14	×	×	×
d_sp	-0.12	0.77	1.40	1.02	1.18	2.06	1.00	1.44	1.54	×	×	×
mn_hwd	50.15	84.99	62.10	62.19	65.92	69.13	61.92	78.73	71.07	×	×	×
sd_hwd	5.01	4.98	5.43	3.49	11.68	9.53	7.37	7.09	8.54	×	×	×
u_hwd	42.84	78.78	52.93	52.17	52.30	53.31	49.21	66.10	51.99	✓	×	×
mn_hwt	2.28	3.69	2.67	2.76	2.87	3.05	2.75	4.07	3.12	✓	✓	✓
sd_hwt	0.32	0.84	0.62	0.58	0.65	0.97	0.63	1.15	0.99	×	✓	×
u_hwt	1.94	3.42	2.28	2.36	2.27	2.35	2.22	3.19	2.26	✓	✓	×
rr_st1	7.38	6.26	9.73	5.03	9.63	8.12	7.51	8.52	8.63	✓	✓	×
rr_st3	2.38	1.85	3.90	2.38	3.69	4.44	2.89	3.45	3.37	✓	✓	×
hi_st	0.36	0.49	0.37	0.53	0.41	0.40	0.46	0.44	0.42	✓	✓	×
mn_gd		0.88	0.72	0.51	0.94	0.52	0.70	0.56	0.70	×	×	×
n_gl		8.27	5.52	7.42	7.76	8.31	7.70	8.66	7.92	✓	×	✓
prc	30.99	36.91	47.18	41.18	37.91	35.47	41.94	41.97	36.00	×	✓	×

9. The Volvo Technology field experiment

9.1. Test site

The experiment was conducted on the motorway E6 outside Gothenburg in Sweden. The participants drove an instrumented Volvo V70, which was equipped with the SafeTRAC lane tracker from Assist Ware Technology, the steering wheel angle sensor LWS 3.1 and the Seeing Machines FaceLAB 3.0 head and eye tracking system (see Figure 125). Speed was logged via the serial bus system Controller Area Network (CAN). PDT data were logged via a separate unit.

Figure 125 – Interior of vehicle with FaceLAB cameras on top of the dashboard (*) and exterior of the instrumented vehicle

9.2. Scenarios and participants

16 drivers in total participated in the experiments (10 women and 6 men). Data from the lane tracker was only analysed for 15 participants due to sensor failure. The average age for the participants was 34 years (range: 25-54) and the average time in which they had held their driving license was 14.19 years (range: 6-31). None of the drivers were or had been professional drivers. Fourteen out of the 16 participants drove more than 10.000 km/year.

All participants drove on a motorway that had two lanes in each direction. The mean width for each lane was 3.8 metres. The test runs were scheduled to avoid rush hours, although dense traffic situations were rather difficult to avoid. The speed limit was 110 km/h.

9.3. IVIS included

Each participant in the experiment performed tasks from System A and System B (for task descriptions see section 2.2). Two out of the nine tasks for System A and one out of the nine tasks for System B were removed because they were considered to be too demanding in real traffic.

9.4. Experimental design

Each participant performed the tasks for both Systems A and B. However, there were no statistical comparisons between the two systems. Task difficulty level was a within subject factor. Three baseline sections were collected in the beginning, middle and end of the experimental drive, in order to try to remove any possible learning effects. The tasks were performed by the driver in between the baseline sections. Half of the group of participants first drove with System A and then continued with System B. The other group did this in the reverse order. The order of the tasks was randomised between all participants. The PDT (peripheral detection task) was run as a separate condition, always after the no PDT condition, in order to be able to compare results with experiments from other partners.

For further details on the general experimental design see section 2.4.

9.5. Procedure

Two observers accompanied the participant throughout the drive. The participant was first provided with instructions by one of the observers, while the other observer prepared the FaceLAB system and the logging equipment. Each participant was requested to practice the tasks in static mode, as well as practice driving the vehicle. The entire experiment took approximately 2.5 hours per participant.

9.6. Measures and analysis method

The mandatory measures for field were computed (see section 2.7 for an overview). Three optional steering, and six lane keeping measures were also included, along with five eye movement related measures and the PDT measures.

9.7. Results

9.7.1. Effects of System A

9.7.1.1. Task length

Task lengths varied between 11.4 and 39.5 seconds (see Table 21) where the baseline was somewhat longer than 90 seconds (mean 101.4 seconds). The longest task (T9), was followed by tasks 4 and 8. Task 1, 2, 3 and 4 were auditory/cognitive and tasks 7, 8 and 9 were visual/manual.

Table 21 – Task length for system A

Task	Task Length (s)
T1	11.4
T2	15.0
T3	13.9
T4	20.4
T7	11.4
T8	18.4
T9	39.5

9.7.1.2. Self-reported driving performance

The order in which the results are presented in the graphs is made according to an a priori assumption, where BL is considered to be easiest and Task 9 the most difficult. According to the self-reported driving performance, the a priori assumption is more or less correct (see Figure 126).

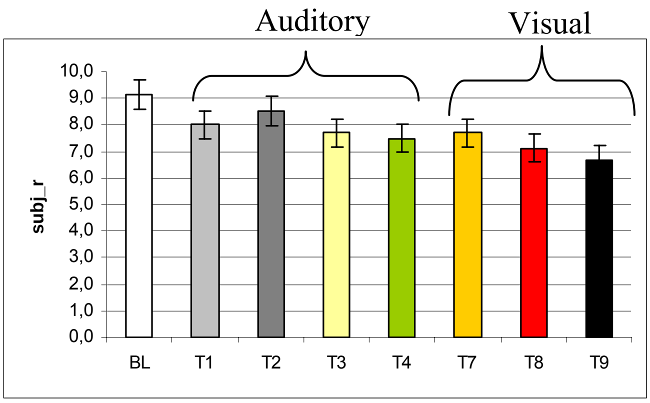


Figure 126 – Subjective rating of driving performance

The ratings were significantly lower for driving with tasks, compared to baseline. Also, all the tasks were rated significantly higher compared to task 9 which might also be due to the fact that task 9 took a very long time to complete.

9.7.1.3. Longitudinal control

Compared to baseline, mean speed was not significantly reduced when participants performed the tasks. Standard deviation of speed (st_sp) shows a significant difference for task 9. However, the corrected measures (st_sp15, st_sp30) show no difference and thus the effect in st_sp is most likely due to the bias of task length in the variation measure.

Minimum speed (u_sp) was higher when performing a task compared to baseline. This was most likely due to the bias of the heavy traffic on the road, where participants were trapped in dense areas of traffic more often, and were forced to adjust their speed. This was more likely for the baseline condition as it lasted longer.

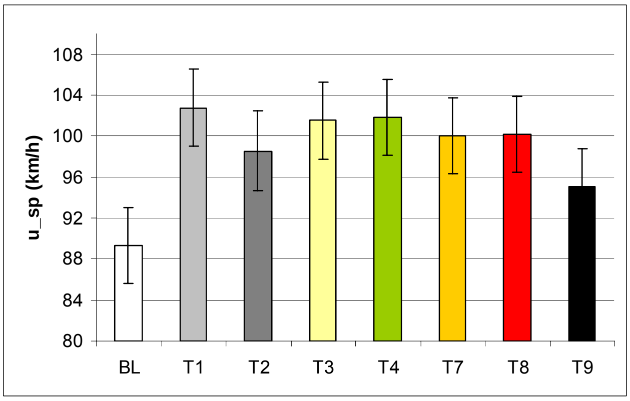


Figure 127 – Minimum speed

9.7.1.4. Lateral control

Results of the high frequency component of steering showed that the visual tasks induced higher values than the auditory tasks (see Figure 128).

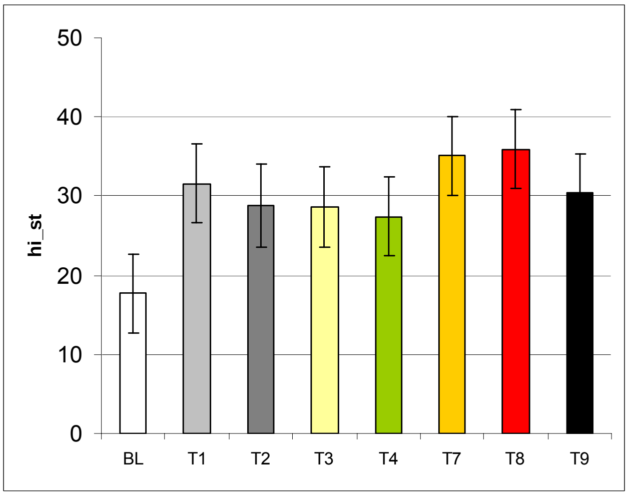


Figure 128 – High frequency component of steering

None of the measures related to time-to-line crossing were sensitive to task versus baseline. For the pr_tlc1 and 2 there were far too few values to rely on. One suggestion might be that future experiments use a different threshold for field studies compared to simulator experiments. Mn_tlc1 and 2 had a large variation within the group. It is likely that the measures related to lane position (as well as speed) were affected by traffic density. Based on WP2 results, we would expect that, as the complexity of tasks increased, there would be a reduction in the mean time-to-line-crossing value, i.e. the driver would drive closer to the line. However, in this particular experiment there was a lot of traffic in the second lane, overtaking the participants, causing our drivers to drive more to the right of the lane.

Standard deviation of lane position (st_lp, st_lp15) was sensitive to the introduction of system A tasks, showing a significant difference from baseline. However, this measure could not distinguish between different tasks (see Figure 129).

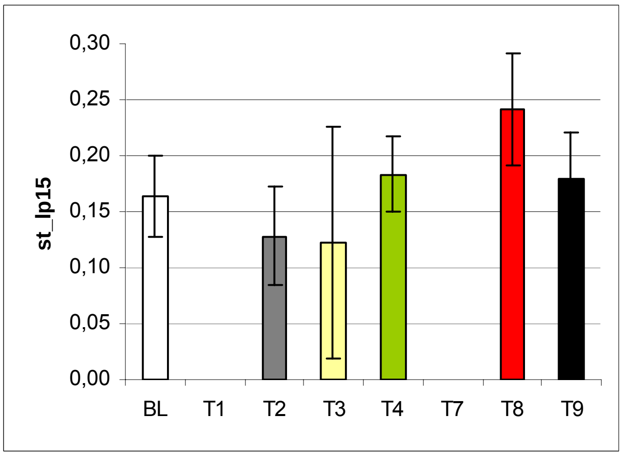


Figure 129 – Standard deviation of lane position with a 15 second long time window

Reversal rate, with a 1° threshold was sensitive to task versus baseline driving. All mean values for visual tasks (T7-9) were higher than baseline, whilst the effect of auditory tasks (T1, T3 and T4) seemed to be similar to baseline or even smaller reversals rate than baseline.

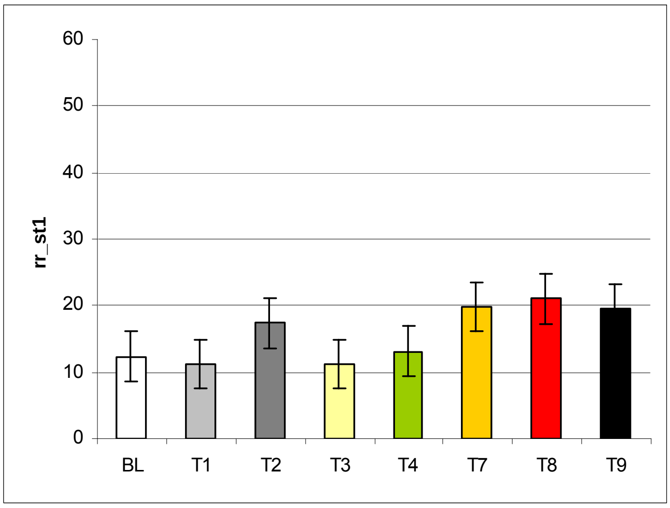


Figure 130 – Reversal rate, 1° threshold

The number of rapid steering wheel movements increased when 5° and 10° were used as a threshold, especially for the visual tasks (see Figure 131).

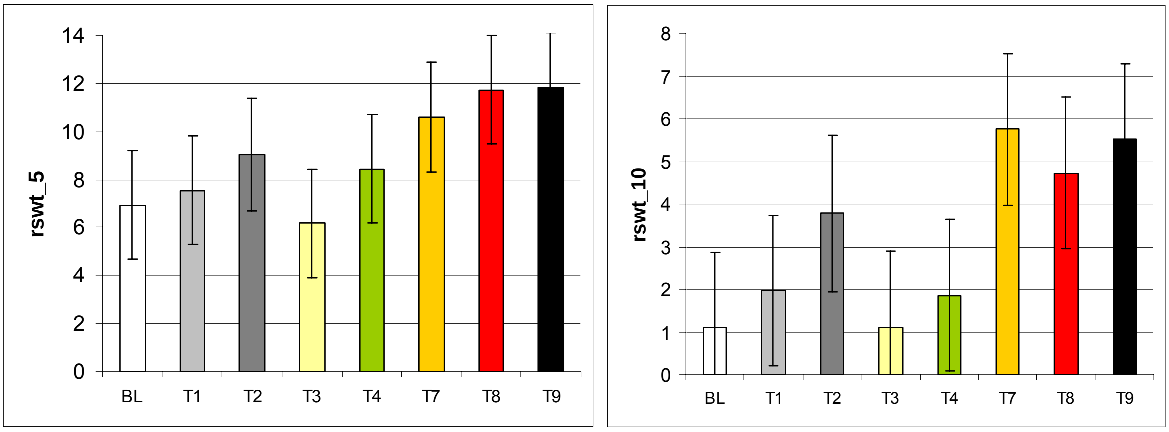


Figure 131 – Number of rapid steering wheel movements (deg/s), using a threshold of 5° (left) and 10° (right)

9.7.1.5. Eye movements

Often, measures related to eye movements are divided into glance-based and non-glance based, with both groups being the outcome of ocular segmentation where for instance fixations and saccades are identified. From fixations and saccades one can then define glances. Non-glance measures are measures where the glance distribution between two areas of interest is not calculated but rather the measures are derived from fixations and saccades.

Glance measures (n_gl, tot_gl, mn_gd) were only calculated for the visual tasks since they are not applicable to baseline and the auditory tasks.

The non-glance measure, standard deviation of gaze angle (st_ga), was only calculated for the auditory/cognitive tasks and baseline while Percent Road Centre (PRC) was calculated for all tasks and baseline. St_ga could not be calculated for the visual tasks, since this measure can be biased by the distance to the IVIS display. Percent road centre is not dependent on the distance to the display and thus works on visual tasks and baseline. This is because this measure determines whether fixations are within a predefined area.

All three glance measures showed an increase in visual demand for tasks 7, 8 and 9. This order fits well to the subjective rating of driving performance, as well as task length. The mean duration of single glances to the IVIS was shown to increase. Results also showed an increase in glance duration as a percentage of total task duration (see Figure 132). Thus, along with the more difficult tasks (or longer tasks) the participants seemed to have longer glance time on the IVIS, as well as spending a higher proportion of their time on the IVIS rather than on the road ahead.

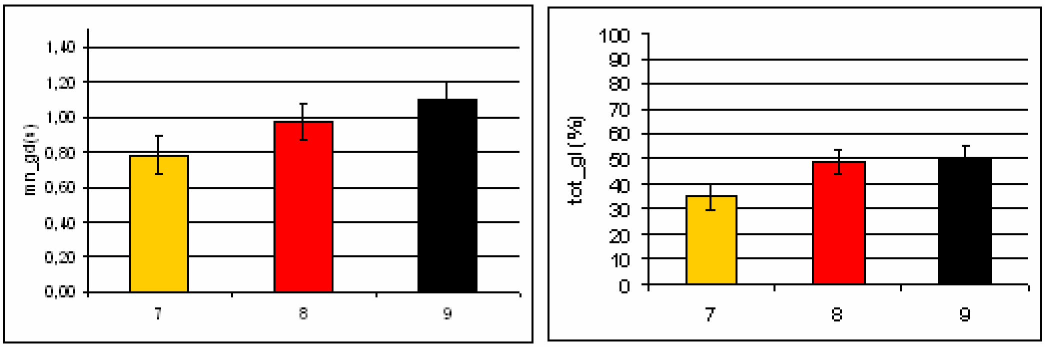


Figure 132 – Mean duration of single glances to IVIS (left) and IVIS glance duration as a percentage of total task duration (right).

The increasing glance frequency to IVIS (see Figure 133) is likely to be related to task length, since for example T9 took nearly 40s to complete.

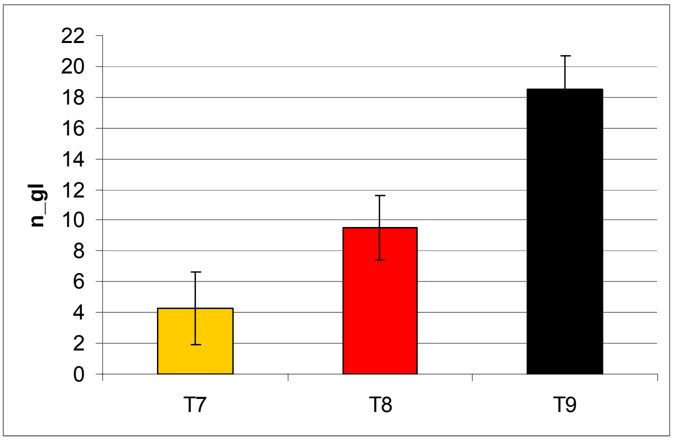


Figure 133 – Number of glances to IVIS

The mean total glance duration values ranged between 3.87 seconds for T7 and up to 20.39 seconds for T9 (Figure 134). This is interesting, since for instance guidelines from JAMA (JAMA, 2004) state that the operation task of a display monitor is prohibited if the task exceeds total glance durations of 8 seconds.

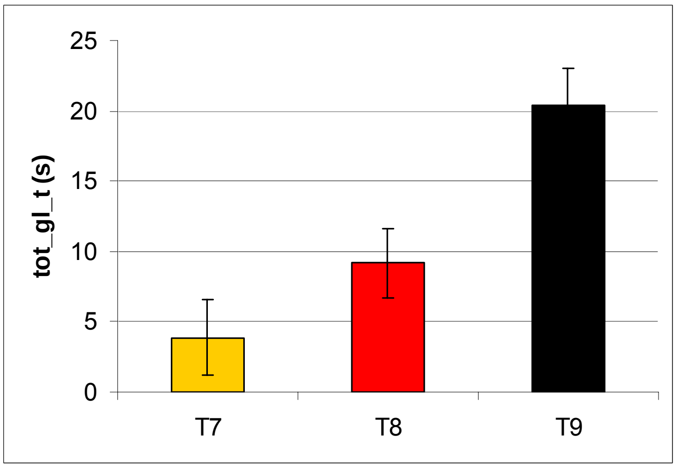


Figure 134 – Total glance duration

Percent road centre, a non-glance based measure that can be applied to all tasks irrespective of modality, showed a clear distinction between the cognitive/auditory and the visual/manual tasks. As can be seen in Figure 135, the gaze concentration on road ahead for cognitive tasks is higher compared to baseline (with a clear gradual change in PRC for T1, T2 and T3) while the concentration on road ahead for visual tasks is lower, compared to baseline.

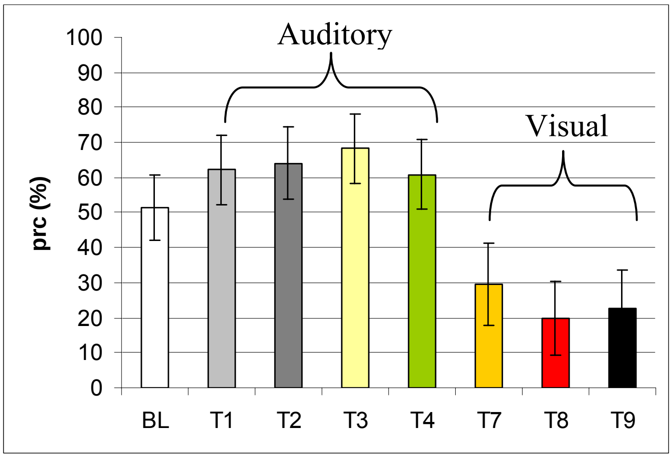


Figure 135 – Percent road centre results for each task in system A

Therefore, PRC as a measure seems to be robust to low data quality. This measure was also shown to work well in the HASTE WP2 experiments, when calculated on data where different fixation algorithms had been used. One remark is that the PRC is very low for baseline compared to previous experiments (e.g. in our simulator based WP2 experiments). One explanation for this could be that, in the field, the cameras had a small off set to the right in order to track participants' faces when they leaned over the right during task completion. This was especially true for System B, which had a smaller screen and since participants had to use a pen as an input device, they were forced to lean more forward towards the screen). This was also true for baseline driving, where participants tended to be on the verge of the cameras' field of view.

9.7.1.6. Event detection

To examine event detection, the PDT (Peripheral Detection Task) was included in this experiment. The percentage of 'hits' in PDT were shown to fall for tasks, in accordance with the a priori task order (ranging from 93% for baseline and 64.5% for T9). The pattern of reaction times was not as clear, however (Figure 136). The spatial cognitive tasks (T1, T2) seem to have a somewhat smaller effect on the PDT reaction time, compared to the arithmetic cognitive tasks (T3, T4). The PDT values for reaction times and hit rate in this experiment were found to be similar to those in previously published work (e.g. Olsson & Burns, 2000).

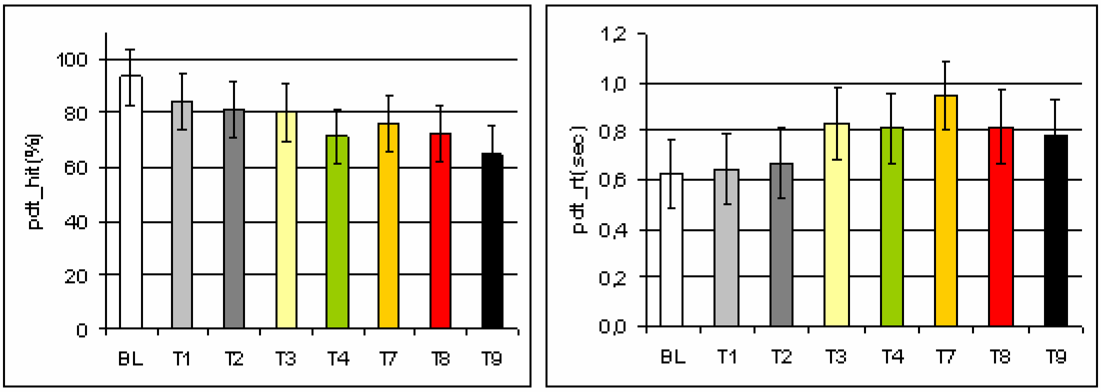


Figure 136 – Percent correct responses or hits (left) and reaction times for correct responses (right)

9.7.1.7. IVIS performance

All participants completed all tasks for System A. However, tasks 5 and 6 were not included at all in the actual experiment because they were considered to be too difficult to perform in field.

9.7.2. Effects of System B

9.7.2.1. Task Length

Mean task lengths varied between 7.2 and 26.7 seconds (Table 22) where the baseline was somewhat longer than 90 seconds. Tasks 2 and 8 had the longest task durations, while T1 and T3 were rather short. Task 1 and 2 were visual tasks only while tasks 3-9 had both visual and manual components. See section 2.2 for further description of these tasks.

Table 22 – Task length in seconds for system B

Task	Task Length (s)
T1	7.2
T2	26.7
T3	11.7
T4	23.7
T7	22.9
T8	22.9
T9	25.9

9.7.2.2. Self-reported driving performance

The subjective rating of driving performance did not correspond fully, to the a priori ordering of the tasks (see Figure 137). Task 7 was ordered as medium difficulty, since manual input was needed and the number of button presses was rather high. However, the task consisted of zooming in and out of the system and was very repetitive, which may have made the task easier than first anticipated.

