INFORMATION SOCIETY TECHNOLOGIES (IST) PROGRAMME

Information Society Technologies

adaptive integrated driver-vehicle interface

AIDE
IST-1-507674-IP

Driving performance assessment – methods and metrics

Deliverable No. (use the number indicated on technical annex)		**D2.2.5**	
SubProject No.	**SP2**	SubProject Title	**Evaluation and assessment methodology**
Workpackage No.	**WP2.2**	Workpackage Title	**Driver workload and distraction assessment methods and tools**
Activity No.	**2.2.5**	Activity Title	**Using performance indicators for driver state assessment**
Authors (per company, if more than one company provide it together)		**Joakim Östlund, Björn Peters, Birgitta Thorslund VTI; Johan Engström & Gustav Markkula, VTEC; Andreas Keinath, BMW; Dorit Horst & Susann Juch, Regienov; Stefan Mattes and Uli Foehl, DC.**	
Status (F: final; D: draft; RD: revised draft):		**RD (deliverable due month 22 (dec 2005))**	
File Name:		**AIDE-VTI-D_WP2.2-R1-V2-Driving Performance Assessment Methods & Metrics.doc**	
Project start date and duration		**01 March 2004, 48 Months**	

Abbreviations and definitions

This report was written according to the AIDE Glossary definitions as described in AIDE deliverable 4.0.1 (Interaction Plan, M 13 -30). However, the table below is only a subset of the complete glossary. Some definitions were excluded and some were added. Abbreviations and terms used in this deliverable but not yet incorporated in the official glossary are listed separately at the end of the table. Furthermore, four columns of the original glossary are excluded in the table below as the table is only intended as an aid for the reader. A more elaborated glossary with e.g. alternative definitions according to the original glossary can be found in Appendix 1.

Term	Definition
Action	An event initiated by the driver or an **application**
ADAS	Systems that interact with the driver with the main purpose of supporting the **driving task** on the tracking and regulating levels.
AIDE design scenario	A driving situation, specified by at least one **action** and one or more **DVE state** parameters, acted upon by **the AIDE system**.
AIDE meta-function	The response of the AIDE system to an **AIDE design scenario**.
AIDE system	The Adaptive Integrated Driver-vehicle Interface targeted by the AIDE IP, implementing the AIDE meta-functions
Application	A program (such as a word processor or a spreadsheet) that performs one of the important tasks for which a computer is used
Application Programming Interface (API)	A software interface that enables **applications** to communicate with each other. An API is the set of programming language constructs or statements that can be coded in an application program to obtain the specific functions and services provided by an underlying operating system or service program.
Behavioural adaptation (BA)	The whole set of behaviour changes that are designed to ensure a balance in relations between the (human) organism and his surroundings, and at the same time the mechanisms and processes that underlie this phenomenon
Data	Information output by a sensing device or organ that includes both useful and irrelevant or redundant information and must be processed to be meaningful.
Device	Functional unit of hardware or software, or both, capable of accomplishing a specified purpose.
Distraction	see **driver distraction**
Domain	A problem space
Driver distraction	Attention given to a non-driving related activity
Driving demand	The demands of the driving task
Driving performance	Performance of the **driving task**
Driving task	All aspects involved in mastering a vehicle to achieve a certain goal (e.g. reach a destination) , including tracking, regulating, monitoring and targeting.
DVE state	A set of **parameters** representing certain aspects of the driver, the vehicle and the environment
Element	A component of a system; may include equipment, a computer program, or a human.
Feature	User-visible aspects or characteristics of a system.
Function	A task, action, or activity that must be accomplished to achieve a desired outcome (EAST-EAA).
Functional Specification	Specification of the normal function of the system.