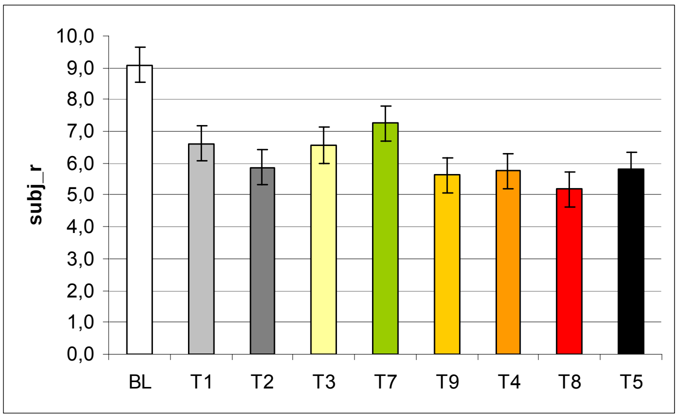


Figure 137 – Subjective rating of driving performance

9.7.2.3. Longitudinal control

Similar to the results for System A, no effect of mean speed was found, although there seems to be a small reduction in mean speed when performing tasks, compared to baseline (see Figure 138). A main effect of standard deviation of speed was found. However, this is most likely an effect of the large difference between baseline length and task length and thus no effects were shown in the corrected measures (st_sp15, st_sp30). Speed measures were very much affected by the high traffic density.

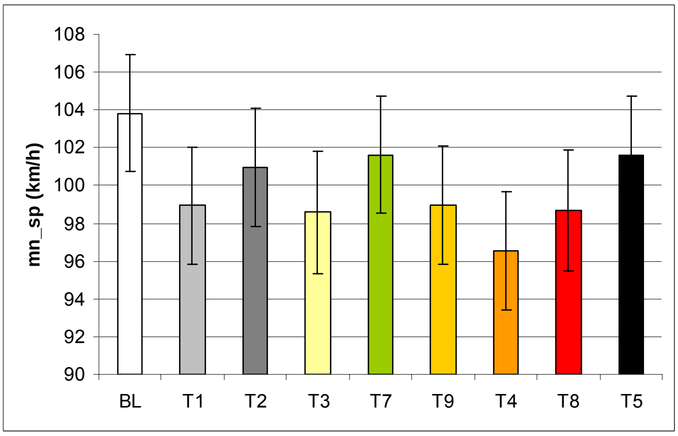


Figure 138 – Mean speed

9.7.2.4. *Lateral control*

Reversal rate, with both 1° and 3° as the threshold, was to some extent sensitive to task vs. baseline (Figure 139). However, not all mean values for each task were significantly higher than baseline.

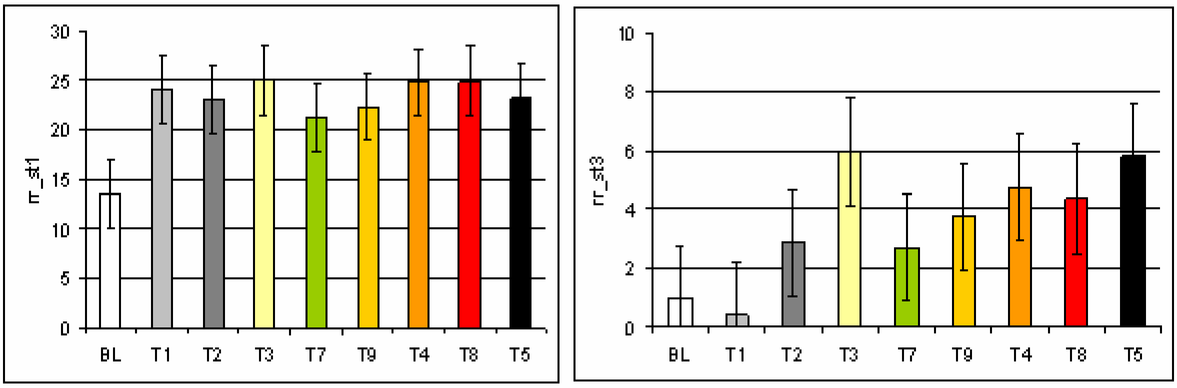


Figure 139 – Reversal rate where 1 (left) and 3 (right) is the amplitude threshold for reversals

High frequency of steering was shown to increase significantly in the presence of all system B tasks (Figure 140).

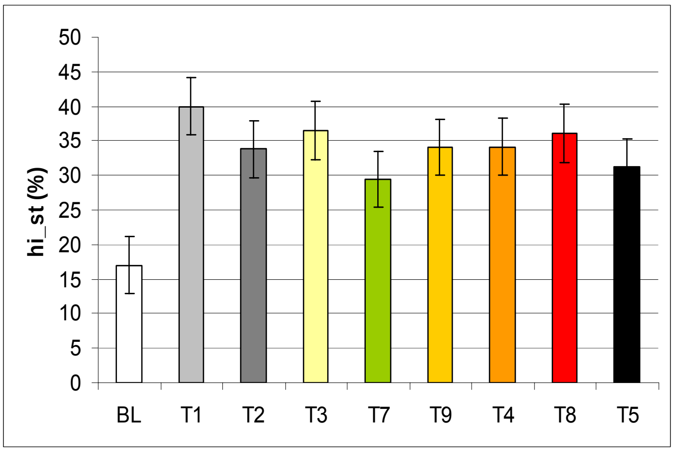


Figure 140 – Hi_st - high frequency component of steering

The number of rapid steering wheel movements was found to be higher when a task was performed, compared to baseline (see Figure 141). However, no trend for task complexity seems to be present. Main effects of task were present for rswt_5 and rswt_10, and Tasks 1 and 3 produced very high values, which again might be due to a bias from task length.

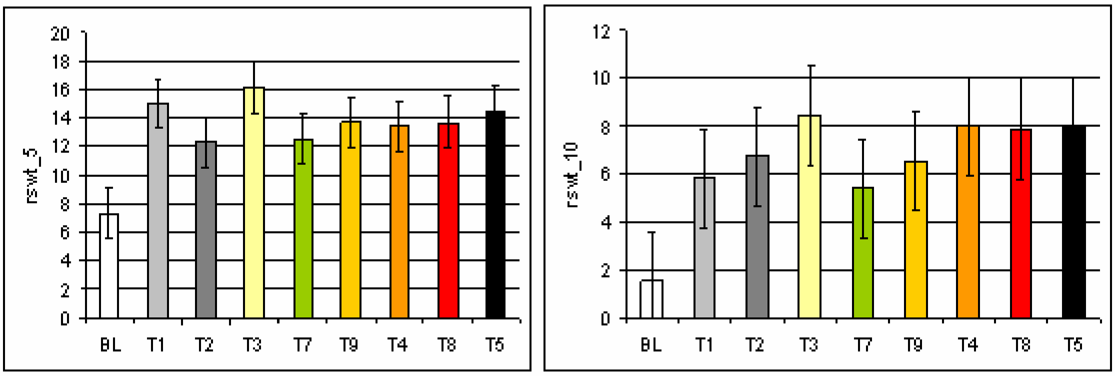


Figure 141 – Number of rapid steering wheel movements (deg/s) using a threshold of 5 (left) and 10 (right).

None of the Time-to-Line-Crossing measures showed main effects. However, this could be due to rather low power in the experiment and the bias of high traffic density. It could be argued that these measures are less applicable in field experiments or might need to be altered, for instance, by using different thresholds.

Lane position measures become problematic since they were very much influenced by the high traffic density and overtaking vehicles.

9.7.2.5. Eye movement

All tasks had a significant lower PRC value compared to baseline. However, no clear effect of task difficulty seems to be present (see Figure 142). The values are much lower in general compared to previous experiments and the reason is the same as discussed in section 9.7.1.5 for System A.

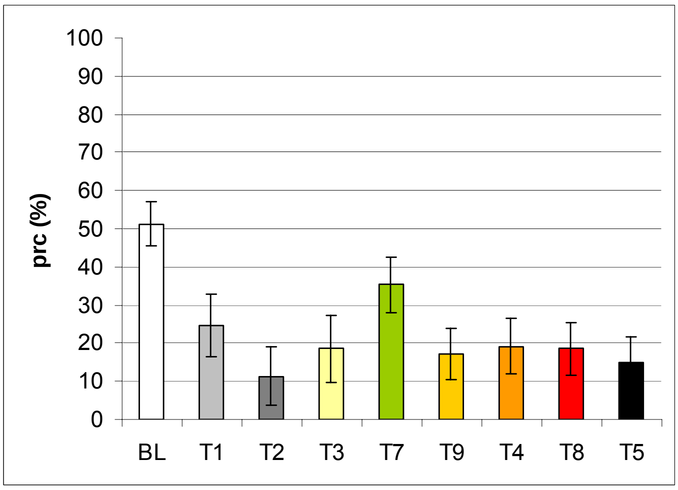


Figure 142 – Percent road centre for each task of system B

The a priori task order does not correspond to the PRC, which is not unusual since the visual measures are often affected by task length. For instance, the lowest PRC values are seen for the pure reading task (T2), even though this task was quite simple (both with regard to a priori ordering as well as its effects on subjective ratings).

Mean duration of single glances to the IVIS seems to indicate an increase with task difficulty. Task 2 is a pure reading task, which might explain the high values (see Figure 143) compared to the tasks included a combination of visual demand and complex manual component.

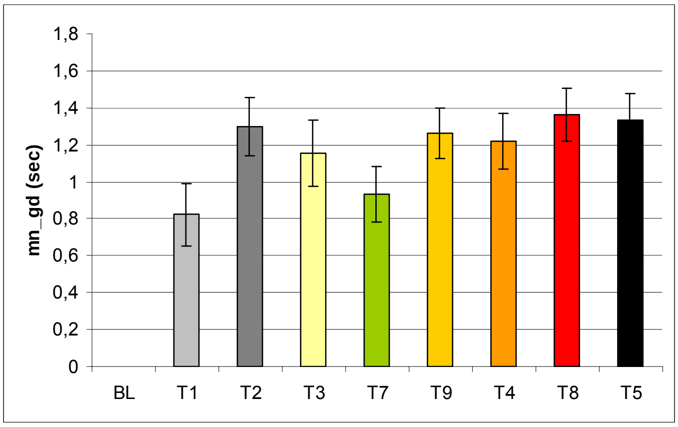


Figure 143 – Mean duration of single glances to IVIS

Glance frequency was found to be sensitive to different task difficulty levels (see Figure 144), and naturally correspond well to task length. Therefore, the reading task (T2) again scores very high.

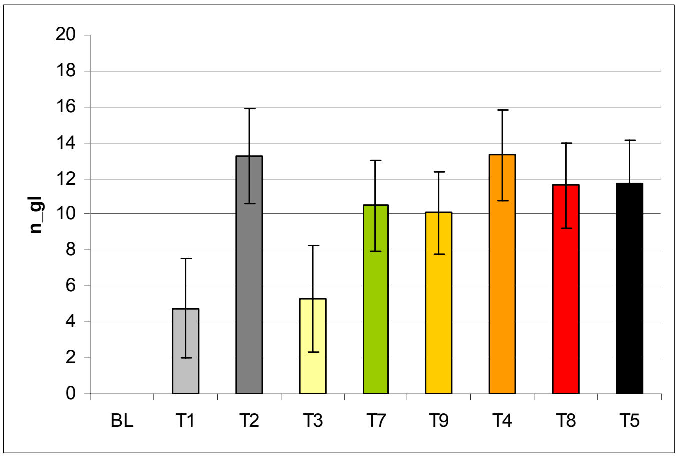


Figure 144 – Number of glances to IVIS to complete task

Glance duration to IVIS as a percentage of total task duration shows similar effects to mn_gd and n_gl (see Figure 145). However, since this is a percentage measure, the short tasks (T1 and T3) might be biased by task length, thus explaining their high values. Glance duration for Task 7 is quite low despite requiring a large number of button presses. This may have been because the buttons were identical to each other (zooming in and out of the map), making Task 7 quite simple.

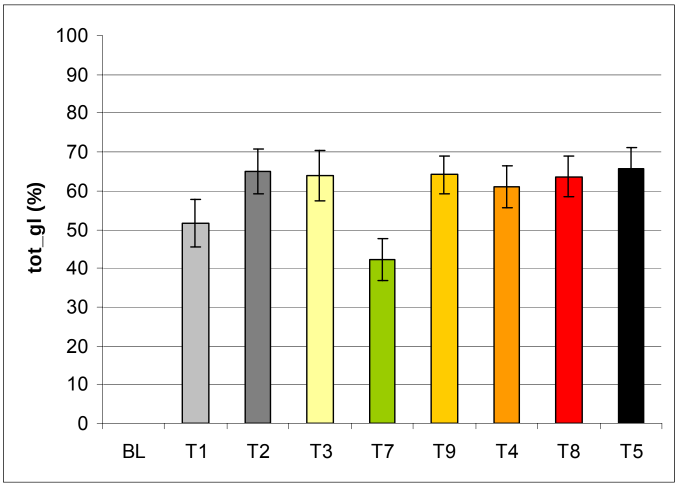


Figure 145 – IVIS glance duration as a percentage of total task duration

Total glance duration is presented below (see Figure 146).

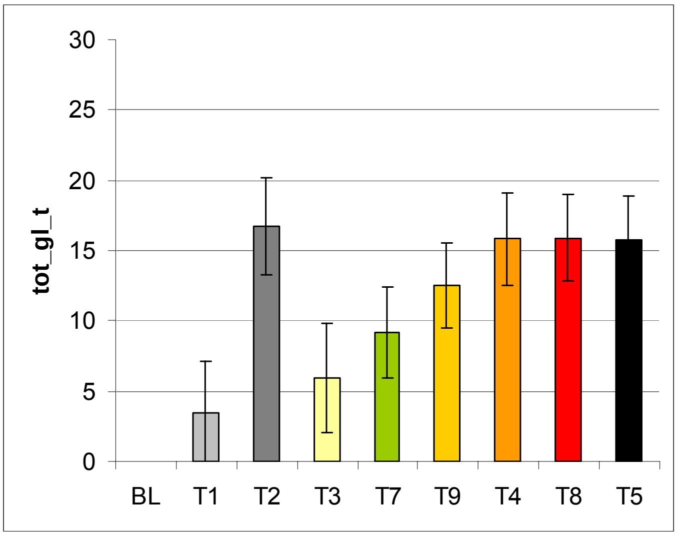


Figure 146 – Total glance time

9.7.2.6. Event detection

The graph for percentage of correct responses in PDT (Figure 147) indicates both an effect of task, compared to baseline, as well as an effect of task difficulty. However, only tasks 3, 5, 8 and 9 differ significantly from the baseline. Task 7 most likely scored high due to its nature (see above). The results are very similar to subjective rating (see Figure 137).

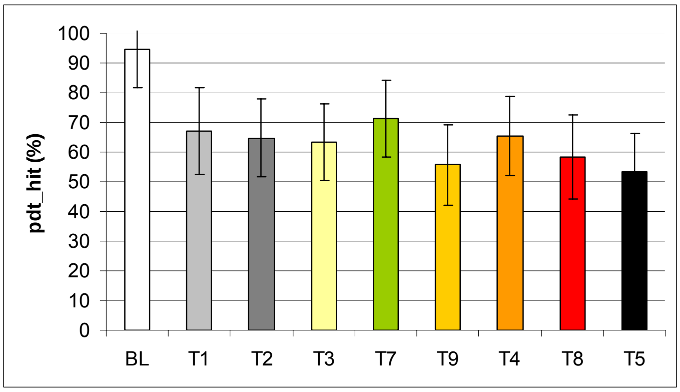


Figure 147 – Percent correct responses

No main effect was found for the PDT reaction times.

9.7.2.7. IVIS performance

All participants completed all tasks. The most complex task (T6) was removed prior to the field experiment due to its complexity.

9.7.3. Comparison between systems and groups of tasks

The two systems included in this experiment differed in a range of ways. System A had both auditory/cognitive tasks as well as visual/manual tasks while System B consisted of either visual only, or visual/manual tasks. System B seemed more difficult than System A simply due to the input device (small pen, high demand on precision) and size of monitor (smaller than for System A).

9.8. Result summary and conclusions

System A consisted of both auditory/cognitive and visual-manual tasks. Results indicated that some of the eye movement related measures may be useful (e.g. PRC) if a future test regime is to investigate differences in driving behaviour due to task modality.

In Deliverable 2 of the HASTE project, more demanding visual/visual-manual tasks produced larger steering wheel reversals, while smaller reversals were observed for demanding cognitive tasks. Certainly, for the cognitive tasks the large reversals were almost as few as for the baseline. In WP3, measures looking at larger reversal rate thresholds were not part of the draft test regime and therefore the WP2 findings could not be validated. However, rswt_70 showed larger reversals for visual tasks compared to the cognitive/auditory tasks.

System B consisted of tasks that were visual only (T1 and 2) as well as tasks which were visual-manual, where the manual input became more and more demanding. According to subjective rating of driving performance, it seems that System A tasks were deemed somewhat easier than System B tasks. This difference may have been partly due to the manual precision needed to perform tasks in System B. Also, the screen for System B was smaller.

In the present experiment, lane position and speed were very much affected by traffic density and therefore not very suitable measures for a field experiment where volume of traffic could not be controlled. One solution may be to change the threshold of pr_tlc for field tests.

Also, it can be argued that speed is perhaps not a very suitable measure for IVIS assessment, since it is mostly reduced during high workload conditions, perhaps because the driver is trying to adapt his/her behaviour to cope with the situation. What this adaptation actually means with regards to safety benefits is not clear.

For both systems A and B, measures related to steering wheel movements, PDT measures, glance measures and percent road centre values seemed most successful in capturing the effect of tasks. PDT is especially recommended since it is quite sensitive to Task/No Task and task difficulty variations. This method is also quite an expensive and easy method to use.

The measures calculated with the moving window technique went some way in solving the problem with time bias, created by the standard deviation calculations. However, new problems were introduced instead. For instance, it was not possible to include as many tasks for the longer time window, since very few tasks were longer than 30 seconds. Also, this technique meant that data from fewer participants were included in the calculation for each task. This is clearly a problem if the reason for exclusion is that these participants were actually more efficient at completing the task.

9.9. Measures summary tables

9.9.1. System A

Significant differences from baseline (BL) are indicated by grey background. The measures marked with (*) mean that the measure is not calculated for baseline. (**) for st_ga indicate that the measure is only calculated for baseline and for auditory/cognitive tasks 1-4.

System A	Mean values										Effect
	Auditory tasks					Visual-Manual tasks					
measure	BL	(T1)	(T2)	(T3)	(T4)	(T7)	(T8)	(T9)	(T5)	(T6)	SLv
task_l	101.4	11.4	15	13.9	20.4	11.4	18.4	39.5	-	-	
subj_r	9.2	8	8.5	7.7	7.5	7.7	7.1	6.7	-	-	✓
mn_sp	100.58	104.52	102.17	103.54	105.08	102.32	102.65	100.47	-	-	✗
st_sp	5.68	0.83	1.92	1.14	1.66	1.37	1.54	3.42	-	-	✓
st_sp15	1.57	-	1.41	0.54	1.31	2.26	0.99	1.61	-	-	✗
st_sp30	2.65	-	-	-	-	2.71	-	2.93	-	-	✗
u_sp	89.32	102.76	98.55	101.53	101.81	100.02	100.15	95.07	-	-	✓
d_sp	-6.83	-5.32	-0.48	6.15	2.37	-1.97	-9.07	-10.07	-	-	✗
hi_st	17.72	31.60	28.8	28.62	27.43	35.07	35.93	30.37	-	-	✓
pr_tlc1	0.0048	-0.0006	-0.0006	0.0002	0	0.0383	0.0177	-0.0024	-	-	✗
pr_tlc2	0.0469	-0.001	-0.0012	0.0007	0.0095	0.1038	0.0495	0.0166	-	-	✗
mn_tlc1	5.45	5.64	6.2	7.49	5.64	5.69	5.33	5.61	-	-	✗
mn_tlc2	3.29	3.96	3.85	4.25	3.24	3.49	3.32	3.50	-	-	✗
mn_lp	1.23	1.19	1.21	1.36	1.19	1.33	1.33	1.33	-	-	✗
st_lp	0.24	0.13	0.13	0.15	0.21	0.16	0.22	0.2	-	-	✓
st_lp15	0.16	-	0.13	0.12	0.18	-	0.24	0.18	-	-	✓
st_lp30	0.18	-	-	-	-	-	-	0.17	-	-	✗
rr_st1	12.36	11.18	17.38	11.24	13.17	19.88	21.00	19.63	-	-	✓
rr_st3	0.63	0.34	1.50	0.27	0.54	2.37	1.62	2.11	-	-	✗
rswt_5	6.92	7.55	9.02	6.18	8.43	10.6	11.74	11.83	-	-	✓
rswt_10	1.11	1.97	3.79	1.11	1.86	5.75	4.73	5.52	-	-	✓
rswt_20	0.03	0	0.24	0.38	0	0.38	0.35	0.51	-	-	✗

System A	Mean values										Effect
	Auditory tasks					Visual-Manual tasks					
rswt_40	0	0	0	0	0	0	0	0	-	-	×
rswt_70	0	0	0	0	0	0	0	0	-	-	×
pdt_hit	0.933	0.843	0.813	0.802	0.714	0.761	0.723	0.645	-	-	✓
pdt_rt	0.624	0.644	0.67	0.834	0.812	0.945	0.818	0.78	-	-	✓
pdt_miss	0.066	0.155	0.167	0.198	0.286	0.239	0.278	0.342	-	-	✓
pdt_cheat	0.001	0.002	0.021	0	0.000	0.000	0	0.013	-	-	×
n_gl*	N/A	N/A	N/A	N/A	N/A	4.3	9.5	18.5	-	-	✓
tot_gl*	N/A	N/A	N/A	N/A	N/A	0.35	0.49	0.50	-	-	✓
mn_gd*	N/A	N/A	N/A	N/A	N/A	0.79	0.97	1.10	-	-	✓
PRC***	51.3	62.2	64.2	68.2	60.9	29.5	19.8	22.6	-	-	✓
st_ga**	0.303	0.374	0.344	0.339	0.343	N/A	N/A	N/A	-	-	✓
tot_gl_t*	N/A	N/A	N/A	N/A	N/A	3.87	9.16	20.39	-	-	✓

9.9.2. System B

Significant differences from baseline (BL) are indicated by grey background. The measures marked with (*) mean that the measure is a glance based measure and not calculated for baseline.

System B		**Visual**		**Visual-Manual**							**Effect**
measure	BL	(T1)	(T2)	(T3)	(T7)	(T9)	(T4)	(T8)	(T5)	(T6)	
task_l	93.6	7.2	26.7	11.7	22.9	18.7	23.7	25.9	22.9	-	
subj_r	9.1	6.6	5.9	6.6	7.3	5.6	5.8	5.2	5.8	-	✓
mn_sp	103.8	98.9	100.9	98.6	101.6	98.9	96.5	98.6	101.6	-	✗
st_sp	5.2	0.9	2.9	1.4	2.7	1.9	2.3	3.1	1.4	-	✓
st_sp15	1.5	-	1.95	0.85	0.29	0.68	0.29	0.88	0.3	-	✗
st_sp30	2.58	-	3.66	7.01	6.15	3.69	2.68	3.59	1.99	-	✗
d_sp	-3.9	3.1	15.9	-12.0	-11.4	-6.9	-11.8	-13.7	-10. 4	-	✓
u_sp	93.2	97.4	95.2	96.6	97.6	95.3	92.9	93.4	99.1	-	✗
pr_tlc1	0.0171	0.0117	-0.0081	0.0098	0.0314	0.0350	0.0311	0.0146	0.0150	-	✗
pr_tlc2	0.0539	0.0138	0.0326	0.0173	0.1097	0.0231	0.0536	0.0715	0.0598	-	✗
mn_tlc1	4.9	7.29	4.58	5.61	5.57	4.88	4.92	4.83	5.11	-	✗
mn_tlc2	2.99	3.34	3.23	3.31	2.93	3.08	2.74	3.23	3.42	-	✗
mn_lp	1.246	1.286	1.177	1.482	1.231	1.314	1.138	1.236	1.317	-	✗
st_lp	0.292	0.141	0.218	0.182	0.215	0.277	0.195	0.270	0.246	-	✓
st_lp15	0.214	-	0.169	-	0.201	0.26	0.184	0.248	0.24	-	✗
st_lp30	0.256	-	0.241	-	0.225	0.348	0.159	0.223	0.192	-	✗
rr_st1	13.59	24.06	23.05	24.96	21.31	22.29	24.85	24.92	23.22	-	✓
rr_st3	1.00	0.40	2.86	5.97	2.71	3.78	4.76	4.37	5.81	-	✓
hi_st	17.03	39.97	33.77	36.46	29.46	34.05	34.17	36.09	31.23	-	✓
rswt_5	7.32	14.98	12.31	16.15	12.52	13.69	13.41	13.68	14.52	-	✓
rswt_10	1.55	5.81	6.72	8.43	5.40	5.57	7.99	7.91	7.95	-	✓
rswt_20	0.03	1.27	0.59	1.91	0.91	0.80	0.82	1.12	1.51	-	✗
rswt_40	0	0	0	0	0	0.26	0	0	0	-	✗
rswt_70	0	0	0	0	0	0	0	0	0	-	✗
pdt_hit	0.95	0.67	0.65	0.63	0.71	0.56	0.65	0.58	0.53	-	✓

System B		Visual		Visual-Manual							Effect
measure	BL	(T1)	(T2)	(T3)	(T7)	(T9)	(T4)	(T8)	(T5)	(T6)	
pdt_rt	0.57	0.96	0.94	0.8	0.76	0.75	0.82	0.93	0.81	-	✗
pdt_miss	0.05	0.33	0.34	0.37	0.29	0.44	0.35	0.42	0.47	-	✓
pdt_cheat	0.39	0	0.02	0	0	0	0	0	0	-	✗
mn_gd*	-	0.82	1.3	1.16	0.93	1.26	1.22	1.36	1.33	-	✓
n_gl*	-	4.8	13.3	5.3	10.5	10.1	13.3	11.6	11.7	-	✓
tot_gl*	-	0.52	0.65	0.64	0.42	0.64	0.61	0.64	0.66	-	✓
PRC	51.3	24.7	11.3	18.6	35.3	17.2	19.2	18.5	14.7	-	✓
tot_gl_t*	-	3.5	16.71	5.94	9.15	12.51	15.85	15.89	15.77	-	✓

10. The VTT field experiment

This field study was designed to investigate and compare the potential or sensitivity of selected assessment methods to reflect the effects of different IVIS on driver behaviour. More specifically, (1) the data were collected in real traffic, (2) the effects of two different IVIS on driver performance were quantified, (3) the road type included motorway driving, and (4) evaluations were based on vehicle data, observations and drivers' reports. The observations were made according to the Wiener Fahrprobe protocol (see Appendix 4).

10.1. Test site

The instrumented vehicle used in the tests was a 1999 Toyota Corolla sedan with manual transmission. The vehicle was equipped with a hidden PC-based measuring system and differential GPS receiver. Data collection frequency was 10 Hz for speed and distance data, and for steering-wheel angle data. The data was transmitted to a computer in the boot of the car.

10.2. Scenarios and participants

The tests were carried out on a motorway section in the Helsinki capital area. The posted speed limit was 120 km/h. The test section was 35 km long. Secondary tasks were performed on the motorway at link sections.

The data included 18 average drivers aged between 28 and 51 years (mean 40 years). Two of them were females and 16 were males. All drivers owned or regularly drove a vehicle of the same type as the one used in the study.

10.3. S-IVIS included

Systems A and C were included. However, with system A, tasks 5 and 6 were excluded, since they were too difficult to perform in real traffic conditions. Therefore system A included seven tasks and system C nine tasks. The drivers performed each task once.

10.4. Experimental design

The test route was driven three times: with system A, with system C, and baseline (driving with no IVIS). The order of the systems and tasks was balanced across participants. The timing of the secondary task was determined and controlled by the distance travelled from fixed points (determined with a GPS receiver).

10.5. Procedure

Participants were told that the study investigated how well drivers perform IVIS tasks while driving. Particularly, participants were instructed to drive safely through the test route and perform a secondary task when it was presented to them.

Since the experiments were conducted in real traffic, an experimenter sat in the front passenger seat, equipped with an extra brake pedal. He also gave directions in order to maintain the correct route. An observer, whom the driver believed to be technical support staff, sat in the back. At no time did the observer interfere with the driving. After completing the drive, participants were told that their driving behaviour had been recorded during the experiment, and their permission to use the data was requested.

Experiments were carried out in May and June 2004. The data was collected on weekdays between 9 a.m. and 3 p.m. The experiments were conducted in good weather and road surface conditions; i.e. there was no precipitation or water on the road surface.

10.6. Measures and analysis method

Speed behaviour and steering wheel position were recorded. Mean (mn_sp) and minimum speed (u_sp), speed change (d_sp) and speed variation (st_sp) were calculated from collected data. Also, braking jerks of more than 8 m/s^3 were counted. One (rr_st1) and three degrees reversal rate (rr_st3) as well as the high-frequency component of steering-wheel movements (hi_st) were computed. Drivers' self-reported driving quality (subj_r) was asked after each IVIS block and at the same locations during the baseline run.

In addition to driver behaviour, the accompanying observer coded drivers' performance and the traffic conditions with respect to:

- presence of vehicle in front
- interaction with vehicles in front
- lane-keeping behaviour
- lane-change behaviour
- speed choice and adaptation
- interaction with other road users (potential conflict etc.)

The observer coded driver performance was based on the Wiener Fahrprobe protocol (Risser, 1985). Observer rating is a method for rating driving performance on a tactical level (Michon's driver model).

10.7. Results

10.7.1. Effects of system A

10.7.1.1. Task length

The task lengths varied between 7 seconds (T1 and T7) and 46 seconds (T9). As shown in Figure 148, the visual/manual tasks took much longer on average than the auditory ones.

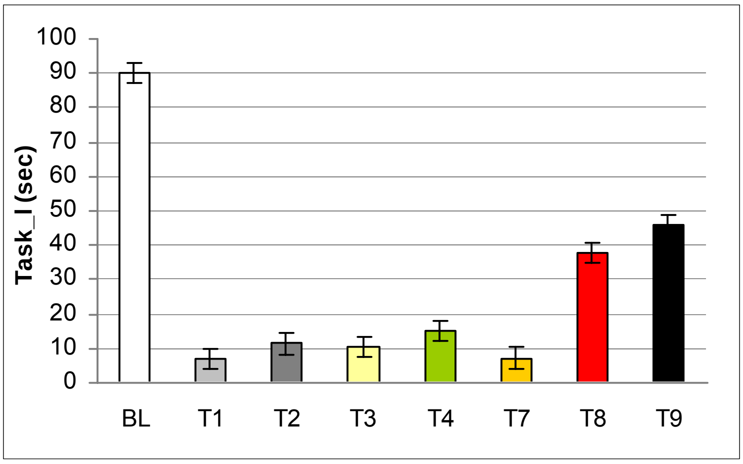


Figure 148 – System A average task lengths

10.7.1.2. Self-reported driving performance

Drivers rated their driving performance on a scale of 1 ("I drove extremely poorly") and 10 ("I drove extremely well"). Drivers rated their driving performance (subj_r) as better in baseline conditions (8.8) than when driving with the system A tasks (7.6) — each task deteriorated driving performance significantly, compared to baseline. Although the differences between the tasks were relatively small, the effects tended to be larger for the visual/manual tasks (average 1.5 decrease) than for the auditory tasks (average 0.8 decrease). Also, as shown in Figure 149, the effects were larger for longer tasks (T8, T9) than for shorter ones.

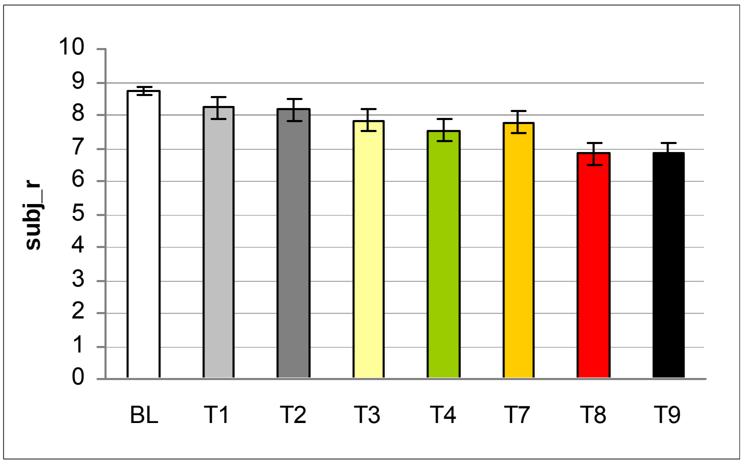


Figure 149 – Self-reported driving quality on observed road sections for system A tasks

10.7.1.3. Longitudinal control

The main results for speed behaviour are given in Table 23. The IVIS task tended to decrease the mean (mn_sp) and minimum speed (u_sp). Also, the speed change (divided by task length) during task completion was found to be negative, indicating a reduction in speed. In addition, the system A tasks had a tendency to increase speed variation. However, none of these effects was found to be statistically significant (on a main level).

Table 23 – Speed behaviour by task (N in brackets)

Secondary task condition	mn_sp (km/h)	u_sp (km/h)	d_sp (km/h/minute)	st_sp (km/h)	st_sp15 (km/h)	st_sp30 (km/h)
Baseline	113.4	108.7	-0.63	2.31	1.09 (18)	1.63 (18)
System A						
Task1	108.2	106.7	-7.82	0.82		
Task2	110.0	108.3	-6.86	1.27		
Task3	109.3	107.5	-12.25	1.17		
Task4	109.5	107.6	-9.88	1.26	1.24 (18)	
Task5	109.7	108.5	-15.84	0.77		
Task8	107.7	104.2	-4.11	2.23	1.28 (18)	2.29 (12)
Task9	108.0	103.7	-3.63	2.46	1.31 (18)	1.92 (16)

10.7.1.4. Lateral control

The IVIS task tended to increase steering-wheel reversal rates (rr_st1 and rr_st3) and high component of steering-wheel movements (Table 24).

Table 24 – Reversal rate (1 and 3 degrees minimum change) and high frequency component of steering-wheel movements by task

Secondary task condition	rr_st1 (1/minute)	rr_st3 (1/minute)	Hi_st (%)
Baseline	23.47	11.37	10.84
System A			
Task1	26.55	11.85	15.29
Task2	22.51	9.52	13.67
Task3	25.11	10.36	13.64
Task4	23.62	10.62	14.55
Task7	28.77	14.46	18.64
Task8	28.91	15.23	16.65
Task9	28.56	14.40	14.17

Post-hoc tests showed that the effects were significant with a one-degree reversal rate between baseline and visual/manual tasks in tasks 7-9 (Figure 150). However, the task duration had no effect on steering-wheel reversal rate or other lateral control measures.

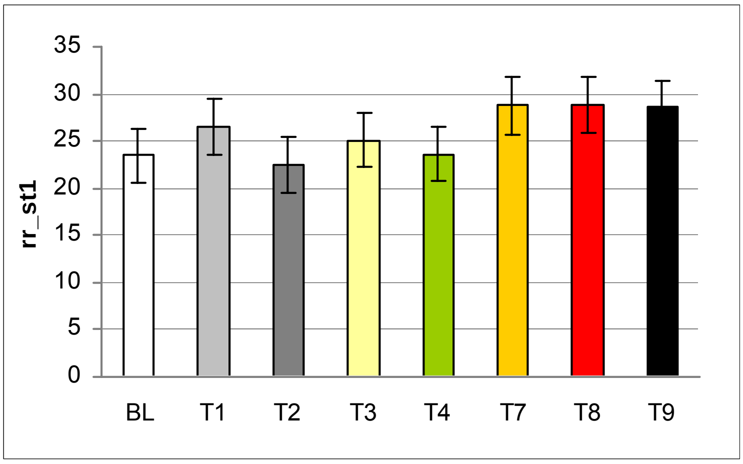


Figure 150 – Steering-wheel one-degree reversal rate for system A tasks

10.7.1.5. IVIS performance

The IVIS tasks were conducted as instructed. Only on two occasions were the participants unable to perform task 7 without some guidance. Those two trials were excluded from data analyses.

10.7.2. Effects of system C

10.7.2.1. Task length

The task length varied between 16 seconds (T1 and T2) and 41 seconds (T9). In general, the tasks that required more scanning, i.e. visual/manual tasks (T3, T6 and T9), took longer than tasks that required less scanning (T1, T4 and T7), as shown in Figure 151.

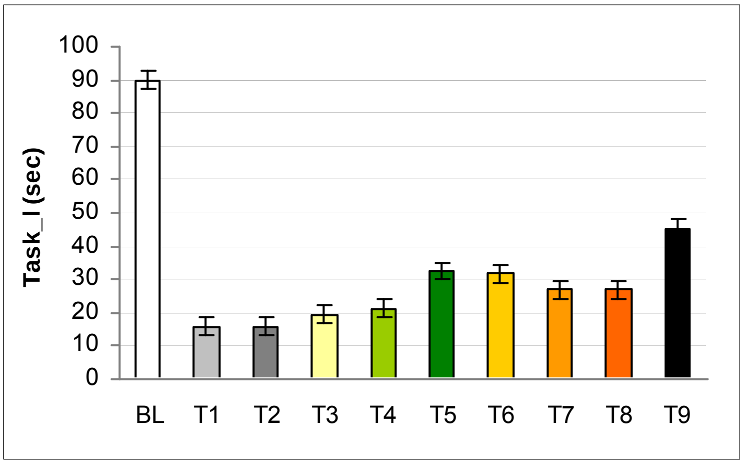


Figure 151 – System C, average task lengths

10.7.2.2. Self-reported driving performance

Figure 152 shows that drivers rated their driving performance as better in baseline conditions (8.8) than when driving with the IVIS task (7.3). The drivers' subjective driving performance was reduced significantly in all tasks. However, differences between tasks were relatively small. Ratings for driving performance fell by only 1.5 during tasks where the driver needed to search for the new message (T4 – T9), compared with a 1.3 decrease for tasks where drivers were asked only to read the new message (T1 – T3).

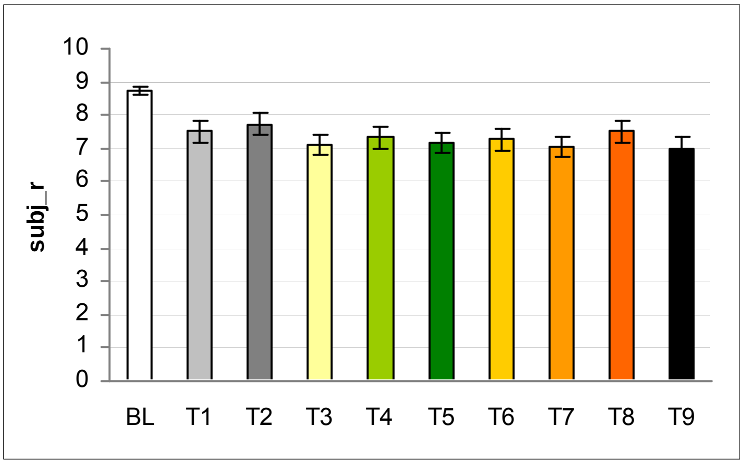


Figure 152 – Self-reported driving quality on observed road sections for system C tasks.

10.7.2.3. Longitudinal control

The main results for speed behaviour when driving with IVIS tasks are given in Table 25. The IVIS tended to decrease the mean speed (mn_sp), and minimum speed (u_sp). Also, the speed change divided by task length (d_sp) was negative, indicating a speed reduction. The speed variation results (st_sp) were not clearly consistent, which might have been due to the different length of the baseline and tasks.

Table 25 – Speed behaviour by task (N in brackets)

Secondary task condition	**mn_sp** (km/h)	**u_sp** (km/h)	**d_sp** (km/h per minute)	**st_sp** (km/h)	**st_sp15** (km/h)	**st_sp30** (km/h)
Baseline	113.4	108.7	-0.63	2.31	1.09 (18)	1.63 (18)
System C						
Task1	110.1	106.7	-3.75	1.89	2.35 (8)	0
Task2	110.2	107.0	-15.99	1.86	2.13 (8)	0
Task3	109.4	105.7	-12.86	2.24	2.18 (13)	1
Task4	108.3	104.7	-11.11	2.50	2.20 (14)	3
Task5	106.5	101.7	-10.38	2.72	1.64 (18)	3.04 (10)
Task6	110.1	106.9	-1.28	1.96	1.43 (18)	2.64 (8)
Task7	106.3	103.0	-8.61	2.23	1.41 (18)	7
Task8	109.8	107.0	0.00	1.63	1.10 (18)	6
Task9	107.1	102.1	-11.03	3.28	1.62 (18)	2.59 (18)

Post-hoc tests showed that the effect of the IVIS task on mean speed (mn_sp) was significant between baseline and tasks 2-5 and 9. The minimum speed (u_sp) was significantly lower for tasks 4, 5, 7 and 9 compared to the baseline (Figure 153).

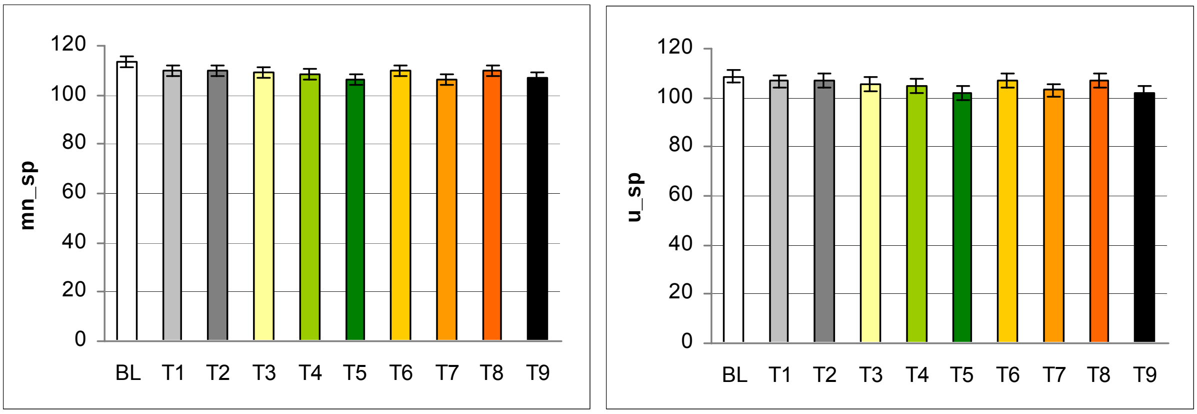


Figure 153 – Mean speed (left) and minimum speed (right) by system C task

Speed reduction during task completion (d_sp) was significantly higher for tasks 2-5 and 9, compared to baseline conditions (Figure 154).

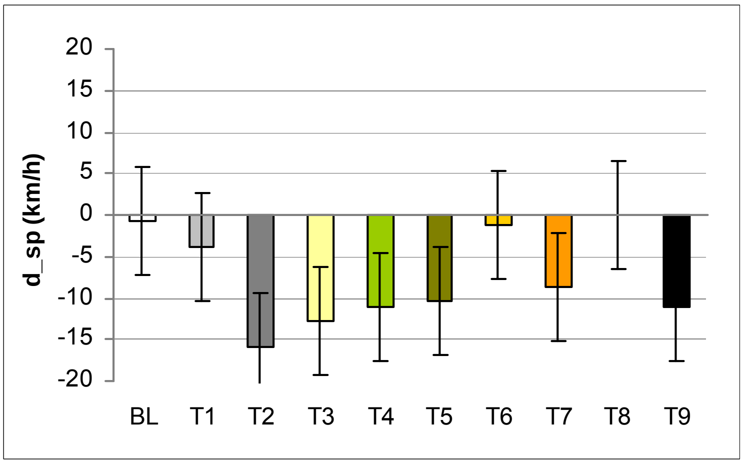


Figure 154 – Speed reduction during task completion (d_sp) by system C task

10.7.2.4. Lateral control

The effects of the IVIS task on lateral control are shown in Table 26. The IVIS task caused a significant increase in both reversal rates (rr_st1 and rr_st3) and tended to increase the high component of steering-wheel movements (hi_st).

Table 26 – Reversal rate (1 and 3 degrees minimum change), high frequency component of steering-wheel movements and number of rapid steering-wheel turns by task

Secondary task condition	rr_st1 (1/minute)	rr_st3 (1/minute)	Hi_st (%)
Baseline	23.47	11.37	10.84
System C			
Task1	35.01	15.69	17.47
Task2	31.90	16.12	13.64
Task3	31.00	13.66	24.63
Task4	28.35	19.42	14.67
Task5	31.23	15.02	20.49
Task6	32.07	15.19	16.41
Task7	30.17	15.23	22.86
Task8	31.98	16.22	17.66
Task9	29.19	17.54	14.79

When comparing the one-degree reversal rate by task, the effect of the IVIS task was significant (between baseline and all tasks, Figure 155). However, the effect by task length was not systematic, i.e. the length of the tasks or the number of button depressions did not necessarily increase the reversal rate.

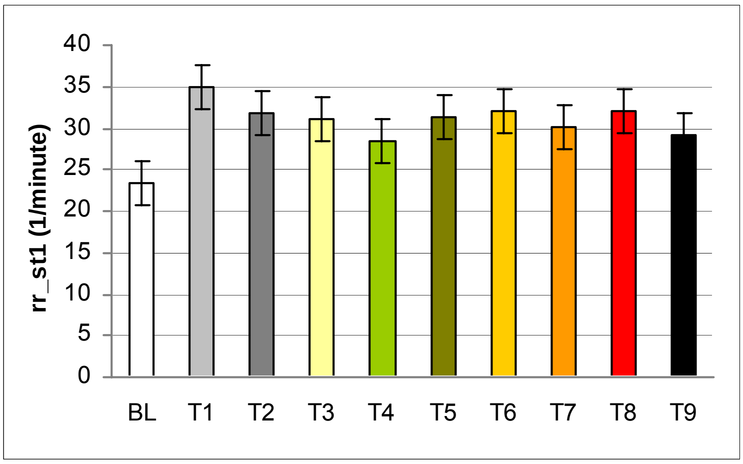


Figure 155 – Steering-wheel one-degree reversal rate by system C task

10.7.2.5. IVIS performance

The IVIS tasks were conducted as instructed. On twelve occasions, the participants could not perform the tasks without some guidance. This happened once with tasks 2, 6 and 7, and three times with tasks 3 and 9. Those trials were not included in data analyses.

10.7.3. Observed effects of system A and system C

The following sections describe the results obtained from the observer ratings, using the Wiener Fahrprobe protocol.

10.7.3.1. Traffic conditions

The traffic conditions by system did not differ in terms of presence of vehicles in front. Specifically, many of the observed sections included a vehicle in front that might have had an effect on driving (reduction in speed, lane change etc).

10.7.3.2. Longitudinal control

In baseline conditions, the speed behaviour was assessed to be inappropriate or the speed changed markedly in 60% of the sections, compared to 78% while performing the system C tasks and 74% while performing the system A tasks. Compared to the baseline condition, the observed slowing down (obs_sp_irr) was more frequent if the driver performed a secondary task while driving. In addition, the proportion of drivers assessed as driving fast was smaller in secondary task conditions (Figure 156).

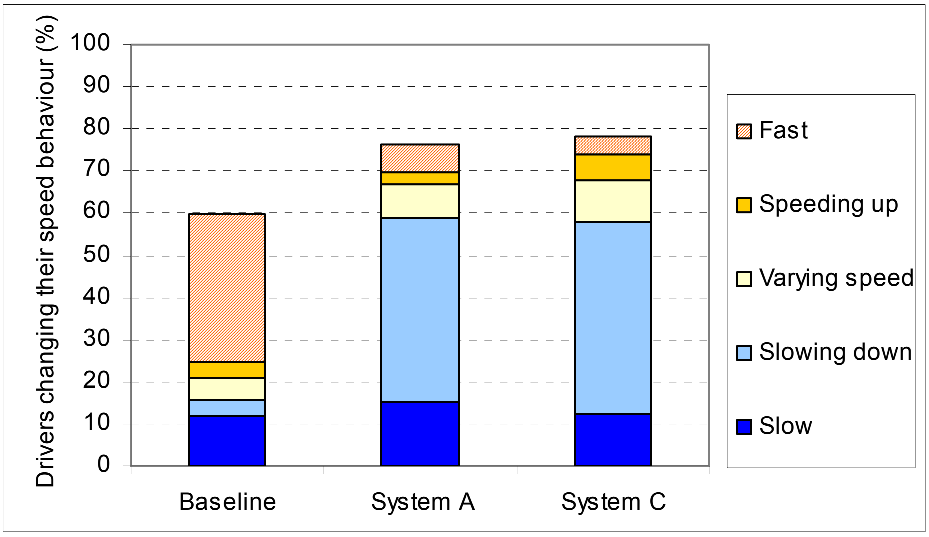


Figure 156 – Changes in speed behaviour by secondary task condition

In general, when increasing the secondary task difficulty, a greater proportion of drivers were observed to have inappropriate speed behaviour (Figure 157).