Functionality	A synthesis of functions to provide a major functional entity of a unit.
Human Machine Interface (HMI)	A set of components that govern the interaction between the user and one or more vehicle systems
Integrated system	Two or more in-vehicle devices, which provide information to, or receive output from, the driver of a motor vehicle, whose output have been combined or harmonised
Interface	Abstraction of a service that only defines the operations supported by that service (publicly accessible variables, procedures, or methods), but not their implementation.
Interface-Specification	A set of language and message formats used for the communication between software components.
IVIS	Systems that interact with the driver with the main purpose to support tasks on the targeting and monitoring levels, or do not support driving at all.
Mental workload	The specification of the amount of information processing capacity that is used for task performance
Message (in architecture)	A message is a group of data values that must be exchanged together. A typical reason for grouping data is the temporal consistency of different data values: a control algorithm may require for example that the temperature and the pressure are measured at the same time. Depending on the size of the message and the maximal size of frames, several messages may be transported by one frame or it may be necessary to split a message into several segments for being able to send it over a network.
Parameter	An independent variable used to express the coordinates of a variable point and function of them.
Parameterisation	To express in terms of **parameters**.
Performance	see **driving performance**
Primary task	The task with the highest priority in a multi-tasking situation.
Real time	System which has to finish the processing within a specific time interval (deadline) dedicated by its environment.
Reference task	**Secondary task** activity which constitutes a primary task performance reference level when performed concurrently with the **primary task**
Requirement	A condition or capability needed by a user to solve a problem or achieve an objective.
Secondary task	A task with lower priority than the **primary task** in a multi-tasking situation.
Specification	Precise (formal if possible) description of an object within the scope of the task.
System	A collection of components organized to accomplish a specific **function** or set of functions.
System response delay (SRD)	Interval during which the driver has to wait for an interface to respond or update in order to continue the task
Task	Process of achieving a specific and measurable goal using a prescribed method
Use case	An intended or desired flow of events or tasks that occur within the vehicle and are directed to or coming from the driver in order to accomplish a certain system-driver interaction.
Visual demand	Degree of visual activity required to extract information from an object to perform a specific **task**

Not yet accepted in the AIDE glossary	
Term	**Definition**
ABS	Anti Blocking System
Accuracy	Closeness with which data approaches the true value of the variable being measured (this can also be termed validity).
ADAM	Advanced Driver Attention Metrics - German national project
Assessment	Process of determining the performance and/or impacts of a candidate application, usually in comparison with a reference case (existing situation or alternative applications), and usually including an experimental process based on real-life or other trials, often involving users. (from WordWeb dictionary). Evaluation and assessment are used rather substitutable in this deliverable.
AWAKE	System for effective Assessment of driver vigilance and Warning According to traffic risk Estimation (IST project 2000-28062)
CARIN system	A Navigation System by Philips
COCOM	Contextual Control Model (see e.g. Hollnagel and Woods, 2005)
DVE	Driver Vehicle Environment
ECOM	Extended Control Model (see e.g. Hollnagel and Woods, 2005)
Evaluation	1. Act of ascertaining or fixing the value or worth of something, 2. An appraisal of the value of something (from WordWeb dictionary). Evaluation and assessment are used rather substitutable in this deliverable.
FCW	Frontal Collision Warning is a system able to warn the drivers if the host-vehicle is approaching too fast an ahead obstacle. The warning information given to the user can be visual, acoustic or both.
Formative Evaluation	Evaluation performed early in the design process with the main goal to improve design. It is thus a key part of an iterative human-centred design process with repeated evaluations and re-designs. (See also chapter 2.5.2)
HASTE	Human Machine Interface And the Safety of Traffic in Europe. HASTE (2002-2005) is a European project that was supported within the specific programme Competitive and Sustainable Growth and is in specific response to Growth Task 2.2.5/6 - Development of methodologies and performance measures to assess long term safety implications of new in-vehicle technologies including HMI for road transport
HF	High Frequency
HUD	Head Up Display
ICA	Interaction and Communication Assistant.
IP	Integrated Project
ISA	Intelligent Speed Adaptation