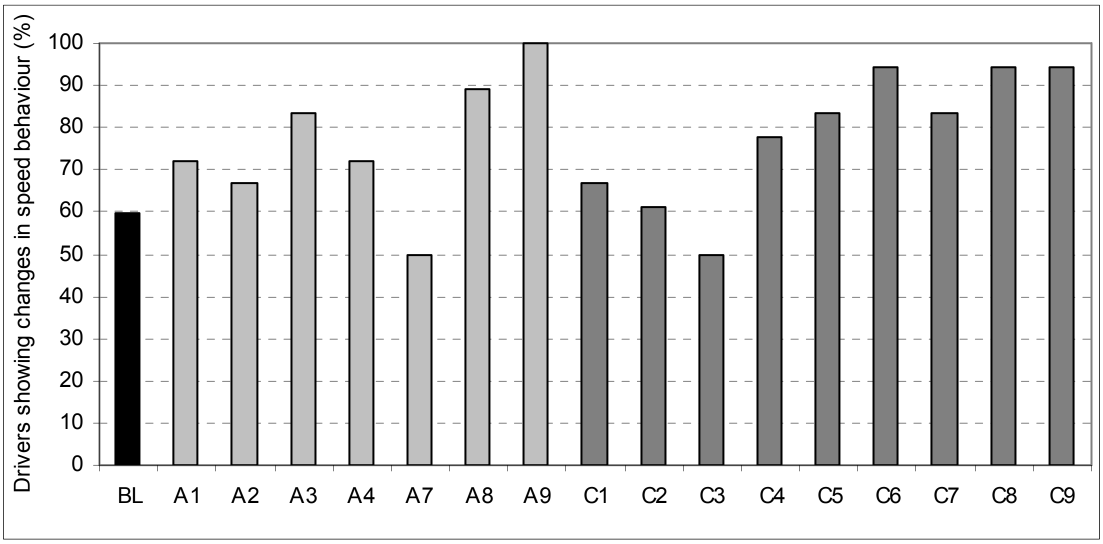


Figure 157 – Percentage of drivers observed to have inappropriate changes in speed behaviour

In addition to monitoring drivers' lateral position variations; their lane-change behaviour was assessed during observations. In the baseline runs, drivers were observed to change lane in 51% of all observations, with 39% for system A tasks and 35% for system C tasks. As seen in Figure 158, the secondary task not only reduced the average number of lane changes, but also increased the number of *inappropriate* ones. The percentage of inappropriate lane changes was six in baseline conditions, 10 with system A tasks and 24 with system C tasks.

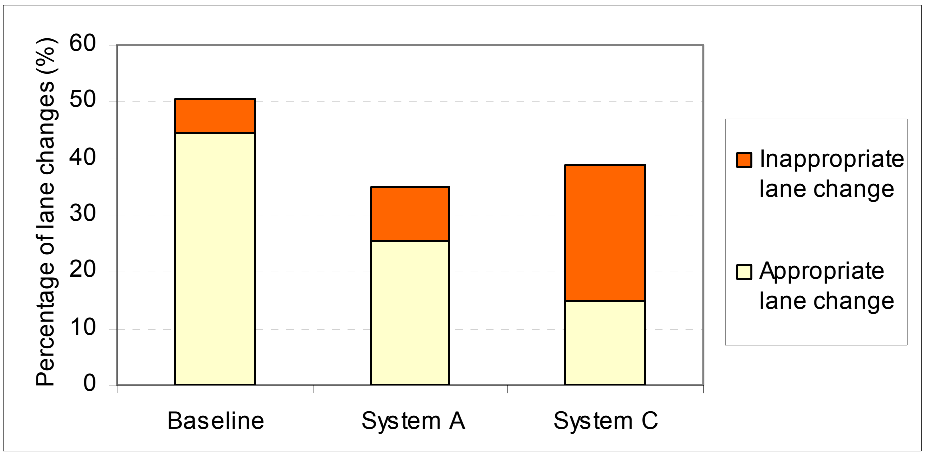


Figure 158 – Observations of lane change behaviour

There were very few sudden-braking events and only one 'incident' occurred while a system A task was being performed.

10.7.3.3. Lateral control

In baseline conditions, the observer coded the lateral control as inappropriate in 9% of observations, compared with 48% for system A conditions and 91% for system C conditions. While engaged in the information tasks, drivers were mainly observed to have lateral movement within their own lane (Figure 159).

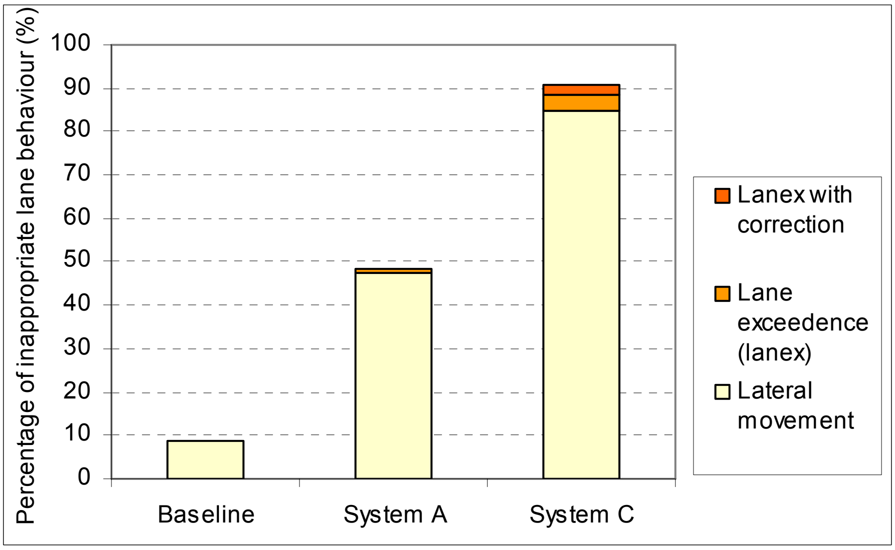


Figure 159 – Type of observed inappropriate lane behaviour by secondary task condition.

The results by task showed that the inappropriate lane behaviour mainly occurred with visual/manual tasks. In other words, with system C, inappropriate lane behaviour occurred with all tasks, whereas with system A, inappropriate lane behaviour was mainly imposed by tasks 8 and 9 (Figure 160).

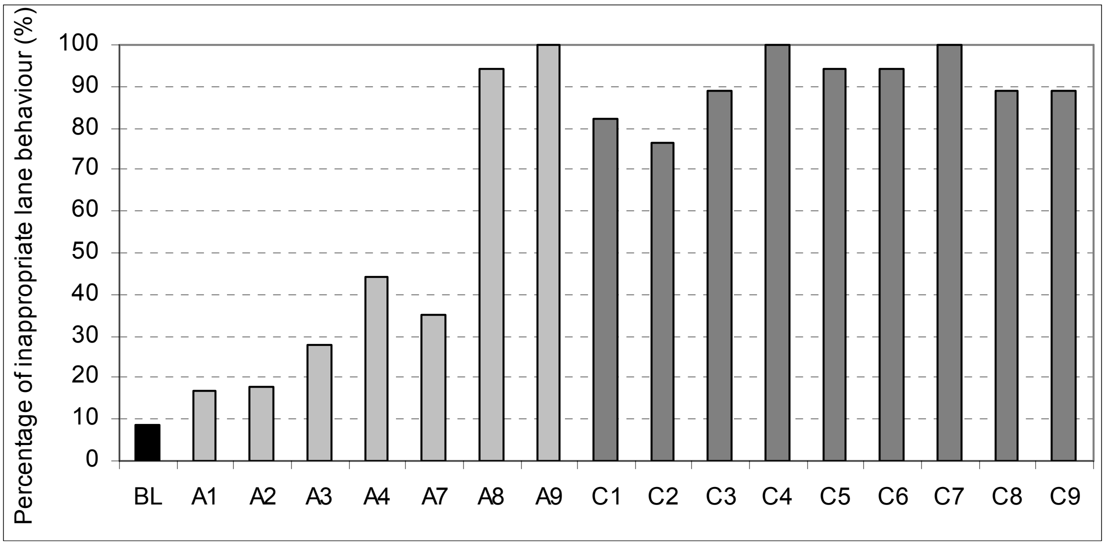


Figure 160 – Percentage of drivers observed to have inappropriate lane behaviour

Roughly 80% of the observation sections included a vehicle in front. In the baseline runs, the drivers were observed to maintain appropriate headway in 91% of car-following situations. The rate was 87% while engaged in system A tasks, and 81% while engaged in system C tasks (Figure 161). As the total number of following-too-close cases was quite small — only three to nine drivers per secondary task — no further analyses were conducted by task.

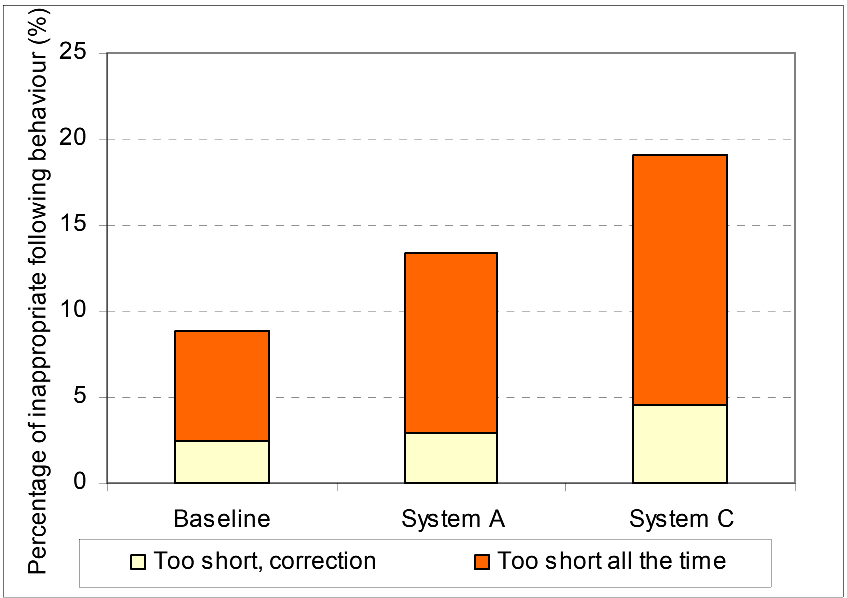


Figure 161 – Percentage of drivers observed to have inappropriate car-following behaviour by secondary task condition.

10.8. Result summary and conclusions

There was no substantial difference in task duration by system. However, visual/manual tasks took longer on average than auditory ones. System A included three visual/manual tasks and all the tasks of system C were visual/manual.

While using each system, drivers reported that their driving performance fell significantly during all tasks. There was no substantial difference by system.

Another measure that led to a significant difference between the baseline and the tasks was the one degree reversal rate (rr_st1). Specifically, three visual/manual tasks of system A and all tasks of system C significantly increased the reversal rate. In addition, the observation of driver behaviour showed that the visual/manual tasks increased inappropriate lane behaviour, the proportion of inappropriate lane changes being 10% with system A and 24% with system C. System C also seemed to increase inappropriate headway more than system A.

In conclusion, the main findings of this study suggest that two measures indicated statistically significant effects of IVIS tasks on driver behaviour. Those measures were driver's self-reported driving quality (subj_r) and the one degree reversal rate (rr_st1). Another main finding suggested that especially visual/manual tasks of the tested systems deteriorated driver behaviour.

10.9. Measures summary tables

10.9.1. System A

Travel Pilot	**Mean values**									**Effect**
	Auditory tasks				Visual-Manual tasks					
measure	BL	T1	T2	T9	T4	T7	T8	T9	T5	T6	SLv
task_l		6.9	11.4	46.1	15.3	7.3	37.9	46.1	-	-	
subj_r	8.7	8.2	8.2	6.8	7.5	7.8	6.8	6.8	-	-	✓
mn_sp	113.4	108.2	110.0	108.0	109.5	109.7	107.7	108.0	-	-	×
u_sp	108.7	106.7	108.3	103.7	107.6	108.5	104.2	103.7	-	-	×
st_sp	2.3	0.8	1.3	2.5	1.3	0.8	2.2	2.5	-	-	×
st_sp15	1.1			1.3	1.2	1.8	1.3	1.3	-	-	
d_sp	-0.6	-7.8	-6.9	-3.6	-9.9	-15.8	-4.1	-3.6	-	-	×
rr_st1	23.5	26.5	22.5	28.6	23.6	28.7	28.9	28.6	-	-	✓
rr_st3	11.4	11.9	9.5	14.4	10.6	14.5	15.2	14.4	-	-	✓
hi_st	10.8	15.3	13.7	14.2	14.6	18.6	16.7	14.2	-	-	×

(significant difference from BL indicated by grey background)

10.9.2. System C

Information system	**Mean values**										**Effect**
measure		Type 1 tasks			Type 2 tasks			Type 3 tasks			
measure	BL	T1	T2	T3	T4	T5	T6	T7	T8	T9	SLv
task_l		15.8	15.8	19.6	21.2	32.6	31.7	26.9	27.0	45.6	
subj_r	8.8	7.5	7.7	7.1	7.3	7.2	7.3	7.1	7.5	7.0	✓
mn_sp	113.4	110.1	110.2	109.4	108.3	106.5	110.1	106.3	109.8	107.1	✓
u_sp	108.7	106.7	107.0	105.7	104.7	101.7	106.9	103.0	107.0	102.1	✓
st_sp	2.3	1.9	1.9	2.2	2.5	2.7	2.0	2.2	1.6	3.3	×
st_sp15	1.1	2.4	2.1	2.2	2.2	1.6	1.4	1.4	1.1	1.6	✓
st_sp30	1.6			4.5	3.2	3.0	2.6	2.8	1.8	2.6	×
d_sp	-0.6	-3.7	-16.0	-12.9	-11.1	-10.4	-1.3	-8.6	0.0	-11.0	✓
rr_st1	23.5	35.0	31.9	31.0	28.4	31.2	32.1	30.2	32.0	29.2	✓
rr_st3	11.4	15.7	16.1	13.7	19.4	15.0	15.2	15.2	16.2	17.5	✓
hi_st	10.8	17.5	13.2	24.6	14.7	20.5	16.4	22.9	17.7	14.8	✓

11. Cross test site comparisons

A meta-analysis of the results of all studies was performed in order to (i) bring out and grasp the common patterns in the experiments, (ii) to identify and select the most powerful parameters for detecting these patterns, and (iii) to check whether the conclusions drawn from the meta-analysis of the earlier S-IVIS studies (WP2) would uphold for real systems. If so, it could be concluded that results of the present studies provide the most important ingredients for a test regime for an IVIS. This chapter mainly focuses on the mandatory measures, i.e. the measures used by all partners. The Peripheral Detection Task measures as well as the measures reflecting eye movements (Percent Road Centre) were also compared across sites, since they appeared to be promising on the basis of earlier results. However, since the equipment for these measures were not available to all partners, they were not tested in all experiments, conclusions on PDT and PRC should therefore be treated with some care.

The analysis was done along the lines of the earlier analysis in WP2. It comprised a sequence of steps, each of them putting a statistical or methodological requirement on a measurement parameter that it could or could not fulfil. If not, the parameter was not retained for the next step.

11.1. Initial check on significance

For each separate study reported in this Deliverable, at first only those parameters were considered that were mandatory and that showed a significant main effect in the study of ‘IVIS Task level’ and were significant to at least the $p = .10$ level. Therefore, over all the experiments, the parameters retained for the next step of the analysis were those that were significant in at least 90% of all studies in which they had been measured (because this is the expected percentage of tests to be found significant at the .10 level). A further criterion applied was that a parameter should have been measured in at least four (out of 12) experiments. There were 11 parameters that met both these criteria.

11.2. Consistency of effects

The next step of the analysis involved considering the *consistency* of the effects. This was based on the reasoning that all task level effects for a given IVIS should be in the same direction, compared to baseline. This test ensures that a behavioral parameter cannot be considered a good indicator of IVIS task difficulty if it varies (in sign) over conditions. The consistency check on the sign of the effects thus looked for parameters that may have been showing differently signed effects (i.e. positive as well as negative) over the task levels, for a given system. This was also done for each experiment. An overall consistency estimate per parameter (′% consistent′) was obtained by noting how often a parameter was consistent in this sense over the total set of experiments.

11.3. Effect sizes and range of effects

These first steps of the analysis served to check whether certain a priori statistical requirements were met by parameters. The next steps looked in more detail into how strong the effects associated with parameters actually were, and whether these effects were sufficiently diverse over the task levels. This is because, ideally, one would only like to work with a parameter that not only shows overall strong effects relative to baseline, but that also shows sufficient differentiating capability between different task levels.

Effect sizes (Cohen´s d) were computed for the set of 11 parameters, which involved calculating the standardized difference score for an IVIS task level effect relative to baseline. This was done per experiment.

The performance of the, as yet, retained parameters were then described by (a) their average effect size (over all IVIS task levels), presented in Table 27 below; and (b) the corresponding range over which the effect sizes varied (see Table 28). This was, again, done per experiment. The values of averages and ranges were computed only for those experiments in which the parameter had been identified as consistent.

Table 27 – Table of average effect sizes

		Average Effect sizes												Grand Average	Consistency Score
	System→	B	B	B	B	B	A	A	A	A	C	C	D		
Measure	**Envt. →**	Lab	SIM	SIM	SIM	Field	SIM	SIM	Field	Field	Lab	Field	SIM		
Subj_R		2.96	2.71	1.97		2.69	1.86		1.41	1.53	0.97	2.1		2.02	1.00
U_SP								1.36	1.41	1.38		1.21	1.13	1.30	0.56
MN_SP		1.25	1.08	1.00			0.8			0.83	0.96	1		0.99	0.70
SD_LP						0.92		1.35	0.98					1.08	0.38
HIST			0.92	0.56		2.10	0.52		1.33				1.04	1.08	0.88
HiST2						1.24								1.24	0.25
RRST1				1.56	0.63	1.09	0.99		0.63		1.11	1.35	0.88	1.03	0.67
RRST3					1.10	1.15		0.79	0.64			0.94	1.29	0.99	0.67
U_HWT		1.28	1.61	0.75	0.74		0.54	1.36			0.42		0.94	0.96	1.00
SD_HWD								2.12					1.39	1.76	0.29
SD_HWT								1.84					1.26	1.55	0.29

Table 28 – Table of effect size ranges

		Range of Effect Sizes												Grand Average	Sum of products for averages and ranges
	System →	B	B	B	B	B	A	A	A	A	C	C	D		
Measure	**Envt. →**	Lab	SIM	SIM	SIM	Field	SIM	SIM	Field	Field	Lab	Field	SIM		
Subj_R		2.22	2.46	1.80		1.84	3.32		1.63	1.85	0.64	1.04		1.87	**3.89**
U_SP								1.98	1.02	0.88		0.73	0.93	1.11	1.35
MN_SP		1.07	1.04	0.93			1.11			0.42	1.41	0.82		0.97	**1.37**
SD_LP						1.73		1.88	1.39					1.67	1.05
HIST			2.05	1.25		1.27	0.91		0.85				0.59	1.15	**1.96**
HiST2						3.32								3.32	1.14
RRST1				0.64	0.47	1.46	1.55		0.75		0.51	1.17	1.25	0.98	1.34
RRST3					0.26	1.15		1.62	0.93			1.17	1.17	1.05	1.36
U_HWT		0.71	1.19	0.27	0.41		0.81	0.49			0.69		0.46	0.63	**1.58**
SD_HWD				0.81				1.66					1.19	1.22	0.86
SD_HWT								1.79					1.22	1.51	0.89

11.4. Obtaining the final index of parameters' discriminative power

The final step in the analysis consisted of adding a parameter's average effect size and its range, and multiplying the sum by its consistency value.

The results of this procedure for the 12 retained parameters are shown below. These final values, therefore, reflect all the elements that should be included in assessing a parameter's discriminative power, i.e.:

- Statistical significance
- Consistent directional effects over task levels
- Sizeable effects compared to baseline; and
- Differentiating power between IVIS task levels.

Table 29 – Final index of parameters' discriminative power

Parameter	Consistency index (= A)	Average effect size (= B)	Range (= C)	Score (A* (B+C))	Remarks
Subj_R	1.00	2.02	1.87	3.89	Consistent, very large effects, very good discrimination
U_SP	.56	1.30	1.11	1.35	Low consistency
MN_SP	.70	.99	.97	1.37	Reasonable consistency fects
SD_LP	.38	1.08	1.67	1.05	Low consistency
Hi_ST	.88	1.08	1.15	1.96	Good consistency, large effects, good discrimination
Hi_ST2	.25	1.24	3.32	1.14	Very low consist.
RR_ST1	.67	1.03	.98	1.34	Reasonable in all respects
RR_ST3	.67	.99	1.05	1.36	Reasonable in all respects
SD_HWT	.29	1.55	1.5110	.89	Very low consistency
U_HWT	1.00	.96	.63	1.58	Good in all aspects
SD_HWD	.29	1.76	1.22	.86	Very low consistency

On the basis of the results contained in this Table it appears that the final set of indicators, i.e., the prime candidates to base a test regime upon, should comprise the indicators presented in the table below.

Table 30 – Indicators proved to be successful in the Meta analysis

Measure	Motivation
subj_r	Average effect size in experiments: 2.02. While this parameter was already found to be a very powerful one in WP there remained some doubt at that time as to whether participants would not let themselves be led in their ratings by the obvious stepwise variation in difficulty level of the S-IVIS tasks, rather than by their 'intrinsic' experience of their own driving quality. In the present studies, however, there was no such obvious element in the tasks that Ss could base their judgments on. It is therefore plausible that the ratings must now indeed have been based on the Ss 'intrinsic' feelings about the quality of their own driving performance. If this is accepted, subj_r is the best we have, and it may in fact be regarded as the only parameter we need.
mn_sp	Average effect size in experiments: 0.99. This parameter appeared to be the best one from among the collection of those describing speed behaviour. The effect sizes associated with it are considerably smaller than for Subj_R, and so is the range of effects found for it. Nevertheless, its discriminative power must be considered as good.
hi_st	Average effect size in experiments: 1.08. This index of high-frequency steering activity has a high score on all dimensions of discriminative power.
u_hwt	Average effect size in experiments: 0.96. This parameter of car-following behaviour is good to excellent in all relevant respects.

A correlation analysis per system and experiment showed that these four indicators are relatively independent of each other. That is, all four would be needed together for a complete evaluation of IVIS effects on driving behavior. The grand average correlations between the parameters over all studies are shown in the table below.

Table 31 – Correlations between the four primary behavioural parameters

Parameter	MN_SP	HI_ST	U_HWT
Subj_R	+ 0.56	-0.37	-0.51
MN_SP		-0.18	-0.31
HI_ST			+ 0.38

A further important consequence of this consideration of discriminative power is that it makes it possible to estimate the number of subjects that would be required to find effect sizes as big as they apparently are with statistical power of 85% or better, the conventional level. The number of participants required for a significant one-tailed (i.e. directional) test for the four parameters is shown below (from Tables in Cohen, 1988).

Table 32 – Effect sizes and minimal number of participants required to find them

Parameter	Effect size found in experiments	Required N to find effect size at .10 level
Subj_R	2.02	Less than 10
MN_SP	.99	15
HI_ST	1.08	13
U_HWT	.96	15

These results suggest that 15 participants are sufficient to show an effect of IVIS (if it exists) compared to baseline, for each of those parameters.