ISO	International Standardisation Organisation
ITS	Intelligent Transportation System
JCS	Joint Cognitive System
LANEX	Lane Exceedance. A metric of vehicle lateral control.
LCT	Lane Change Test
LDW	Lane Departure Warning is a system able to warn the driver if the host-vehicle is about to leave the lane. The warning can be visual, acoustic or both.
LKA	Lane Keeping Aid is a system supporting the driver keeping the vehicle in the lane.
LP	Lateral Position
LPF	Low Pass Filtering
MSDLP	A Modified way to calculate Standard Deviation of Lateral Position (see Chapter 7)
NASA	National Aeronautics and Space Administration
NASA-TLX	NASA Task Load indeX
OECD	Organisation for Economic Co-operation and Development
PC	Personal Computer
PDA	Personal Digital Assistant, a type of handheld computer
PDT	Peripheral Detection Task: Method whose purpose is to measure the driver's mental workload and visual demands by means of a visual stimulus presented at the periphery of the ocular field; the user is asked to press a button in response to the stimulus.
Precision	A measure of the reproducibility of the measurements; i.e., given a fixed value of a variable, precision is a measure of the degree to which successive measurements differ from another.
Reliability	The reproducibility of measurements over time; it refers to the consistency of the measure on different occasions or with different sets of equivalent tasks.
Resolution	The smallest change in measured value to which the instrument (recording device) will respond.
SD	Standard Deviation
SDLP	Standard Deviation in Lateral Position
Sensitivity	The ability to measure small changes [ISO WG8 N266].
S-IVIS	Surrogate In Vehicle Information System. Secondary tasks representing IVIS in terms of cognitive and visual load.
SMS	Short Messaging System
SP	Sub Project
Summative Evaluation	The purpose of summative evaluation is to verify an almost finalized system according to certain criteria, mainly those specified in the requirements phase. (see also chapter 2.5.2)
SW	Steering Wheel (metrics)

TLC	Time to Line Crossing
TTC	Time to Collision
TTT	Total Task Times
Usability	The extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use. (ISO/IEC 9241-11: Guidance on Usability (1998)).
Validation	Process of verifying that an application performs as expected, often based on assessment results. In this sense, validation is usually considered as an extension to the assessment process, and sometimes the generic term assessment will be used to encapsulate validation.
Validity	The extent to which the variable is diagnostic for the concept being investigated. A measure is valid if it measures what it intends to measure. A reliable measure is not necessarily valid.
WP	Work Package

Table of content

List of figures

List of tables

EXECUTIVE SUMMARY

This deliverable is based on the work performed in T2.2.5. It is a continuation of the work initiated and reported in D2.2.1 (Johansson et al., 2004) on driver performance methods and metrics. The work constitutes a thorough investigation of a selection of driving performance and Lane Change Test (LCT) based metrics, the outcome is a "user's guide" to a set of vehicle control metrics.

Analyses and description of metrics were based on a theoretical framework defined in chapter 2, providing the basis for understanding the concepts related to driving and distraction. The mechanisms of distraction and driving behaviour are complex. Mainly two mechanisms explaining the effects of distraction were identified: Visual distraction leads to a more distributed gaze, which results in deteriorated lateral control. However, the driver tries to compensate by reducing the travel speed and thus reduces the information flow and risk of accident. Cognitive load leads to less capability to monitor and assess the current driving situation, and accordingly to less efforts to search for unexpected events. The result is a forward gaze concentration, which leads to more stable lateral control. However, there is a risk that the event detection capability is deteriorated and that the driver is more likely to make poor decisions in the interaction with the road environment and with other road users. This can be found as occasional speeding behaviour and very short headways to lead vehicles.

When choosing vehicle control metrics for driving related studies, it must be considered that one single isolated metric seldom can be used to identify any negative effect of IVIS/ADAS on driving performance. Instead, several metrics are required. Also, the route or scenario influences the possibility to apply different metrics. In rural road and motorway with few overtakings, turnings and conflict situations, there are many metrics to apply. In built up areas and in conflict situations, the applicability of many measures is non existent. Therefore, the temptation to choose critical situations as driving scenarios should be weighted against the number of applicable metrics.

The work presented in this deliverable was based on previously reported studies, reanalysis of data from previous experiments, and new small scale experiments. An exhaustive investigation of various vehicle control metrics was carried out. As a result a set of control metrics were chosen and specified in terms of definition, description of use, interpretation guidelines and limits. Primarily, two metrics are proposed: *modified lateral position variation* and *steering reversal rate*. New metrics for the Lane Change Test were evaluated, as well some amendments to the LCT-method. The LCT was also investigated in terms of different setups i.e. comparing different simulator with the traditional PC version. It was found that the up-to-date version of the LCT method, including metrics, is the most feasible one, the modified version could not be used to discriminate between visual or cognitive workload. Furthermore, the LCT provides comparable results independent of simulator fidelity (from PC-mock up to real vehicle cabin with projector visualisation). In the discussion and conclusion some examples are given on how to apply the selected vehicle control metrics.