11.5. What do the results say about the safety of a specific IVIS?

Taking the four prominent parameters combined as the tool for getting to a judgement, it should now be possible to qualify the four IVIS used in WP3, in terms of what their safety effects would be. The most appropriate way of doing so is to establish the quantitative link between the behaviour parameters and safety risk through the procedure developed in another part of HASTE, using the Stated Preference technique which is under development at The University of Leeds). However, by grouping the results on the four parameters for the investigated IVISs in a systematic manner, it should now also be possible to obtain at least an impression of how to qualify them. The table below contains the grand averages on the four parameters, i.e., over all tools and environments, for the investigated IVISs.

Table 33 – Grand average of effect sizes for the four systems

Parameter	System A	System B	System C	System D
Subj_R	1.60	2.58	1.54	-
MN_SP	0.82	1.11	0.98	-
HI_ST	0.93	1.19	-	1.04
U_HWT	0.95	1.10	0.42	0.94

These numbers should be viewed as preliminary results only, and on the assumption that the tasks selected from each system are a fair representative of the system. Bearing this caveat in mind, it is apparent that System B was associated with the largest effects, followed by Systems A and System C. Therefore, in terms of its consequences on safety, System B would presumably be the most detrimental. System D, which was only evaluated in a single study, is hard to classify. It would probably take an intermediate position between Systems A and B.

11.6. The non-mandatory PDT parameters

A similar analysis as the one described above was done for the PDT measures recorded by some partners HASTE. The number of hits and reaction time to PDT were both found to be significant (at the 10% level) in 5 out of 6 studies. Formally, these parameters therefore failed the very first criterion for including them in the further steps of the meta-analysis, which is that they should have been shown to be significant in at least 90% of the studies. Nevertheless, in the studies in which they were significant the two parameters showed good discriminative power.

Table 34 – Discriminative power of PDT_HIT and PDT_RT in studies that found them significant (5 out of 6)

Parameter	Consistency index	Average effect size	Range of effect sizes	Score
PDT_HIT	1.00	.93	.99	1.91
PDT_RT	1.00	1.16	1.28	2.44

Formally, we would have to withhold a judgment on PDT-related measures because (1) they could not be measured in as many studies as the other parameters; and (2) they just missed the requested level of overall significance in the studies that used them (1 of 6 non-significant). However, in practical terms the PDT-related parameters appear to be on a par with the four parameters identified above.

11.7. The non-mandatory PRC parameter

Percent Road Centre (PRC) was measured in the two VTEC field studies and at Leeds. The meta analysis showed that the measures were significant in all three, and consistent in two studies. However, this parameter is a special case in that it specifically captures visual activity rather than driving performance. Thus, we would have to look into the modality of the underlying tasks in order to see what 'consistency' would mean in this case: see the next paragraph.

11.8. Effect sizes of parameters in relation to task modality

The table below presents the effect sizes in relation to modality type. In order to make a fair comparison effect size now has been taken to be 0 in case of a non-significant result.

Table 35 – Effect sizes in relation to modality type

	Visual	Visual-manual	Cognitive
Subj_R	-2.19	-2.49	-0.97
MN_SP	-0.62	-0.84	-0.54
HI_ST	0.84	0.88	0.71
U_HWT	0.98	1.00	0.91
PDT_RT	0.81	0.82	0.60
PRC	-2.20	-1.91	0.65

Two things are apparent from this table:

(1) For most parameters task modality does not make a difference.

(2) The exceptions are Subj_R and PRC. The latter is maybe not surprising, since PRC specifically captures visual activity (in case of cognitive tasks PRC even goes the opposite way, explaining the non-consistency in the results mentioned in the preceding paragraph). The former appears not to have an easy explanation.

11.9. Choice of tool and environment

The results of the meta-analysis as presented can also serve to reach a judgment about the best – in the sense of most discriminative – environment in which to evaluate a system handed over for assessment.

11.9.1. Lab versus simulator

This comparison can be made by assessing how the four measures, identified as the overall most discriminative, performed in the Leeds laboratory (LabSim) vs. simulator studies with System B. The following observations were made:

- The HI_ST parameter only showed an effect in the full simulator environment. Indeed, all steering-related measures showed inconsistent results in the LabSim.
- The car-following parameters were more powerful in the full simulator environment.

It can therefore be concluded that the full simulator was more discriminative than the LabSim.

11.9.2. Simulator versus field

The comparison here is for both Systems A and B, for which the results of several experiments in full-blown simulators can be compared to those of several field experiments. The following observations were made:

- HI_ST, the most prominent measure of steering activity, had a more noticeable average effect in the field experiments. On the other hand, its range of effects, the second element of its discriminative power, appeared to be slightly better in the simulator studies.
- The MN_SP measure looked slightly better in its discriminative power in the simulator than in the field studies.
- The same appears to be the case for the Subj_R parameter.
- No car-following measures were obtained in the field, and so the simulator studies provided something extra here that the field study could not. As a matter of fact, this is a category of parameters that is very difficult to handle in field studies, because it is hardly ever possible to present well-defined and controlled car-following situations to drivers. Clearly, this is an inherent characteristic that is to the benefit of simulators (and that, as we have seen, yields a category of dependent variables that adds essence to an evaluation procedure).

All in all, the conclusion must be that the field studies did not add substantial insights that the simulator studies missed.

11.9.3. Lab versus field

System C was tested both under lab and field conditions, from which it appeared that:

- Car-following effects were present in the lab, while as outlined above they were not – measured in the field. However, the effect sizes for these parameters were rather small.
- On the other hand, all other essential parameters (Subj_R, MN_SP, and HI_ST) showed clearly better discriminative performance in the field.
- In all, the results of the field study should probably be regarded as slightly superior in their discriminative capacities than the lab.

11.9.4. Conclusion preferred tools and environments

A full-blown driving simulator appears to be the preferred evaluation environment for an IVIS about which little a priori information is known. It is the simulator that provides the most in terms of discriminative parameters that should be applied, i.e., the four parameters identified above.

On the basis of the results from WP2, the conclusion was already drawn that it is a rural environment that should be presented in the simulator. Thus, combined use of a simulator in the rural road seems to be the preferred set-up when assessing the safety implications of an IVIS.

12. Expert assessment of Systems

While the one of the main aims of HASTE was to explicitly focus on driver-behaviour related parameters for the safety assessment of systems there are clearly alternative approaches, specifically those that rely on expert assessment. The TRL checklist (Stevens, et al., 1999) is one of the most prominent of these. The systems used in HASTE WP3 were therefore also tested with the TRL checklist, to establish whether the two approaches could be compared.

12.1. TRL checklist

The TRL Checklist comprises six major categories on which a system is to be evaluated by an expert: Documentation; Installation and integration; Driver input controls; Auditory properties; Visual properties of display and display screen; and Dialogue between user and system. This is very much the categorization also followed in the (slightly) later European Statement of Principles. The 'Dialogue etc.' category has a subcategory that directly checks the 'Safety-related aspects of information.' All other safety-related aspects are assumed to be implicit in the aspects to be judged within a given category, which would traditionally have been classified more as 'usability' aspects. The scoring proforma can be received by contacting TRL.

12.2. Procedure

The TRL checklist was completed for every system by the experimenter who had been using that system in a HASTE experiment.

12.3. Results

In this section, the summary of findings for each assessment will be given, so as to see whether the findings could be linked to the experimental results.

12.3.1. System A

12.3.1.1. Summary for VTI

Serious Concerns / reasons

System A has serious interface problems. The menus, layout design and visual properties are cluttered and indistinct. The information presented is excessive. The system is a little slow and demands several keystrokes to complete a function, which can cause frustration and loss of focus for the driver.

Minor Concerns / reasons

The system has problems with the colouring of the display and its functions. The colours are not conventional and unambiguous. The Swedish language adaptation is not perfect neither in the visual nor the auditory mode.

Overall Assessment

System A has serious interface problems.

Additional Comments

The remote control is good – it has few buttons, which are distinct from each other

Recommendations

You should be able to choose auditory instructions.
The icons are not self explanatory – these should be redesigned
There should be explicit feed-back prompted by driver action
The menus should be clearer
Some soft buttons are excessive
Sequences that are often recurring should be "prioritised" in the system, e.g. choice of destination should be made accessible with very few keystrokes.

12.3.1.2. Summary for TC

Serious concerns/reasons

- Input and feedback delays are too long
- It is very easy to get lost in menus, causing frustration
- After inactivity in some screens, systems will default to map
- Use of remote while driving is physically/visually distracting
- Map is hard to read (red/orange colour combinations, looks washed out)

Minor concerns/reasons

- Blue/yellow colour combinations
- Icons/aesthetics are confusing and not pleasing

Overall assessment

- Interface feels clumsy. Navigating is cumbersome and confusing (double button presses cause user to skip steps unintentionally)

Additional Comments

None

Recommendations

- Menu hierarchy could be improved to be more intuitive, as you often do nt know where a selection will take you
- Better 'bread crumb' trail so that users know where they are
- Quicker response times and feedback

12.3.2. System B

12.3.2.1. Summary for VTEC

Serious Concerns / reasons

- The only possible position for the IVIS (due to the suction cup solution) is a really poor ergonomic solution. It also obstructs the vision ahead.
- Usage of the "pen" means difficulties to interact with the system; the graphic demands high precision and the pen means that one hand is occupied in an emergency situation.
- The auditory feedback (turn directions) is barely audible, which instead makes visual attention to the display important.
- System response time creates confusion. It is too long and makes the user wonder if he/she even succeeded to point at the right spot in the graphical interface. This requires great visual attention.
- Menus are accessible during driving which may affect attention to the road.

Minor Concerns / reasons

- The size of the display is a little small to display complex maps.
- The attaching device is not stable enough on uneven surfaces.
- The physical controls do not communicate their function clearly enough (legibility of symbols is poor), and feedback on activation is not clear enough.

Overall Assessment
The instruction says that the system should not be operated during driving. However, since it is possible to use it during driving, drivers will probably do so, hence the concerns above. The system is consistent with current standards, which is good. But it is still not good enough to be used during driving (i.e. to interact with during driving).

12.3.2.2. Summary for TNO

Serious concerns / reasons
First I want to stress that the supplier of B clearly states not to operate the system while driving. Therefore, I assume that drivers plan their route before they start driving. So I did not look at these issues. However, while driving there are few options that the driver can use, like zooming in and out or changing the way the direction is indicated. This is done by pressing a small button under the screen. This button is clearly located but is too small to be used for different options. For example, by pressing the left side of the button the driver would go to the menu. However, he should never get there while driving because the only way out is by pressing a small icon with the stylus. Pressing the same button at the top or bottom allows zooming in or out of the display, while pressing the right side of the button changes the manner in which the direction indication to the driver is given.

The auditory output is far too low. However, the auditory direction indications are very useful.

The system does not react very quickly when zooming in and out.

Minor Concerns / reasons
Colour coding can be (much) better

Overall assessment
Nice PDA and route guidance looks rather nice. However, in its present form it should not be used in the car unless there is a passenger to operate the system while driving. The volume level is too low and the control for zooming in and out may lead to too many errors.

Additional comments
I stress once more that I only looked at the controls that may be used while driving. Of course there are other features like entering a destination. If we assume that it is normal for a driver to enter a destination while driving I think it will be clear that this system (or any other system that needs visual guidance to select information) is unsuitable. Another problem with the present system is that it is not a dedicated route guidance system and, therefore, offers too many functions while driving.

Recommendations
The control for adjusting the zoom level of the direction indicator must be adjusted. These should be separated to avoid errors. The volume level should be increased. The route

guidance system is GPS-based and therefore it detects movement. If so the software might be programmed such as to exclude the possibility of using unwanted functions while driving.

12.3.3. System C

12.3.3.1. Summary for MINHO

Serious Concerns / reasons

- Brightness influence – difficult control during the day/illuminated environments

Minor Concerns / reasons

- Poor feedback to driver about the accuracy of the task

Additional Comments

System simulation – hard to assess various items (for example inserting information into the system was very limited)

Recommendations

None

12.3.3.2. Summary for VTT

Serious Concerns / reasons

None

Minor Concerns / reasons

- Reflection (monitor) in bright sunshine,
- Control layout –touch screen (visual information only),
- Momentary uncertainty of selected response button

Overall Assessment

None

Additional comments

- Message divided in several rows, but only two rows are within sight at the same time
- Comprehensibility of messages when using road numbers: general problem of these type of messages

Recommendations

None

12.3.4. System D

12.3.4.1. Summary for TC

Serious Concerns / reasons

- Use of stylus while driving is physically/visually distracting
- Very difficult to perform the precise selections required by the system while driving
- Screen is too small to present the amount of information required to be an in-vehicle navigation system

- Graffiti letters are difficult to do, causing user to enter the wrong letter repeatedly (frustration occurs)

Minor Concerns / reasons

- Colours cause legibility problems
- Auditory output should be louder

Overall Assessment

- Nice device for use by passenger (lots of additional PDA features), but definitely too small and distracting to use safely while driving

Additional comments

- Each type of task often requires multiple actions (e.g., entering an address requires user to tap, use Graffiti and scroll)

Recommendations

- Reduce number of steps required to perform task
- Use a keyboard layout by default instead of Graffiti. It is not clear to user how to pull up the keyboard for data entry
- Use less detailed icons and more vibrant colours

12.4. Comparison with experimental results

The checklist assessment yields judgements that are naturally qualitative. For that reason alone they are hard to compare with the quantitative results of the HASTE studies. On the other hand, the checklist judgments may lead to conclusions – on the criticality of driving with the system, in particular – that can be compared to those one would reach on the basis of quantitative statements.

If we look for such 'strong' statements in the summaries given above we find the following:

- System A is said to have a serious interface problem. While this is not followed up by stating that it should not be used whilst driving, this is what one should be expected to follow from it.
- System B is explicitly condemned by one assessor as not being usable while driving.
- System D is judged to be used safely only by passengers.
- System C is received a relatively kind judgment

Thus, a rank ordering of systems in terms of being fit for use while driving, perhaps in terms of a simple pass-fail criterion, would possibly look as follows:

- Systems B and D are worst, followed closely by System A.
- System C is relatively acceptable.

Thus the ordering would be (B,D); (A); (C).

The rank ordering of systems on the basis of the HASTE experiments, as summarized in Table 32, was roughly as follows: (B); (D); (A); (C). Although there are all kinds of difficulties associated with comparing such wildly different and ordinal assessment dimensions, it appears that the two rank orders are quite similar, i.e. not blatantly dissimilar.

12.5. Discussion

The comparison between the HASTE results and the TRL checklist results had to be based on very rough blocks of material, each of them summarizing a detailed body of measurements, expert ratings, specifications, etc. Therefore, the results of the comparison must be taken with even more caution than is already usual in this type of research. Taking this in mind, however, the results of the comparison are encouraging at what might be called the zero-order level. The result may then point the way to an approach in which checklists and behavioural measurement procedures are seen as complementing each other.

13. Recommendations for a future test regime

The outcome of HASTE should be a prototype of a test regime for the assessment of In Vehicle Information Systems. A test regime could be defined and used in a variety of ways. In HASTE, it has been suggested that the test regime should be used both for pre-deployment phases (e.g. throughout early design phases and iteratively during the system development) as well as for safety validations of final systems. The test regime could be presented in a variety of ways, including code-of-practice, ISO standard, Pass/Fail criteria or as a testing procedure within the primary new car assessment programme (PNCAP).

Test regimes have been defined differently in different projects and working groups. This often depends on the fact that the intended use differs. These include factors such as purpose of the study: formative or summative, system and task characteristics: visual or auditory output modalities, as well as hypothesis to be tested and available resources.

HASTE's intention has been to recommend a test regime that is as cost effective and valid as possible, to be used both as a *pre-deployment* regime as well as for *final verification* of IVIS tasks (Roskam et al., 2002). The regime is intended to be used both by governmental organizations as well as OEMs.

Another project with similar intentions as to HASTE is CAMP (The Crash Avoidance Metrics Partnership). The primary difference is in that CAMP incorporates an anchor-task (radio tuning) with which they use as their safety criteria and with which they compare all other tasks. Most partners involved in the project are vehicle manufacturers and the measures and test procedure within the regime are meant to be used by the OEMs *throughout the design phase* (Shulman et al. 2004). The regime should be practical, meaningful, repeatable during product development e.g. in order to decide which in-vehicles task a driver might reasonably be allowed to access and perform while driving.

The ADAM project (Advanced Driver Attention Metrics) has mainly worked on the Lane Change Task (LCT), the Occlusion method, the PDT, the Combined Probe test and the static driving simulation. Thus, the effort has been on finding fast and efficient verification procedures to assess effects of IVIS. The focus has been on direct measurement of driver distraction mainly with surrogate methods.

One issue to consider when constructing a test regime is how it should be wrapped up. There are several possible ways to go. Some examples could be:
(1) ISO
(2) Primary NCAP procedure
(3) Define strict pass/fail criteria for e.g. legislative organisations
(4) Code of practice

In order for the test regime to be as constructive as possible it would be of particular interest to have a recommended tool which can be seen as a tool for vehicle and system manufacturers.

If the test regime is to serve as a tool during the design phase the focus of the test regime should ideally be on tasks rather than systems. However, to define what a task is could be problematic. In 'Statement of Principles, Criteria and Verification Procedures on Driver

Interactions with Advanced In-Vehicle Information and Communication Systems', version 3, by AAM (Alliance of Automobile Manufacturers) a task is defined in the following way:

> *Task is defined as a sequence of control operations (i.e., a specific method) leading to a goal at which the driver will normally persist until the goal is reached. Example: Obtaining guidance by entering a street address using the scrolling list method until route guidance is initiated. A goal is defined as a system state sought by a driver. Examples include: obtaining guidance to a particular destination; greater magnification of a map display; determining the location of a point of interest; and canceling route guidance.*

Most projects similar to HASTE assess systems on the basis of task. However, one could also consider an assessment of a system in a more naturalistic way. For example, a navigation system might be better assessed while in use for a longer period and the criteria would instead be that the system should not exceed more than x number of glances longer than 2 seconds.

13.1. Dependent measures

The primary behavioural measures to be included in the test regime as a result of the meta-analysis have identified above. As far as these results go, they have shown that these four parameters are sufficient (and necessary) to evaluate any system that is offered for assessment.

Cognitive distraction has been shown to be very difficult to capture when tasks are "auditory/cognitive only" (e.g. route guidance measures). However, based on the validation results in WP3, PDT reaction time (pdt_rt) along with the measure indicating gaze concentration (PRC) are sensitive for cognitive load.

In the table below some things to consider when specifying a final test regime are presented.

Table 36 – SWAT analysis of the selected dependent measures

	Strengths	Weaknesses	Opportunities	Threats
Subjective Rating (Subj_R)	Fast, cheap	Subjective, perceptions of driving performance may not be the same as actual performance	Different rating scales can be developed	Manipulation of data from instructions to participants
Mean Speed (MN_SP)	Easy signal to measure, on-road/sim	Safety interpretation of speed effects - speeding vs slowing down. Speed needs to stabilize to normal level again between tasks.		Slowing down may not be a relevant criteria for classification as unsafe.
Steering (HI_ST)	Easy signal to measure, on-road/sim, relevant			May reflect increased effort or sensitivity to steering error and not necessarily represent a threat to traffic safety.
Minimum headway (U_HMT)	Relevant	Needs lead vehicle, needs distance sensor		Resource demanding in real traffic
Percent road centre (PRC)	Measures perceptual performance, relevant, high face validity, easy to calculate (much easier than glance measures)	Currently expensive hardware, Not calculated in all studies (Haste), needs eye tracker	Can be developed as inexpensive, easy to use tool. Can easily be used in product development.	
Peripheral Detection Task (PDT_RT; PDT_HIT)	Measures perceptual performance and reaction time, relevant, high face validity, easy to calculate	Somewhat intrusive, may effect other measures. Not calculated in all studies and not sufficient statistical reliability (Haste)	Can be augmented with other event detection stimuli	

With those things in mind, the final recommendation of dependent measures in a final test regime would be:
(1) Subjective rating of driving performance (Subj_R)
(2) Mean speed (MN_SP)
(3) High frequency steering (HI_ST)
(4) Minimum Headway (U_HWT)
(5) PDT reaction time (PDT_RT)
(6) Percent Road Centre (PRC)

13.2. Scenarios

The rural road type used in the WP 3 experiments showed sufficient discriminative power, in terms of the parameters emerging from the meta-analysis, so that it should be considered as the recommended one for the test regime. A result from the WP 2 experiments to be mentioned in this respect is that this road type need only comprise *straight* sections: it was found in the WP 2 studies that, while there may be main effects associated with road curvature, there were no massive interactions between straight/curved section effects and IVIS task level.

13.3. Practice

Before assessment the participants should be well acquainted with:

- The test vehicle and test environment: Each participant will practice driving for approximately 10 minutes in the environment to get well acquainted with the vehicle and the environment.
- The tasks: both in static mode (when this is possible considering the nature of the tasks) as well as when driving (in order to avoid results to be representative of naïve users performance only). Each subject will practice all tasks in static setting (sitting in vehicle) just before the experiment drive with that specific system. The practice will last for approx. 5-10 minutes per system (depending on the number of tasks assessed) where subject should feel that he/she will be able to recall what he/she should do when instructions are given.

13.4. Experimental design and Analysis

The experimental design as well as the statistical analysis on parameters to be done in the test regime would have to be the same as done in the WP 3 studies. It is also important to define how many tasks are included in the same analysis. However, this has not been done yet.

13.5. Participants

As explained above, no more than 15 participants would be needed to demonstrate IVIS effects by the selected behavioral measures if such effects existed. Average drivers defined as in the WP3 experiments should be included.

13.6. Task selection

During a formative assessment, the goal should be to assess as many tasks as possible throughout the design and development process of the system. For a summative assessment further discussion is needed to decide on how tasks are chosen. It is important that system manufacturers will not begin to sub-optimise their systems.

13.7. Interpreting and weighting the results from the single measure/methods into a final assessment.

In order to put the results for the separate behavioural parameters in one bin, so to say, and to derive a common overall safety effect from them, we need the relationships existing between behavioural parameters and risk. These are to be derived from the Leeds Stated Preference study.

13.8. Summary of recommendations

The following can be summed up from the results in WP3:

1. The test regime can be used both in the design stage as well as in final assessment
2. Number of subjects: only 10-15
3. Age between 25 and 50, M&F, sufficient driving experience (10.000 km annually, at least 5 yrs licence)
4. Environment: at least medium-range simulator; rural road type
5. Duration per task: about 10 min
6. A single baseline ride is required (10 min)

References

Alliance of Automobile Manufacturers (2003 June 17; Version 3). Statement of Principles, Criteria and Verification Procedures on Driver Interactions with Advanced In-Vehicle Information and Communication Systems, Washington, D.C.: Alliance of Automobile Manufacturers

Brookhuis, K.A., De Vries, G. and De Waard, D. (1991). The effects of mobile telephoning on driving performance. Accident Analysis and Prevention, 23, 309-316.

Godthelp, J. (1984). Studies of human vehicle control. (PhD thesis). Institute for Perception TNO. Soesterberg, The Netherlands.

Haakamies-Blomqvist, L., Östlund, J., Henriksson, P., & Heikkinen, S. (2000). Äldre bilförare i simulator – en valideringsstudie (VTI Rapport 464). Linköping: Väg och transportforskningsinstitutet.

Harbluk, J.L., Lalande, S. (2005). The impact of auditory tasks on visual detection: Email reading. Accepted for Presentation at Driving Assessment, June 2005.

Harms, L. (1996). Driving performance on a real road and in a driving simulator: Results of a validation study. Paper presented at the Vision in Vehicles V, Glasgow.

JAMA (Japan Automobile Manufacturers Association) guidelines, presentation Nagoya, 2004

McLean, J.R. and Hoffmann, E.R. (1975). Steering reversals as a measure of driver performance and steering task difficulty. Human Factors, 17, 248-256.

Olsson, S. (2000). Measuring driver visual distraction with a peripheral detection task. Master Thesis. ISRN LIU-KOGVET-D-0031-SE. Linköping University, Department of Behavioural Science.

Olsson, S., Burns, P.C. (2000). Measuring Driver Visual Distraction with a Peripheral Detection Task. NHTSA Internet Forum on Driver Distraction**.** http://www-nrd.nhtsa.dot.gov/departments/nrd-13/driver-distraction/Topics033080006.htm#A6

Risser, R. (1985). Behaviour in traffic conflict situations. Accident Analysis and Prevention, 17(2), 179-197.

Roskam, A. J., Brookhuis, K. A., de Waard, D., Carsten, O. M. J., Read, L., Jamson, S., Ostlund, J., Bolling, A., Nilsson, L., Antilla, V., Hoedemaker, M., Janssen W., Harbluk, J., Johansson, E., Tevell, M., Fowkes, M., Engström, J. and Victor, T. (2002). Development of Experimental Protocol. Deliverable 1. EU project HASTE, project no. GRD1/2000/25361.

Shulman, M., Deering, R. K. (2004) Second Annual Report of the Crash Avoidance Metrics Partnership, April 2002 - March 2003

Van Winsum, W., Martens, M. & Herland, L. (1999). The effects of speech versus tactile driver support messages on workload, driver behaviour and user acceptance. TNO-report, TM-00-C003. Soesterberg, The Netherlands.

Victor, T., Johansson, E. (2005). Using eye movements to measure driving information loss while performing secondary tasks. Manuscript submitted for publication.

Östlund, J., Nilsson, L., Carsten, O., Merat, H., Jamson, H., Jamson, S., Mouta, S., Carvalhais, J., Santos, J., Anttila, V., Sandberg, H., Louma, J., de Waard, D., Brookhuis, K., Johansson, E., Engström, J., Victor, T., Harbluk, J., Janssen, W., & Brouwer, R. (2004). HMI and Safety-Related Driver Performance, Deliverable 2. EU project HASTE, project no. GRD1/2000/25361.

Appendix 1: List of indicators

Complete indicator name	Abbreviation
Self-reported driving performance	
self-reported driving performance	subj_r
Longitudinal control	
speed [km/h]	mn_sp
speed variation [km/h]	st_sp15
speed variation calculated with a sliding window of 15s [km/h]	st_sp30
speed variation calculated with a sliding window of 30s [km/h]	st_sp
speed change [km/h]	d_sp
min speed [km/h]	u_sp
min Time To Collision [s]	u_ttc
mean of TTC minima [s]	mn_ttc
mean distance headway [m]	mn_hwd
mean of distance headway minima [m]	u_mn_hwd
distance headway variation [m]	sd_hwd
min distance headway [m]	u_hwd
mean time headway [s]	mn_hwt
proportion of time spent between x and y of the total time headway	hwt_x_y (%)
mean of time headway minima [s]	u_mn_hwt
time headway variation [s]	sd_hwt
min time headway [s]	u_hwt
Lateral control	
lateral position [m]	mn_lp
lateral position variation [m]	st_lp
lateral position variation calculated with a sliding window of 15s [m]	st_lp15
lateral position variation calculated with a sliding window of 30s [m]	st_lp30
mean TLC minima [s]	mn_tlc
min TLC [s]	u_tlc
proportion of time outside lane, lanex [%]	lnx
1 deg reversal rate [1/minute]	rr_st1
2 deg reversal rate [1/minute]	rr_st2
high frequency steering [deg]	hi_st
rapid steering wheel turnings > 10deg [1/minute]	rswt_10
rapid steering wheel turnings > 40deg [1/minute]	rswt_40
rapid steering wheel turnings > 70deg [1/minute]	rswt_70
Workload measures	
glance frequency	n_gl
glance duration [s]	mn_gd
glance duration variation [s]	sd_gd
gaze angle variation [deg]	st_ga
total glance duration	tot_gl_t
glance duration as a percentage of total task duration	tot_gl
PDT measures	
hit responses between 200 and 2000 ms (%)	pdt_hit
reaction time to hit responses (s)	pdt_rt
responses faster than hit threshold (%)	pdt_cheat
missed responses (no response) (%)	pdt_miss
Other	
observer ratings	oObs_r
subjective workload	subj_wl
task length	task_l

Appendix 2: Detailed report on included indicators

Overview of indicators

The table below presents a summary of the included indicators in the WP3 experiments.

Table 37 – Overview of measures

Lane-position and TLC measures	
The lane position is defined as the distance between the right hand part of the front right wheel to the left part of the right hand lane marking [m]. When the right hand lane marking is crossed, the lateral position becomes negative. The lane boundaries are defined as the inner edges of the lane markings. Left-hand wheel and left-hand lane marking are used in the UK. Thus, when the vehicle crosses lane boundaries, discontinuities appear in the data. For the rural road data, the lane position data was recalculated so that the right lane marking was always used as the reference (which removes the discontinuities).	
Variable short name	**Description**
mn_lp(m)	Mean lane position (in metres).
st_lp(m)	Standard deviation of lane position.
st_lp15(m)	Standard deviation of lane position calculated with a sliding window with the length of 15s. The sliding window is used to minimize the effect of different IVIS lengths and facilitate the comparison between IVIS.
st_lp30(m)	Standard deviation of lane position calculated with a sliding window with the length of 30s
mn_tlc (s)	Mean of the TLC local minima values.
lnx(%)	Proportion of time that any part of the vehicle is outside the lane boundaries.
pr_tlc(%)	The proportion of Time-to-Line-Crossing (TLC) local minima values less than 1 second.
Steering wheel measures	
Variable short name	**Description**
rr_st1(1/minute)	Reversal rate, where 1 is the amplitude threshold for reversals. The reversals should be computed before segmenting the data and normalised with the task length in minutes.
rr_st3(1/minute)	Reversal rate, where 3 is the amplitude threshold for reversals. The reversals should be computed before segmenting the data and normalised with the task length in minutes..
hi_st (%)	The high frequency component of steering is defined as the ratio between the power of the 0.3-0.6 Hz component and all steering activity. High_steering shall be calculated as following. The steering signal is filtered with a second order Butterworth low pass filter with cutoff frequency 0.6 Hz. This results in the "all steering activity" signal. The signal is further filtered with a 0.3 Hz second order Butterworth high pass filter, which results in the high frequency steering component. The power of the signals is calculated as the root mean square.
hi_st2	High frequency component of steering wheel movements. The measure is computed by means of a power spectral analysis in the 0.3-0.6 Hz interval. The final hi_st2 value is the natural logarithm of the power in this interval. This measure is defined and calculated by VTEC in WP2.
en_st(-)	Steering Entropy, implemented according to Nakayama et al. (1999).
rswt_5 (1/minute)	Number of rapid steering wheel movements (deg./s), implemented according to the specification provided by VTT, see definition below. The number (5, 10, 20, 40 or 70) represents the velocity thresholds(s) for rapid steering wheel movement. The number of rapid steering wheel movements are then divided by the task length in minutes. The rapid steering wheel movements in rswt_5 are larger than 5 deg/s.
rswt_10(1/minute)	Number of rapid steering wheel movements > 10 deg/s.
rswt_20(1/minute)	Number of rapid steering wheel movements > 20 deg/s.
rswt_40(1/minute)	Number of rapid steering wheel movements > 40 deg/s and <= 70 deg/s.

rswt_70(1/minute)	Number of rapid steering wheel movements >70 deg/s.
Speed and Headway-related measures	
Variable short name	**Description**
mn_sp(km/h)	Mean speed (km/h)
u_sp(km/h)	Minimum speed.
st_sp(km/h)	Standard deviation of speed.
st_sp15(km/h)	Standard deviation of speed calculated with a sliding window with the length of 15s. The sliding window is used to minimize the effect of different IVIS lengths.
st_sp30(km/h)	Standard deviation of speed calculated with a sliding window with the length of 30s.
d_sp(km/h)	Speed change from start to end, divided by time duration. The values are calculated by fitting a linear function to the speed signal for each IVIS segment, where d_sp equals the difference between the initial and the end value of the linear function.
sd_hwd(m)	Standard deviation of distance headway (in metres).
mn_hwt(s)	Mean time headway (in seconds)
mn_hwd	Mean value of distance headway
sd_hwt(s)	Standard deviation of time headway
u_hwt(s)	Minimum time headway
hwt_0_1(%)	The proportion of time spend between 0 and 1s of the total headway time.
hwt_1_2(%)	The proportion of time spend between 1 and 2s of the total headway time
hwt_2_3(%)	The proportion of time spend between 2 and 3s of the total headway time
hwt_3_4(%)	The proportion of time spend between 3 and 4s of the total headway time
hwt_4_5(%)	The proportion of time spend between 4 and 5s of the total headway time
hwt_5_6(%)	The proportion of time spend between 5 and 6s of the total headway time
hwt_6(%)	The proportion of time spend above 6s of the total headway time
Eye Movements	
General: The eye movement measures could be divided into two general types: (1) Glance-based measures and (2) basic ocular measures. The former represent the properties of glances to a defined target, e.g. frequency and duration. Thus, glance based measures are only applicable to the tasks with a visual part (i.e. not possible for auditory/cognitive tasks). Moreover, a baseline condition does not make sense in this case, since no glances are expected to a task when no task is performed. Basic ocular measures operate on the gaze data (before it is classified into glances) and, thus, a baseline condition should be included. Basic ocular measures are mainly useful for quantifying the effects of cognitive load. The basic ocular measures to use is here is st_ga and PRC(%).	
Variable short name	***Description***
st_ga(deg)	Standard deviation of gaze angle. This is computed as the standard deviation of the combined pitch and yaw components of the gaze direction (gacomb), where gacomb is given by. **Error! Objects cannot be created from editing field codes.**
n_gl	Number of glances to IVIS to complete task. Often referred to as Glance Frequency.
tot_gl	IVIS glance duration percentage of total task duration. In the literature, this measure is often referred to as total glance time when expressed in seconds.
tot_gl_t	Total glance duration
PRC(%)	Percent road centre. This is computed as the percentage of driver gaze fixations within one minute that fall within a specified area representing the road centre. The road centre area is defined as a circle with a radius of 8 degrees around the road centre point. The road centre point is the Mode or most frequent fixation position (see Victor and Johansson, 2005, for details).
mn_gd(s)	Mean duration of single glances to IVIS.

Task measurement	
Variable short name	**Description**
task_l (s)	Task duration

PDT measures	
Variable short name	**Description**
pdt_rt (s)	Response time for response data which are defined as hits.
pdt_hit (%)	CHEAT_THRESHOLD<response<RESPONSE_THRESHOLD Response time (time lapsed between stimulus presentation and response). pdt_hit, pdt_miss and pdt_cheat adds up to 100%.
pdt_miss (%)	No response is given within RESPONSE_THRESHOLD. -1
pdt_cheat (%)	The response is faster than CHEAT_THRESHOLD. -3

Subjective ratings	
Variable short name	**Description**
subj_r	Subjective ratings of own driving performance given on a scale from 1-10.
obs_r	Ratings by experimenter leader in field according to Wiener Fahrprobe protocol
com_t (y/n)	Ratings by experimenter leader in field and sim/lab. Has the participant completed task (yes, no)

Detailed specification of indicators

In the following sections the indicators are specified in more detail.

<u>Speed</u>

Definition

Speed is defined as the travel speed in km/h relative to the road surface [km/h].

Value

Increased speed during the influence of distracting factors has been used as an indicator of decreased speed control. Since increase in speed correlates to increase in accidents, an increase in speed can be used as in indicator of decreased performance. The value of speed as a performance measure is based on the assumption that the measured speed is driver paced. However, in high traffic density speed is affected by other road users to a higher extent than if the traffic density is low. The driver may reduce the speed as a compensatory action due to increased mental load or distraction by e.g. an IVIS. This is however more often used as an indication of increased mental load rather than change in driving performance.

Technical considerations

It should be possible to relate the vehicle's speed to current signposted speed limits. The table below describes requirements for speed data.

Table 38 – Description of speed data

Measurement range	**20 km/h to 180 km/h**
Accuracy	**± 2km/h**
Precision	**2 km/h**
Sampling rate	**100 ms (10 Hz)**

Speed variation

Definition
Speed variation is defined as the speed standard deviation [km/h].

Value
Speed variation is often used as a measure of driving performance for driving on high way and rural road. High variation has been considered as an indicator of poor driving performance that reflects involuntary speed variation; speed instability. Variation is usually calculated as standard deviation. A deficiency of this parameter is that it does not differ between involuntary speed changes and speed variation due to the interaction with other road users or adaptation to the road conditions (curvature, visibility).

Calculation
In WP 3 has the length of the IVIS varied between different participants. To eliminate the effect of IVIS length in the variation measure has a variation standardising procedure been used when the task length is larger than 15 seconds. The standardising procedure in this case is a sliding window technique. The sliding window technique calculates the variation measures, e.g. standard deviation, of the values inside the window for each position of the window. The window is moved one sample at a time. The final variation measurement is calculated from the mean of the calculated window variation measures. The window length used in WP3 is 15 and 30 seconds. See the figure below for at better understanding.

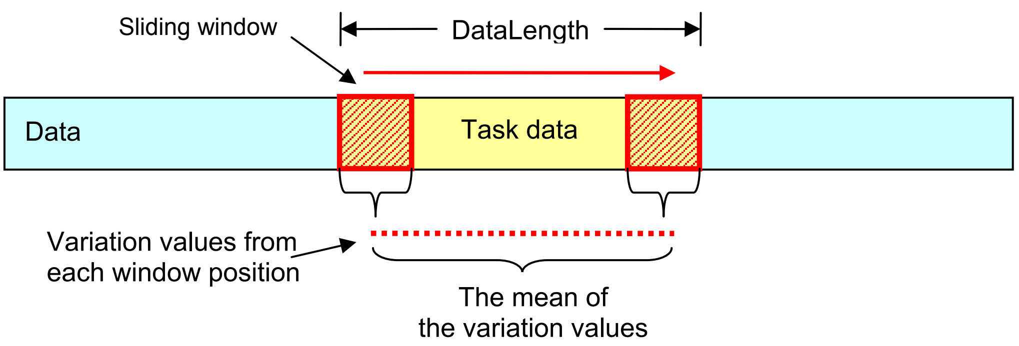


Figure: The figure shows the sliding window technique used when eliminating the effect of different task lengths. The sliding window is moved one sample at a time and the standard deviation is calculated for each window position. The final variation value is calculated from the mean of the standard variation values from each window position.
This technique can only be used if the IVIS length is longer then the window length. If the IVIS length is shorter than is the variation measure calculated from the standard deviation of the signal without any regards to the task length.

Technical considerations

See **Table 38** for data requirements. Speed standard deviation should only be calculated over sections of equally signposted speed limits.

Lateral position

Definition

Lateral position is defined as the distance between the right hand part of the front right wheel to the left part of the right hand lane marking [m]. When the line is crossed, the lateral position it becomes negative. The lane boundaries are defined as the inner edges of the lane markings. Left-hand wheel and left-hand lane marking are used in the UK.

Value

Lateral position reflects strategy. For instance, Brookhuis found that under the influence of sedative drugs drivers drove more towards the relatively safe emergency shoulder compared with a control condition (i.e. they adapted their safety margins).

Technical considerations

Lateral position is used to calculate both lateral position variation and TLC and thus, it is important to get precise data. Target accuracy for on-the-road pilots is set to ± 10 cm. In driving simulators will be at least ten times better. See table below for data requirements.

Table 39 – Description of lateral position data

Measurement range	**From 0 m to lane width**
Accuracy (while driving ; including yaw, roll, pitch, height variations)	**± 10 cm or better when LP is within lane width**
Precision (while driving)	**5 cm or better when LP is within lane width**
Rate	**100 ms (10 Hz)**
Marked line characteristics :	**Well marked White/yellow continuous or dashed lines.**

Lateral position variation

Definitions

Lateral position variation is defined as the lateral position standard deviation [m]. Lateral position variation is derived from lateral position data. When the length of the lateral position signal for each task segment varies have a standardising procedure based on a sliding window technique to eliminate the effect of task length been used. The sliding window technique has been used when the task length is larger than 15s. This technique has been further described in the speed variation measurement definition.

Value

Less lateral control may be observed as an increase in lateral position variation. In several studies, driver impairment (drugs, sleepiness) and time on task have been shown to cause increase in SDLP; the steering control has become less stable. However, SDLP is influenced by take-overs and voluntary changes in lateral position due to road curvature; effects that may or may not be related to driving performance.

Calculation
In WP 3 has the IVIS length varied. The lateral position variation measure has been calculated with the sliding window technique (described in the speed variation section above) to eliminate the effect of the different IVIS lengths. The window length used in WP3 is 15 and 30 seconds. This technique can only be used if the IVIS length is longer then the window length.

For more information on standard deviation of lateral position see Brookhuis et al. (1991).

Lane exceedences

Definition
A lane exceedence (LANEX) is defined as the proportion of a time any part of the vehicle is outside the lane boundary [%]. The lane boundaries are defined as the inner edges of the lane markings. The vehicle boundaries are defined as the outer edges of the front wheels.

Value
LANEX has been used as a measure of lateral control, e.g. by Tijerina et al (1999).

Technical considerations
Lateral position data is required.

Time-to-line-crossing

Definition
Time-to-line-crossing (TLC) is defined as the time to cross either lane boundary with any of the wheels of the vehicle if speed and steering wheel angle are kept constant. As the vehicle approaches the line TLC will decrease until it reaches a minimum. Under “normal” conditions this will occur when the motion of the car is changed from going towards one line to the other. During this change the car will pass a situation where it momentarily will not move toward any of the line but follow the road perfectly this will result in an indefinite or undefined TLC. In order to determine the safety margins we have to look for the TLC minima, which is also the case for TTC. A TLC min value is defined as the min TLC within a TLC waveform. TLC values higher than 20 seconds are ignored. Also TLC waveforms of duration less than one second are ignored.

See Godthelp (1984) for more information.

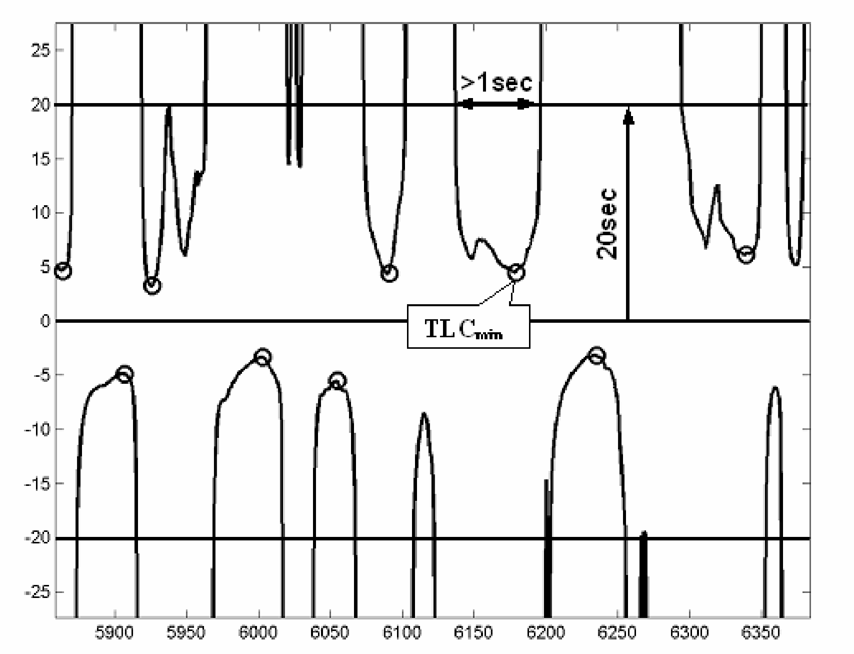


Figure 162 – Principles used to identify relevant TLCmin values as described above.

The graph shows how TLC values less than 20 seconds and TLC wave duration > 1 second are defined. Time to cross the right line is represented by negative values.
Figure:

Included measures are:

- *The proportion of TLC min values less than one second [%]*
- *Mean value of the min TLC values [s]*

Value

Time-to-line-crossing was first proposed by Godthelp and Konings (1981) to describe steering behaviour. According to Godthelp et al, TLC reflects the time available for error neglecting, assumed a fixed steering strategy. In other words; TLC reflects a lateral control safety margin. Godthelp's proposed calculation of TLC included a complex mathematical definition, based on vehicle speed, steering wheel angle, heading angle and lateral position. In this calculation, it is assumed that the road is straight. Van Winsum et al (1996) proposed an alternative method of calculating TLC that considered road curvature. Due to problems achieving all necessary data for exact calculation, approximations are often used based on lateral position and lateral velocity and in simulator studies also lateral acceleration in relation to the road.

Calculations

Within the HASTE project one trigonometric method and two approximations of TLC will be used in the simulator experiment, and one or if possible both approximations in the field experiments. The lane boundaries are defined as the inner edges of the lane markings. The vehicle boundaries are defined as the outer edges of the front wheels.

For the trigonometric method, TLC is based on the vehicle speed and the instaneous circular path of the vehicle. At the intersection of this curve and the edge/centre line distance to line crossing (arc segment length) is calculated. Then this arc segment length is divided with travel speed in order to get TLC. The calculations are based on the instantaneous curve radius. The calculations are described in van Winsum et al (van Winsum, Brookhuis, & de Waard, 1997).

The first approximation (TLC1) assumes that the lateral motion is linear. Thus, TLC is calculated as lateral distance divided by lateral velocity. The lateral distance to line in the TLC calculation will be different depending on which direction the vehicle is moving (towards the right or left line (lane) marker. When the lateral velocity is:

- *Negative (moving to the right), then the lateral distance to right line will be equal to lateral position as previously defined.*
- *Positive (moving to the left), then the lateral distance to left line will be defined as (lane width - (lateral position + vehicle width)),*
- *Zero, then TLC is infinite.*

The second approximation (TLC2) includes road relative lateral acceleration and is calculated as the lateral distance to line divided by the sum of lateral velocity and acceleration. The lateral distance to line in the TLC calculation will be different depending on which direction the vehicle is moving (towards that right or left line (lane) marker. When (lateral velocity + change in lateral velocity) is:

- *Negative (moving to the right), then the lateral distance to right line will be equal to lateral position (see footnote).*
- *Positive (moving to the left), then the lateral distance to left line will be defined as (lane width - (lateral position + vehicle width)).*
- *Zero, then TLC is infinite.*

Technical considerations

Of course, the measurement of lateral position is crucial for TLC. In simulator experiments, this should not be a problem.

Reversal rate

Definition

Reversal rate is defined as the number of changes in steering wheel direction per minute [turns/minute]. A specified angle difference threshold between steering end values is required for the reversal to count. In HASTE Wp3 are the angle difference threshold value *1 and 3 degrees* used. Higher values may be used. See the figure below.

For more information on steering reversals see McLean et. al (1975).

Value

The number of changes in steering wheel rotational direction reflects the frequency of steering corrections, not the magnitude.

Calculation

Reversal rate is calculated as follows. First, the steering signal is low pass filtered with a second order Butterworth low pass filter of cut off frequency 0.6 Hz. Then, local minima and maxima are identified with a peak detection algorithm; within a moving window of 0.8 seconds length, the values have to increase/decrease monotonically towards the centre value to classify the centre value as a local maximum, and of course the opposite to be a minimum. Then the differences between adjacent minima and maxima are calculated. If the difference is larger or equal to the threshold value, then there is one reversal. Note that it is actually the local minimums and maximums that are counted. The number is divided with the IVIS length in minutes for the purpose of normalizing the value to the data length.

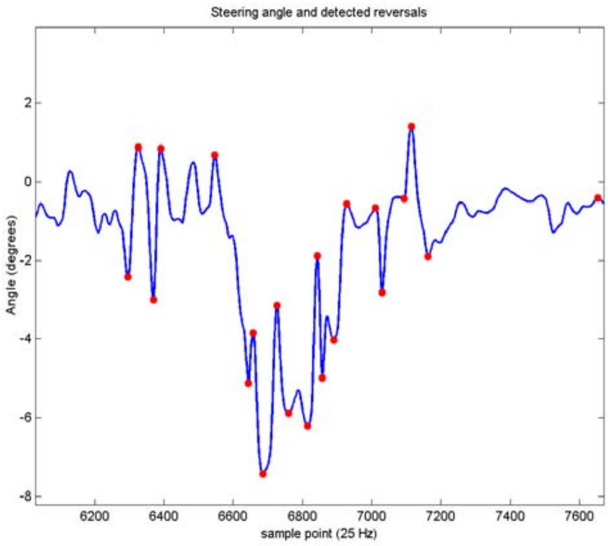


Figure 163 – Steering angle (blue) and reversals (red). Threshold 2 degrees.