Finally, in task 2.2.7, the provided performance metrics will be combined and compared with secondary task methods and workload metrics. In task 2.1.4, the metrics will be included in the compilation of a final battery of test procedures for the evaluation of adaptive IVIS/ADAS in field and simulators, specifically in WP2.4.

1 Introduction

The purpose of D2.2.5 is to deliver a set of driving control and Lane Change Test metrics to be applicable in the AIDE test regime in order to investigate workload and distraction. Thus, it should be useful in both formative and summative evaluations efforts to be carried out. Thus, it constitutes a continuation of the work initiated and reported in D2.2.1 (Johansson et al. , 2004). Recommendations are given in the form of a user's guide.

The development and/or choice of metrics are based on:

- new and existing analyses on previously collected data, e.g. from the HASTE project and previous LCT-experiments in e.g. the ADAM project.
- theoretical discussions.
- new small scale experiments.

It is of high importance that the suggested driving control and LCT metrics:

1. are well defined in order to promote consistency and avoid different implementations and interpretations in different experiments,
2. provide results that can be interpreted deterministically in terms of driving performance and be used a probe of distraction and workload,
3. are applicable in typical driving scenarios, such as motorway, rural road and urban driving including "normal" driving and critical events,
4. cover the evaluation of the effects of visual and cognitive workload and distraction on driving performance in formative and summative evaluation efforts.

Since there are no detailed driving scenarios determined as a definite prerequisite to this deliverable a final choice of metrics cannot be made for all scenarios; selection of metrics are strongly dependent on driving scenario and experimental design. A reduction of the number of metrics will be made in an empirical comparison of approaches for evaluating driving performance and driver state (Task 2.2.7, deliverable due month 26) and in the compilation of a final battery of test procedures for the evaluation of adaptive IVIS/ADAS in field and simulators (Task 2.1.4, due month 36). However, in case of redundant measures, only the best one is given.

1.1 Input from previous AIDE work

The key input from previous AIDE work to the present task was the review of existing performance metrics included in D2.1.1 (Johansson et al., 2004). Moreover, the theoretical framework outlined in Chapter 2 is based on the general conceptual framework described in the AIDE Interaction Plan (D4.0.1; Engström, 2005).

1.2 Structure

The structure of the deliverable is as follows: Chapter 2 provides a theoretical framework explaining the concepts of driving behaviour, driving performance, distraction, risk, behavioural adaptation etc. Also, mechanisms of the effects of distraction on driving behaviour and driving performance are explained. These concepts and models are required for the theoretical reasoning concerning the effects of distraction on the driving control and LCT metrics.

Chapter 3 includes studies on *vehicle control metrics*, and a selection of metrics. The outcomes of previous projects have been used, previously collected data was reanalysed, and some new data was collected and analysed.

The following chapters (4, 5 and 6) include studies on the Lane Change Test with the focus on new metrics (previously collected data was reanalysed), a modification of the method and finally study on the effect of set up (simulator/mock up/field) on driving behaviour in the LCT.

In the next two chapters (7 and 8), the metrics selected for inclusion in IVIS and ADAS evaluations (the AIDE test regime) are described in a “user’s guide” format. All metrics are given with **definitions**, **descriptions of use** (where and how to use the metric), **interpretation guidelines** and **limitations** (e.g. “not applicable in urban driving”).

Finally, the last chapter provides a discussion about the applicability of the selected metrics in different types of evaluations (formative/summative) and different scenarios (motorway/rural road/urban road etc.). This part is of high importance for the further use of the metrics in the combination with scenarios and test environments, since there are complex settings in which very few metrics are applicable and since there are problems related to measuring driving behaviour in real traffic.

1.3 Final introductory word

When designing an experiment, you should always ask yourself: **What is the main purpose of the experiment?, What do we need to know and thus measure?, and What can we really measure and what will it say?** Most certainly, the answers will generate changes to your design. The deliverable was written in collaboration between VTI, VTEC, DC, Regienov, and BMW.