Technical considerations

Care has to be taken in the calculation of this indicator so that only driver-induced changes in steering wheel angle are recognised and not artefacts caused by noise. Technical specifications for the measurement of steering wheel angle are listed in the table below.

Table 40 – Description of steering wheel angle data

Measurement range	± 45° or more
Accuracy	± 0.5°
Precision	0.5°
Sampling Rate	100 ms (10 Hz)

Time To Collision, Time headway and Distance headway

Definitions

Time To Collision (TTC) [seconds] is defined as the distance to the lead vehicle (bumper to bumper) divided by the speed difference to the lead vehicle. TTC is only defined if the distance between the vehicles decreases. As with TLC, TTC generates wave formed data. TTC values larger than 15 seconds are ignored. Also TTC wave forms of duration less than one second are ignored.

Time Headway [seconds] to lead vehicle is defined as the distance to the lead vehicle (from bumper to bumper) divided by own momentary travel speed. Distance Headway [metres] to a lead vehicle is defined as the distance to lead vehicle, defined as the distance from bumper to bumper. Time headway values larger than 3 seconds are ignored. Distance headway values larger than 50 metres are ignored.

TTC and headway are measures of longitudinal risk margin. Included measures are:

- Proportion of time of which the TTC is less than 4 seconds. This measure is called *Time Exposed Time-to-collision* (TET).
- The proportion of TTC local minima less than 4 seconds.
- Mean of TTC local minima.
- The proportion of time headway local minima less than 1 second.
- Mean of time headway local minima
- The proportion of distance headway local minima less than 20 metres.
- Mean of distance headway local minima

Value

The closer and faster a participant travels behind a lead vehicle, the less is the chance to manage avoiding a collision in case of the lead vehicle reduces the speed. For a small TTC or headway, the time a participant may be distracted by another task without a highly increased risk of accident is much less than if the time headway is large.

Technical considerations

Requirements on headway data is listed in the table below.

Table 41 – Description of distance headway data

Measurement range	**From 0 to 50 metres**
Accuracy	**± 0.5 m**
Precision	**0.1 m**
Sampling Rate	**100 ms (10 Hz)**

Speed change

Definition

Speed change during attention to IVIS [km/h]. The least-square-method is used for adapting a straight line to the speed profile. Speed change is calculated as the end value minus the start value for the adapted line.

Value

Speed change may better reflect the impact of IVIS on speed behaviour than mean speed since it reflects the adaptation of speed during the time period of distraction.

Steering angle variation

Definition

The steering angle variation is defined as the standard deviation of the steering angle [deg].

Value

This measure is very easy to calculate and may provide a simple measure on steering corrections. It will however only reflect steering corrections if the road is not very curvy. This measure is thus not feasible in urban environments.

High frequency component of steering wheel angle variation (hi_st)

Definition
The high frequency component of steering is defined as the ratio between the power of the 0.3-0.6 Hz component and all steering activity.

Value
As with the standard deviation, the proportion of the high frequency component of steering wheel angle reflects steering corrections. However, this method aims at excluding the effect of open loop behavior and only focus on corrections. McDonald and Hoffman (1980) support that steering corrections are reflected by high frequency components.

Calculation
High_steering shall be calculated as following. The steering signal is filtered with a second order Butterworth low pass filter with cutoff frequency 0.6 Hz. This results in the “all steering activity” signal. The signal is further filtered with a 0.3 Hz second order Butterworth high pass filter, which results in the high frequency steering component. The power of the signals is calculated as the root mean square.

Technical considerations
For all frequency related calculations, the tolerance for artifacts is low, but this should not be a problem since measuring steering wheel angle is not very difficult.

High frequency component of steering wheel angle variation (hi_st2)

Definition
The measure is computed by means of a *power spectral analysis* in the 0.3-0.6 Hz interval. The final value is the natural logarithm of the spectral power in this interval.

Value
As with the standard deviation, the proportion of the high frequency component of steering wheel angle reflects steering corrections. However, this method aims at excluding the effect of open loop behavior and only focus on corrections. McDonald and Hoffman (1980) support that steering corrections are reflected by high frequency components.

Calculation
The hi_st2 value is calculated as follows. The power spectral density of the steering wheel angle data is estimated using Welch’s averaged periodogram method with a Hanning window. The FFT length used when estimating the power spectral density is 256. An integration between 0.3Hz and 0.6Hz in the estimated periodogram gives a value of the spectral power of the steering wheel signal. The natural logarithm of the power spectral value is then calculated.

Technical considerations
For all frequency related calculations, the tolerance for artifacts is low, but this should not be a problem since measuring steering wheel angle is not very difficult.

Rapid steering wheel turnings (RSWT)

Definition
Number of RSWT within the interval a specified interval, e.g. $40 < \text{RSWT} \leq 70$ degrees per minute.

Value
When in highly critical situations, the driver may perform rapid steering wheel turnings to avoid driving off the road or colliding into other vehicles. RSWT may be sensitive to this behaviour.

Calculation
The rapid steering wheel turnings are calculated as follows. The data corresponding to the steering wheel should be measured (in degrees) or sub sampled to 10 times per second (10 Hz). The first step is to compute the absolute value of the difference of two consecutive samples. The second step is to identify values higher than 4.0 among these differences (the value of 4.0 is based on experience). If there are sequences including several consecutive values higher than that threshold (that is usually the case), a maximum value will be identified. This is the actual value of the RSWT. If there are several RSWTs in the sequence of 2 seconds, these are interpreted as one RSWT that has a maximum value of those turnings. Finally, the values are multiplied by 10 to receive the values in degrees per s. The value is divided with the IVIS length in minutes for the purpose of normalizing the value to the data length.

Technical considerations
This indicator requires a baseline measurement of each participant's steering behaviour, which have influence on the design of the study.

Self-reported driving performance

Definition
After each S-IVIS block or at the corresponding road sections in the baseline drive, the participants are asked to report their driving performance. The scale is vertical from 1 to 10, where 1 corresponds to extremely poor and 10 to extremely well. Response is verbal.

Value
This measure is very simple to use and takes advantage of the fact that the drivers in most situations have an opinion about his/her own driving performance. A deficiency is of course that the driver's opinion may not reflect actual risk of accident.The driver may be unaware of the risk, which is supported by the fact that speeding is a quite common behaviour although strongly linked to risk of accident. However this could be considered to be exactly what one would like to know, i.e. do the drivers realise how dangerous they are? If so, the countermeasures are different than if not so.

Eye movement measures – background definitions

Table 42 – Definitions (ISO 15007-1)

Glance	Defined as the moment the eyes start their transition to the target area until the moment it leaves the target area.
Dwell time	Sum of consecutive individual fixation and saccade times to a target in a single glance (i.e. same as 'glance' but not including the transitional saccade(s)).
Fixation	Alignment of the eyes so that the image of the fixated targets falls on the fovea for a given time period.
Saccade	Brief movement of the eyes between fixations.

Glance frequency (n_gl)

Definition

Glance frequency is defined as the number of glances to a target during a pre-defined task, where each glance is separated by at least one glance to a different target.

Value

Depending on the complexity of the task, typically between 1 and 7 glances are needed to acquire and process the information. Because it is related to the overall complexity of the display, it is a highly sensitive measure of visual attention or visual workload

Technical considerations

The SAE J-2396 standard provides the glance definition '*A glance is considered as a series of fixations at a target area until the eye is directed at a new area*'. However, it does not consider fixations, smooth pursuits and saccades which are the bricks forming a glance. For the glance frequency measure, a smooth pursuit is to be classified as a fixation (smooth pursuits are series of short fixations separated by short, to many systems immeasurable, saccades).

Mean glance duration (mn_gd)

Definition

The mean glance duration is the calculated mean duration of all individual glances to a target during one task.

Value

Long glance durations associated with a target may be indicative of high workload demand, posed by that location (or task involving that location). Mean glance duration should always be considered together with glance frequency or total glance duration.

Technical considerations

The same technical considerations as for glance frequency apply here.

Total glance duration (tot_gl_t)

Definition
The total glance duration is defined as the sum of the glance durations of each individual glance to a target during a task.

Value
Long total glance time associated with a target may be indicative of high workload demand, posed by that location (or task involving that location).

Technical considerations
The same technical considerations as for glance frequency apply here.

IVIS glance duration proportion of total task duration (tot_gl)

Definition
IVIS glance duration percentage of total task duration. In the literature, this measure is often referred to as total glance time when expressed in seconds. Computationally, the total glance duration to the IVIS during the task is divided by the total task duration.

Value
A high value indicates that the participants spend a higher proportion of glance time on the IVIS rather than the road ahead. As the complexity of the task increases, the demand for longer and more glances increases.
Whereas mean glance duration give good indications of the demand imposed by the IVIS, this measure puts the glance demand in relation to the time spent on the road (and other targets).

Technical considerations
The same technical considerations as for glance frequency apply here.

Percent road centre (PRC)

Definition
PRC is defined as the percentage of driver gaze fixations that fall within a specified area representing the road centre. The road centre area is defined as a circle with a radius of 8 degrees around the road centre point. The road centre point is the Mode or most frequent fixation point (see Victor and Johansson, 2005, for details). In the HASTE analysis, PRC was computed by first identifying saccades and fixations in the gaze signal. Second, the sum total of all fixations which had a Euclidean distance to the road centre point of less than or equal to 8 degrees, was taken. This sum was divided by the total number of fixations during the task to yield the proportion of fixations directed at the road centre area. This number is then multiplied by 100 to express PRC in percent.

Value
Percent road centre has been shown to correlate with cognitive, visual, and auditory demand. Percent road centre typically decreases from baseline levels when a visual task is performed and decreases further as task difficulty increases. An increase in PRC in relation to baseline

driving indicates a narrowing of the field of vision which is typically seen with high cognitive and auditory load.

Technical consideration

Here the gaze signal was not segmented into glances (PRC is a non-glance based measure) rather PRC was calculated as the proportion of the gaze signal encoded as fixations that were directed at the road centre area. However, using the raw gaze signal without breaking it down in saccades and fixations should yield approximately the same result.

Standard deviation gaze angle

Definition

Standard deviation of gaze angle is computed as the standard deviation of the combined pitch and yaw components of the gaze direction (ga_comb) of all fixations for a given task, where ga_comb is given by:

$$ga_{comb} = \sqrt{ga_{pitch}^2 + ga_{yaw}^2}$$

Next the standard deviation of the resulting ga_comb values is taken.

Value

The standard deviation gaze angle provides a measure of the dispersion of fixations for non-visual tasks. For the gaze concentration phenomenon the standard deviation gaze angle is reduced. The measure gives a higher resolution value of the dispersion than PRC.

Technical consideration

Standard deviation gaze angle is a non-glance based measure and does not require glance segmentation of the gaze signal. Moreover, since the IVIS location strongly affects the standard deviation gaze angle measure it is to be used in baseline driving and for cognitive and auditive tasks.

PDT performance

Definition

PDT performance is defined as how well the participant manages the PDT task. Included measures in WP3 are:

- Pdt_rt
- Pdt_hit
- Pdt_miss
- Pdt_cheat

The percentage measures pdt_hit, pdt_miss and pdt_cheat adds up 100%.

Appendix 3: Sliding Window Standardising Technique

Problem

Some measures, e.g. Lateral Position variation, Time Headway variation, Distance Headway variation and Speed variation, are accumulative with respect to how much (time duration) of data is used. An extreme situation is if a really distracting task, active during 5 seconds, is compared to a no task condition of 30 seconds duration. The lateral position variation may very well be less in the distracting task than in the no task condition. What is found here is only an effect of time duration and thus contaminating the data. A standardising procedure to get eliminate the effect of task length is necessary.

Approach

The approach is to identify those measures affected by this problem, and define minimum lengths of data that is required to achieve reliable measures. If less data is available than the minimum data length, the measure is not calculated. For “too” short tasks one will thus not be able to calculate the measures concerned simply due to the fact that the dynamic driving behaviour cannot entirely be represented within a few seconds.

Method description

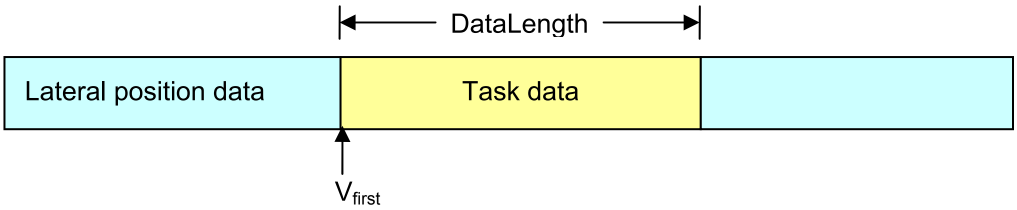
The method has to be applied to the following measures (indicated is also the lengths of the windows; two versions per measure):

- Lateral position variation: window lengths 15 & 30 seconds, short names: st_lp15, st_lp30
- Time headway variation: window lengths 15 & 30 seconds, short names: sd_hwt15, sd_hwt30
- Distance headway variation: window lengths 15 & 30 seconds, short names: sd_hwd15, sd_hwd30
- Speed variation: window lengths 15 & 30 seconds, short names: st_sp15, st_sp30

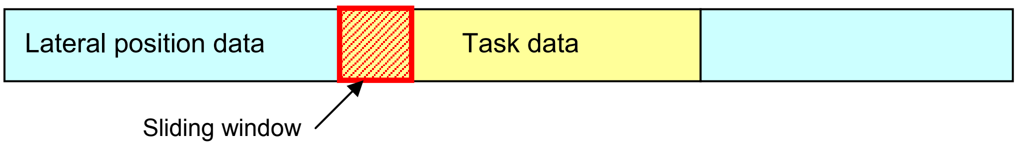
The method is based on a sliding window function which calculates the variation measures, e.g. standard deviation, of the values inside the window for each position of the window. The final variation measurement is calculated from the mean of the calculated window variation measures.

The different steps when calculating the variation measures is described below. In this description the standard deviation of lateral position data (SDLP) is calculated.

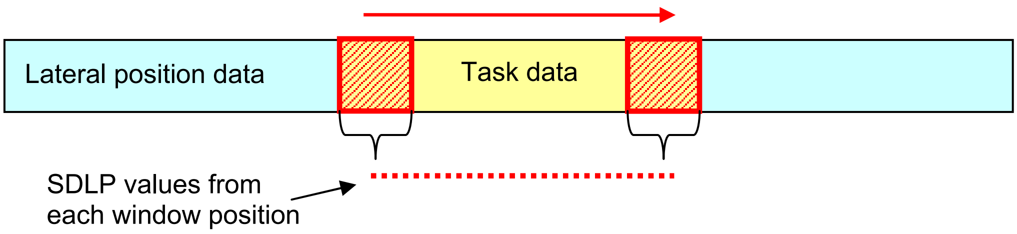
1. The first step is to identify the beginning of the task data in the Lateral position data, V_{first} and then length of the task, DataLength.



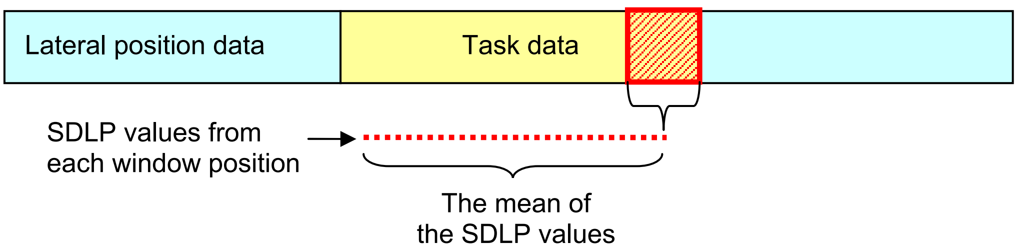
2. Then decide which window size to use, 15s or 30s. The first window calculation will start with the first value of the task data, V_{first}.



3. Calculate the standard deviation of the values inside the window and then move the window one step, to the next data value. The procedure of calculating the standard deviation of the values inside the window and then moving the window one step will be repeated until the window reaches the end of the task data.



4. The final SDLP value is calculated from the mean of the SDLP window values calculated in the step 3.



Algorithm example

An example of the sliding window algorithm when calculating the SDLP is shown below in pseudo code. The following abbreviations have been used to simplify the description:

TD = Task data, which means that TD(1) is the first value of the task data and T(i) is the i:th value of the data.

SD = Standard deviation

Pseudo code is presented below with explanatory comments marked in red:

Input values:
SampleRate = Number of values per second [Hz]
WindowSize = 15s or 30s (st_lp15 or st_lp30)
DataLength = Number of values in the task data

Initialization:
Acc_SDLP =0 %(Accumulated SDLP)
Sd_tmp=0; %(Standard deviation variable)

The position of the first value in the last window

loop from i=V_{first} to i=(V_{first}+DataLength-WindowSize)

Calculate the SD of the data, Sd_tmp, from TD(i) to TD(i+WindowSize-1)

Acc_ SDLP =Acc_ SDLP +Sd_tmp;

first window value

last window value

end

SDLP=Acc_ SDLP/(DataLength-WindowSize+1)

Calculating final SDLP value by mean calculation

Number of SD window values

Comments

When calculating the variation data with the above described procedure the first and the last values of the task data will contribute less to the variation measure. The first value of the task data will only contribute to one window variation measurement while the last value will contribute to more window variation measurements depending of the length of the task data. If the window size is half of the task data length then a value in the middle of the task data will contribute to almost all window variation measures and the first value will only contribute to one window variation measure.

Thoughts concerning how to increase the smaller contribution of the first and the last values have been many. It has been suggested to weight these values, to consider the task data as a closed circle when applying the sliding window or to include data before and after the task duration to increase the contribution. The negative aspects of these suggestions are that the final variation measure might have been distorted when increasing the contribution. When weighting the first and the last values the data will be distorted and does not reflect the actual events. When enclosing the task data as a circle there might be a big difference between the first and the last value in the task data and the variation measurement will then be dependent on the difference between the first and the last value. When including data before and after the task data other events outside the task will affect the variation measurement. These are some of the reasons why the method described above does not increase the contribution of the first and the last values in any way. The final variation measurement is then only based on the driver's actions during the task and not other events. One thought is also that the first values in the task data reflect some of the events before the task was triggered and are also affected by how fast the test person understands the task and should then not contribute as much as the values in the middle of the task data. A variation measure solution where each value contributes the same to the final value and is independent of task duration has not been solved during this project.

The dependency between standard deviation and window size are examined on lateral position data from simulation experiments on two different rural roads. Rural road data is used to avoid lane shifting and no task being performed during the data acquisition, baseline driving. The standard deviation of the lateral position data have been calculated with the use of the sliding window technique with windows of varying lengths. The standard deviation has first been calculated for all values inside the window for each window position and then has the mean standard deviation of the lateral position been calculated from all the window standard deviation measures. The results from the dependency measurement for the first rural road is displayed in **Figure 164** with corresponding lateral position data in **Figure 165** and for the second rural road in

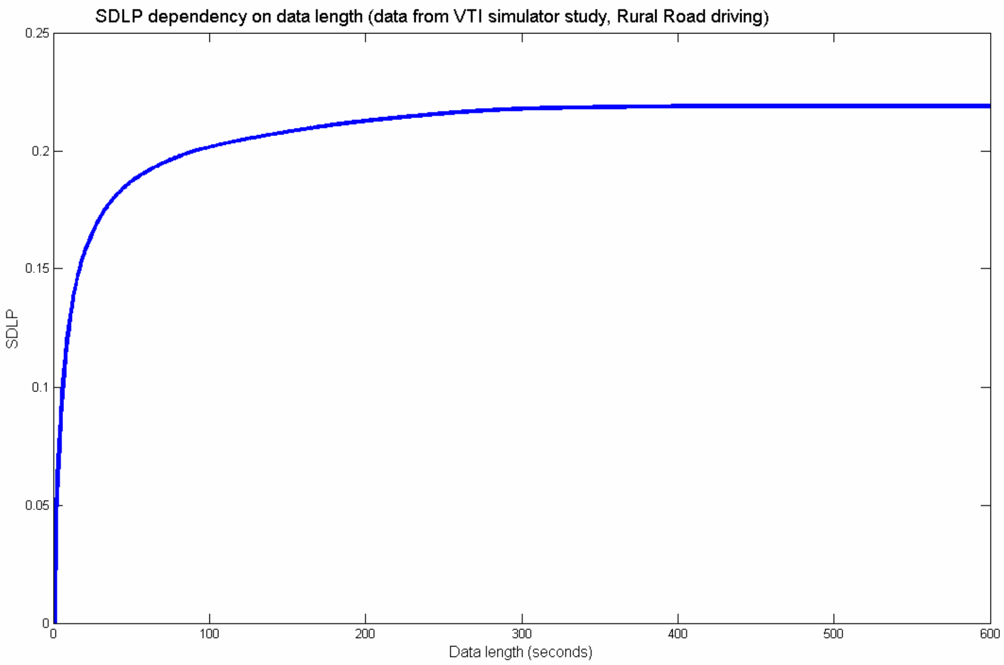


Figure 166 with corresponding lateral position data in **Figure 167**. It can be seen in the figures below that large standard deviation measures corresponds to large swerves, for example the curve of tp8 in

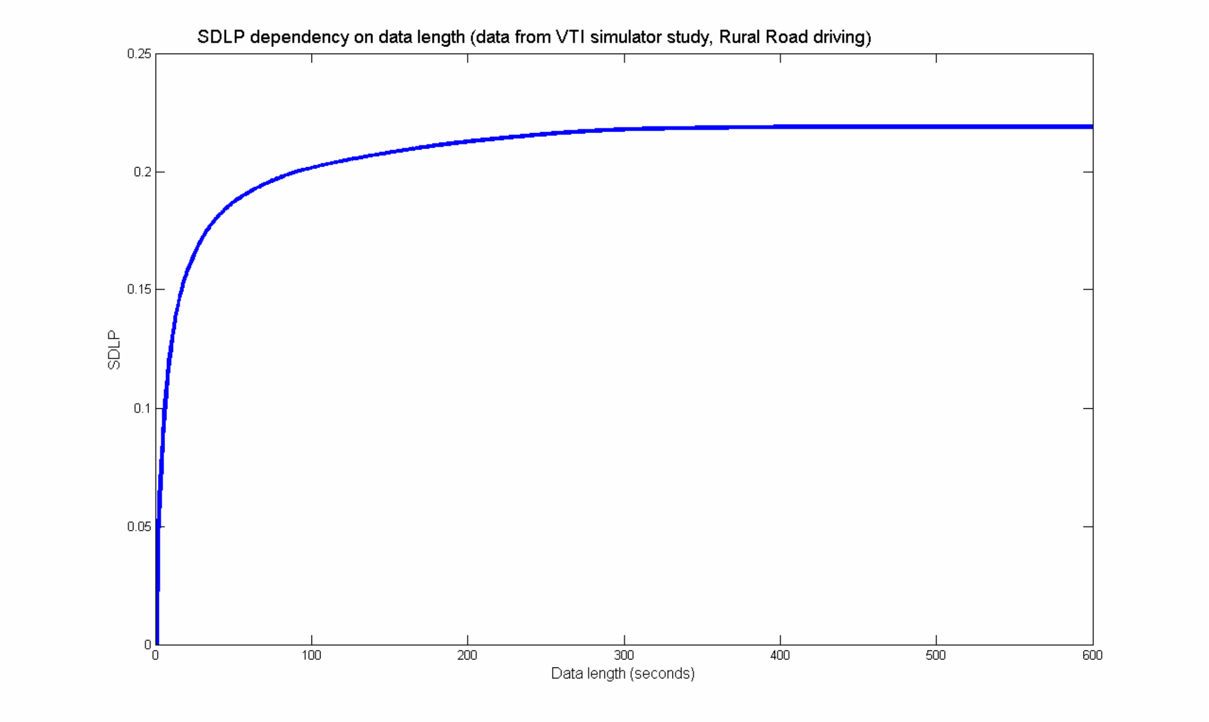


Figure 166 corresponds to the large swerves of tp8 in **Figure 167**.

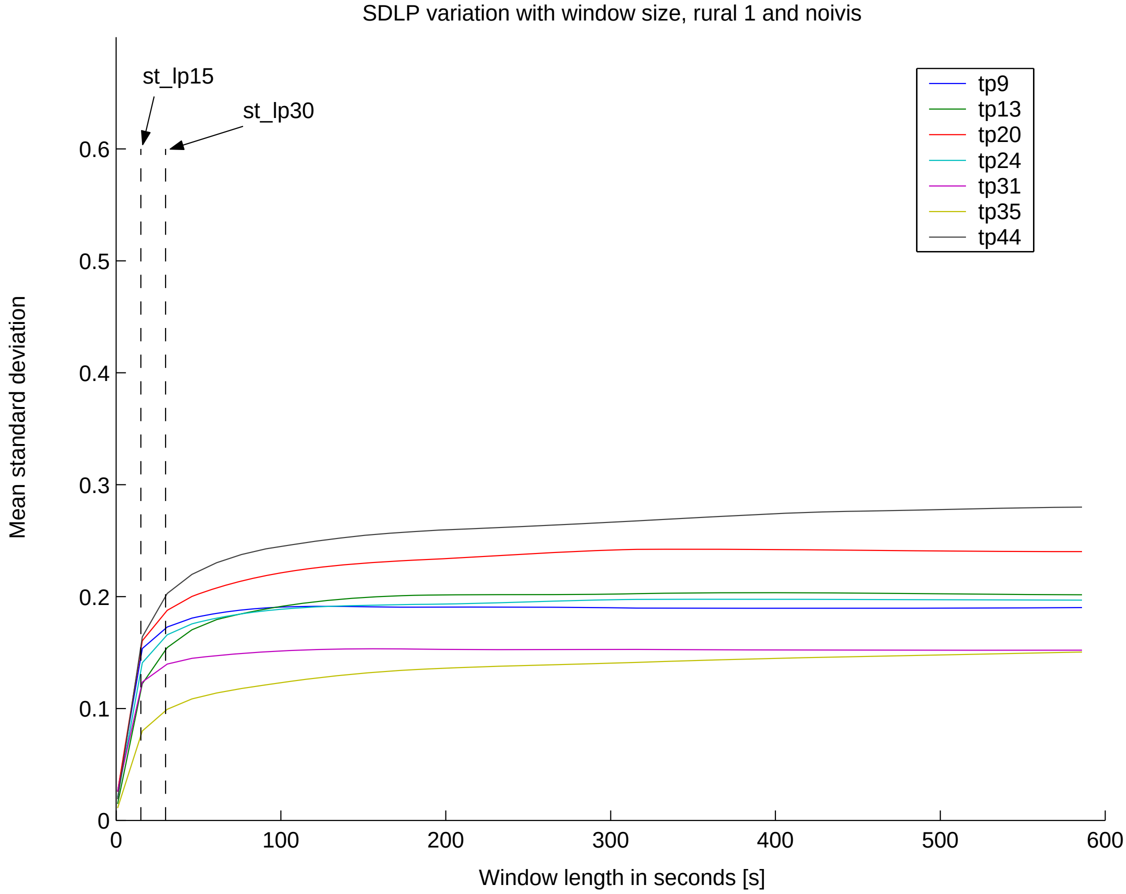


Figure 164 – The standard deviation of lateral position based on the sliding window technique as a function of window length on rural road 1, baseline data.

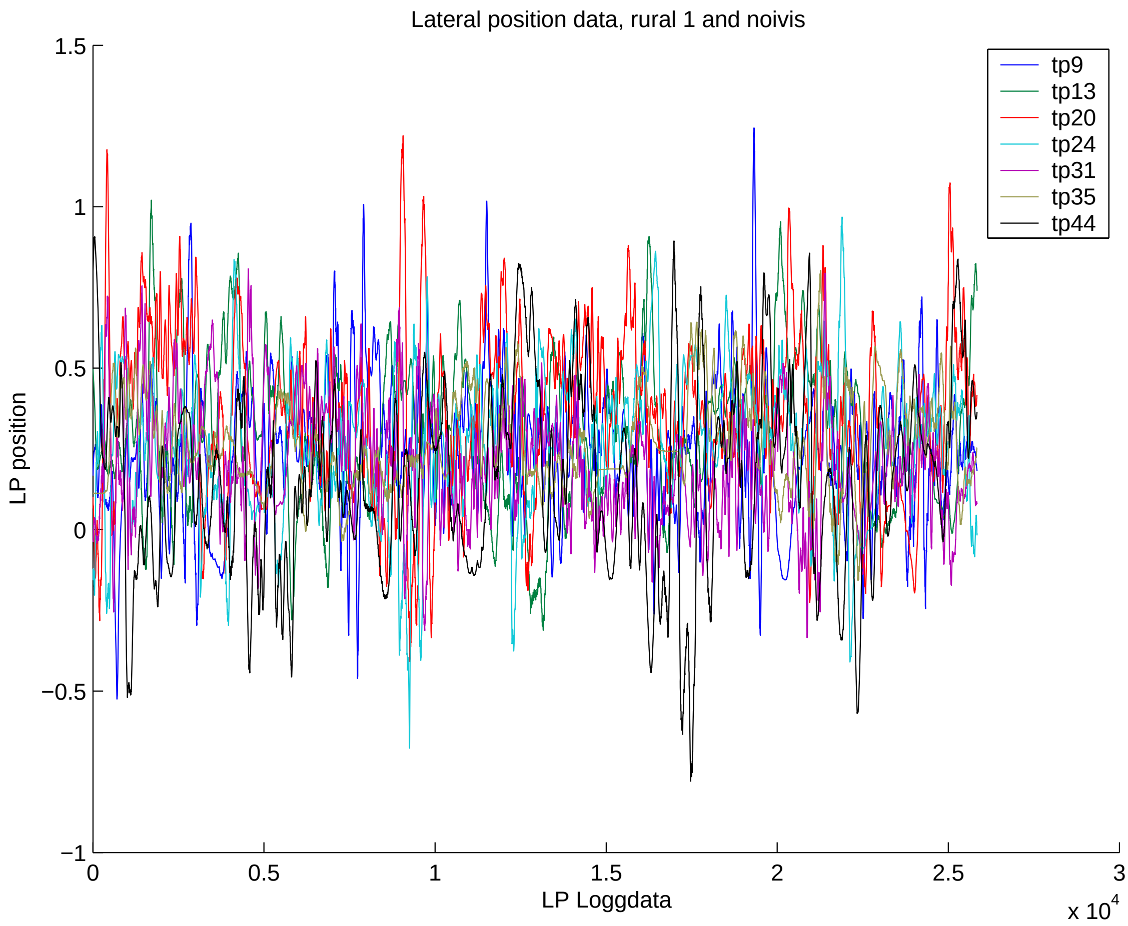


Figure 165 – Lateral position data for rural road 1 baseline data.

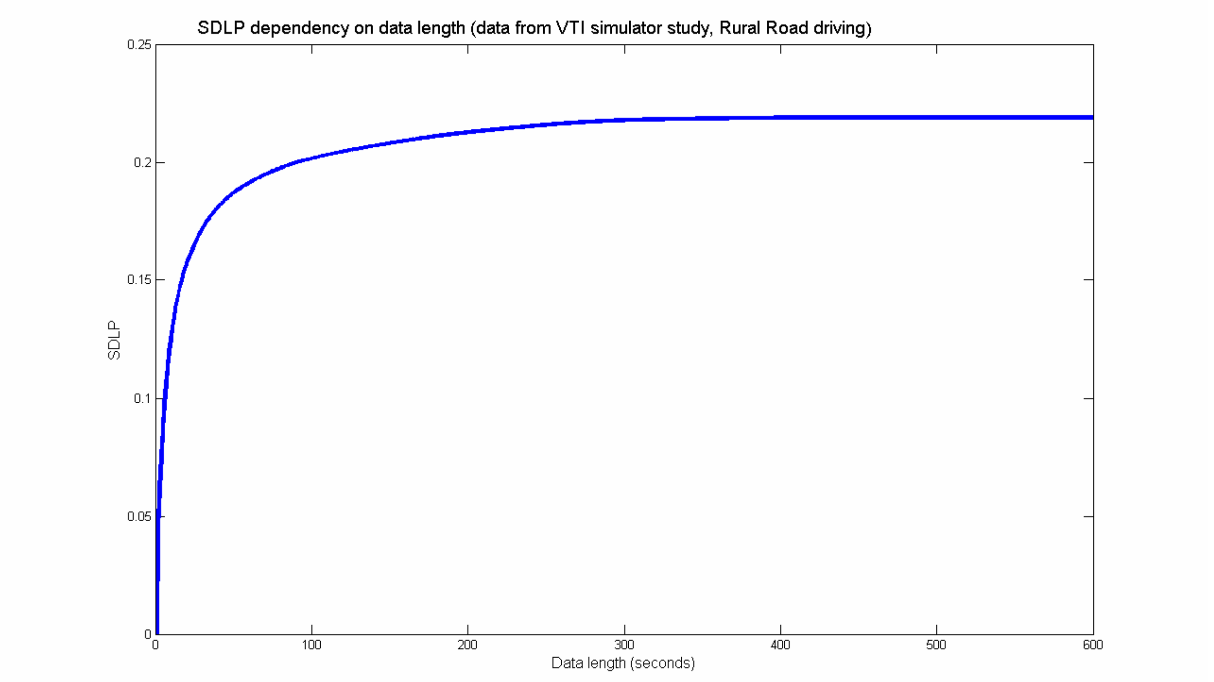


Figure 166 – The standard deviation of lateral position based on the sliding window technique as a function of window length on rural road 2, baseline data.

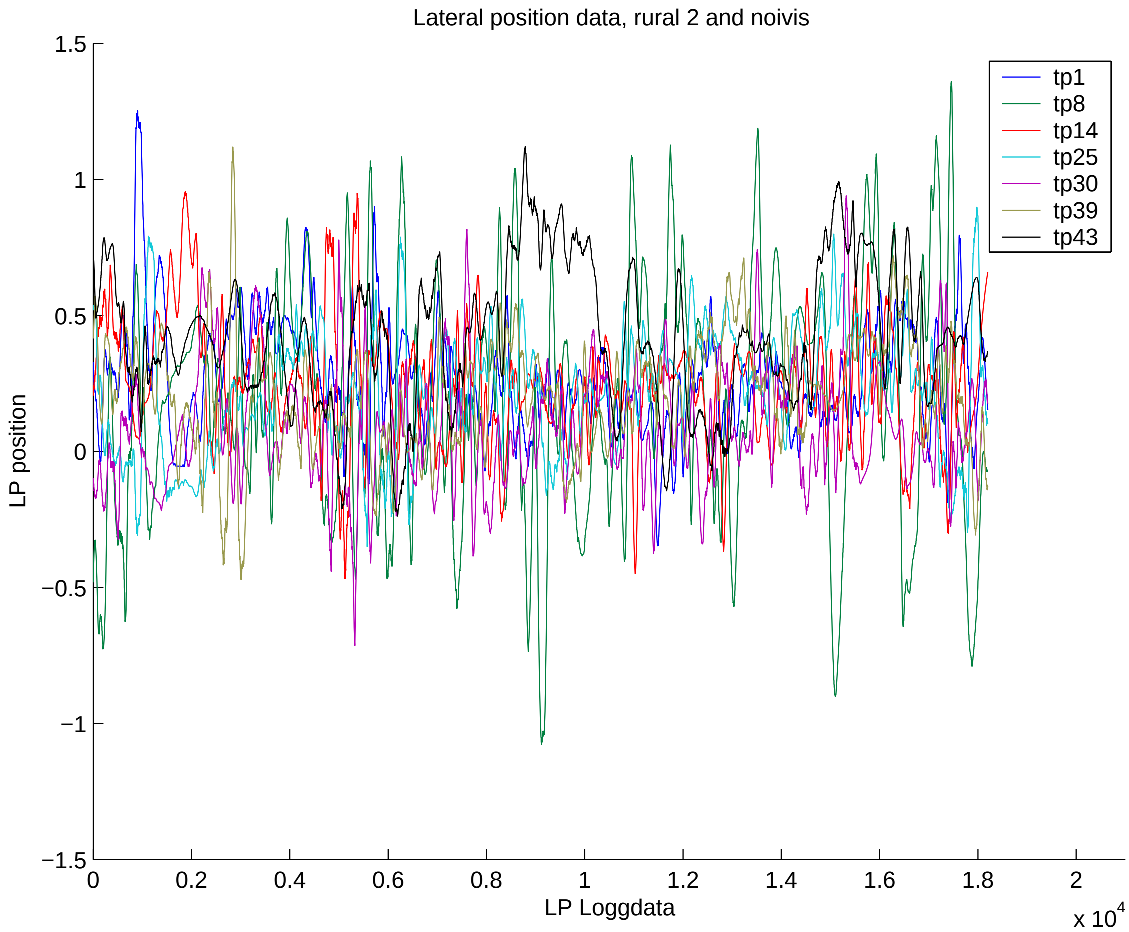


Figure 167 – Lateral position data for rural road 2 baseline data.

The chosen window sizes 15 seconds and 30 seconds are marked in **Figure 164** and in

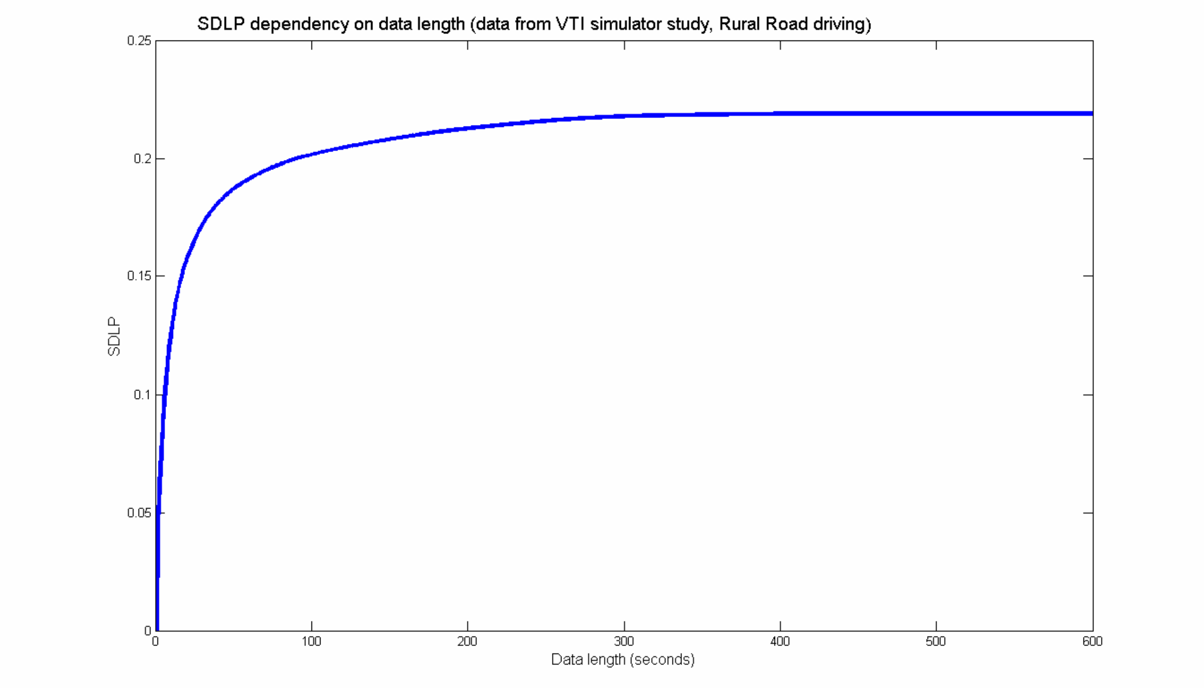


Figure 166.It can be seen in those figures that the variation measures increases up to window

sizes of approximately 45 seconds before it stabilizes. Since the only task length is these experiments longer than 45 seconds are baselines were instead 15 and 30 seconds window sizes used. The problem with these sizes is that many tasks are shorter than 15s and will then not be able to use the sliding window technique, therefore the many blanks in the result tables where sliding window measures are stated.

Appendix 4: Analysis plan

Objectives of the analysis

The analyses are to result in rankings of the IVIS tasks (within each system), i.e. significant differences between the tasks are to be identified.

Method

Separate analysis for each combination of IVIS, road type and assessment method are to be conducted.

Simulator and laboratory experiments

Table 43 – Data points for each IVIS and participant in the simulator experiments

	BL	1	2	3	4	5	6	7	8	9
Straight	1	1	1	1	1	1	1	1	1	1
Curve	1	1	1	1	1	1	1	1	1	1

The statistical model for this analysis is:

List of factors

Factor	Levels	Type
IVIS task (IVt)	0 (baseline); 1-9;	fixed
Road complexity (RLv)	1 (straight); 2 (curved);	fixed
Study object (α)	1; 2; 3; ...; 20	random

$$y_{ijk} = \mu + IVt_i + RLv_j + \alpha_k + (IVt \times RLv)_{ij} + (IVt \times \alpha)_{ik} + (RLv \times \alpha)_{jk} + \varepsilon_{ijk}$$

SPSS syntax

```
UNIANOVA
  y BY IVt RLv α
  /RANDOM = α
  /PLOT = PROFILE( IVt*RLv )
  /EMMEANS = TABLES(IVt) COMPARE ADJ(LSD) 1)
  /EMMEANS = TABLES(RLv) COMPARE ADJ(LSD)
  /DESIGN = IVt RLv α IVt*RLv IVt*α RLv*α.
```

1) Note that the option LSD means that no adjustment is made for multiple comparisons, SIDAK or Bonferroni adjustments could be chosen instead. Furthermore, the comparisons between estimated marginal means (EMMEANS) are made as if all factors are fixed.

Field experiments

In the field experiment analyses, the following factors are included:

1. IVIS task, BL + 6-9 levels, nominal
2. Participants, 15 levels, random

The statistical model for this analysis is:

List of factors

Factor	Levels	Type
IVIS task (IVt)	0 (baseline); 1-6 (or 9);	fixed
Study object (α)	1; 2; 3; ...; 20	random

$$y_{ijk} = \mu + IVt_i + \alpha_k + \varepsilon_{ijk}$$

SPSS syntax

```
UNIANOVA
  y BY IVt α
  /RANDOM = α
  /PLOT = PROFILE( IVt )
  /EMMEANS = TABLES(IVt) COMPARE ADJ(LSD) 1)
  /DESIGN = IVt α
```

1) Note that the option LSD means that no adjustment is made for multiple comparisons, SIDAK or Bonferroni adjustments could be chosen instead. Furthermore, the comparisons between estimated marginal means (EMMEANS) are made as if all factors are fixed.

Appendix 5: Wiener Fahrprobe protocol

The Wiener Fahrprobe protocol used in e.g. the VTT field experiment is presented below.

What is it?

The Wiener Fahrprobe is based on the Austrian driving test and can be used to assess driver behaviour using observers in the vehicle (Risser, R., 1985. Behaviour in traffic-conflict situations. Accident Analysis and Prevention, 17). Each of the observers records a variety of different observations. The total set of variables recorded is intended to be a reflection of the observed driving behaviour or driving style.

Method

How many observers?
This technique usually requires the presence of two observers in the car, one in the front passenger seat and the other in the rear. It is possible to do it with one, although it can be difficult to record everything and provide directions. Whilst it can become tedious doing the same task, you are probably better off sticking to being either the observer in the front or the back and not swapping between the two. This provides more reliability in my opinion (and is safer).

What do the observers record?
The Wiener Fahrprobe allows two sets of variables to be collected:

1. Standardised variables – recorded by the coding observer in the rear

The list of standardised variables consists of those types of behaviour that can be specified and be expected to appear in advance, for example speed choice.

Within these variables, errors might occur, such as "driving above the speed limit". These types of errors are predictable, although not as regards their frequency. Other, and more severe errors, are not easily predicted, either with respect to where they will appear, or under which conditions, or whether they will appear at all. The Highway Code can be used for the identification of erroneous behaviour.

In addition, communication aspects probably influence driver behaviour, especially in urban areas between car drivers and vulnerable road users, e.g. pedestrians. Both positive and negative communication are recorded.

2. Traffic conflicts - recorded by the observer in the front

Traffic conflicts are situations where the driver and other road users are on a collision course and have to react in order to avoid an accident. Such conflicts may include rear end or head on conflicts, conflicts at turns, lane change conflicts and conflicts with vulnerable road users.

Practicalities

- The "coding observer" sits in the rear of the car records the standardised variables. One standardised observation sheet per road section is completed.
- The "conflict observer" in the front passenger seat records traffic conflicts. One coding sheet per error is recorded along with the name of the road section on which it occurred.

- The technique is best used on a predefined experimental route, thereby minimising variance in road and traffic conditions between drivers (although of course this can never be entirely reduced).
- You would, of course, have to do two runs per driver, one without the system and one without.
- The most important thing to do in advance of the trials is to go out round the route a few times and familiarise yourselves with the road sections, and the traffic.
- You need to be able to provide drivers with clear directions well in advance of them having to turn etc.
- Get used to using hand signals for indicating left and right (people invariably muddle these up!).
- Take a map and be prepared to divert and abandon the trials.
- Try and avoid the rush hours.
- Try and ensure that each driver goes out at the same time.
- Keep your eyes peeled for advanced warnings of roadworks that might affect the journey.

Analysis

This is very straightforward, and simply requires a total count of the number of negative behaviours including:

- Unsafe merging/gap acceptance at junctions
- Incorrect lane changes
- Ignores other road users e.g. by not adapting their speed
- Unsafe overtaking manoeuvres
- Adoption of short headways

In addition, the total number of conflicts can be calculated.

Example form for the coding observer

Section no. Start Finish
1

	JUNCTIONS		
	Lane choice for proceeding		**Behaviour at traffic lights**
	correct		drives against red
	in time		drives against amber
	at the last moment		does not start when it is green
	incorrect		starts too early
	Use of the indicator		**Gap Acceptance**
	indicates in time		safe
	does not indicate		unsafe
	does not indicate in time		with traffic
	indicates ambiguously		without traffic
	Checks the situation with respect to other road users		inappropriate speed
	yes		aggressive
	no		

	LINK		
	Overtaking or lane change		**Speed**
	correctly		Inappropriate
	not correct		Inappropriate for road geometry
	in spite of oncoming traffic		too fast near VRUs
	without sufficient vision		brakes abruptly
	while forbidden		unsteady speed
	because of a stationary obstacle		**Distance to the road user ahead**
	lane change in time		too close
	Use of the indicator		**Checks the situation with respect to other road users**
	indicates in time		yes
	does not indicate		no
	does not indicate in time		**Behaviour when merging**
	indicates ambiguously		safe
	Lane use		unsafe
	inaccurate, weaving		with traffic
	extremely on the right side of the lane		without traffic
	extremely on the left side of the lane		inappropriate speed
	cuts the curve		aggressive

Lane use		**Overtaking**	
Left Lane		Over Took	Overtaken By
Centre Lane			
Right Lane			
Following			
% Journey			
% Too Close			

Example form for the conflict observer

	Approaching a place of interaction
	checks the situation
	drives with anticipation
	does not drive with anticipation
	inappropriate speed
	inaccurate lane choice

Interaction			
	insists on right of way		does not insist on right of way
	does not allow to continue/merge		allows to continue/merge
	does not reduce speed		reduces speed
	presses other cars		
	obstructs others (e.g. at crossings, etc.)		
	others move into the safety distance of the subject		
	turns right near oncoming traffic		
	obstructs others when turning right		
	obstructs others when turning left		
	makes other road users decelerate		
	makes others accelerate		
	impedes cyclists/pedestrians		
	endangers cyclists/pedestrians		

	Overtakes or changes lane
	cuts up
	too small lateral distance
	Aborted

	Conflict
	subject provokes conflict
	subject does not provoke conflict

	Communication		comments
	Positive		positive
	Negative		negative

Description