2 Theoretical Framework

2.1 Introduction

The goal of T2.2.5 is to develop and specify suitable metrics for IVIS/ADAS evaluation with respect to the systems' effect on *driving performance*. Thus, it is of key importance to be clear about what is meant by "driving" and how performance of the driving task can be conceptualised. The framework outlined here is mainly intended as a "common language" to describe key phenomena and concepts relevant to driving performance measurement. It is based on the COCOM/ECOM model (Hollnagel and Woods, 2005) identified in SP1 (D1.1.1a) as the prime candidate for the Generic (G-) DVE model in AIDE. The COCOM/ECOM has also been adopted as the general conceptual framework for AIDE. For other descriptions of potential applications of COCOM/ECOM, see Engström and Hollnagel (2005) and Peters and Nilsson (2005).

This chapter is structured as follows: First, the concepts of driving and driver behaviour are defined and the general hierarchical control framework outlined. Based on this, a taxonomy for IVIS/ADAS functions is proposed, where functions are classified with respect to the aspect of driving that they support (if any). This taxonomy is then used to identify and categorise potential effects that these functions may have on driving performance. Finally, different types of driving performance metrics are identified based on the framework.

2.2 Driving as hierarchical control

2.2.1 Control

Driving could be considered as an instance of the more general notion of *driver behaviour*. Behaviour, as opposed to mere bodily movements, can be conceived of as goal-directed activity. Thus, *driving* can be thought of as a set of behaviours directed towards goals associated with vehicle operation. *Control* is a useful concept for describing the dynamics of goal-directed behaviour, in human as well as in machines. Controlling a process means control actions are determined by the aim to achieve a consistent goal state (often called the reference or target value), e.g. by means of countering effects of external disturbances. Control is thus closely related to orderliness or predictability, i.e. a controlled system is ordered, stable and predictable while a system that is out-of-control is disordered, unpredictable and unstable.

Control comes in different varieties: In *feedback control*, the controller performs corrective actions based on the deviation between a desired outcome (the goal) and the actual state. A prototypical example of a feedback control is the thermostat. Another type of control is *feed forward*-, or anticipatory, control. In this case, control actions are based on predictions of future states, and not purely reactive as in feedback control. Driving behaviour involves both feedback, feed forward and mixed control loops. This mixture of control was nicely described by McRuer et al. (1977) for the driver's steering control.

Control theory has been widely applied to the modelling of operator vehicle handling, in automotive and other domains (e.g. Weir and McRuer, 1968; Weir and Chao, 2005; Donges, 1978). In these types of models, vehicle control is mainly modelled in terms of feedback control mechanisms where control inputs (steering wheel/pedal movement) are translated, via the vehicle dynamics, into vehicle position/heading values that are fed back to the human controller. An example of such a model is illustrated in *Figure 1*.

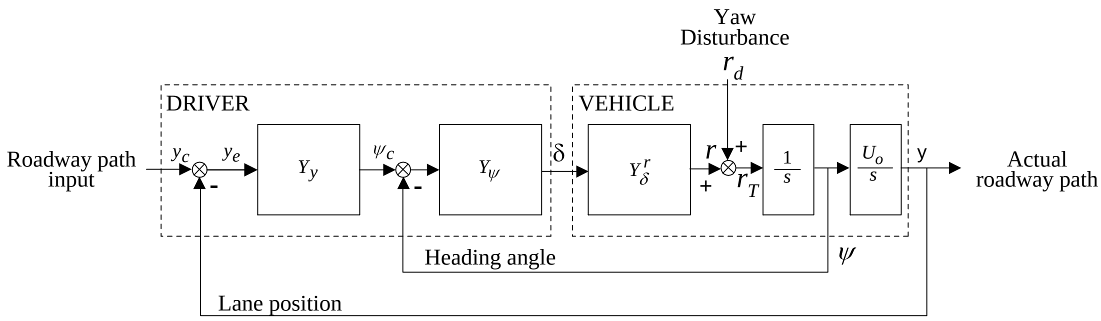


Figure 1 Example of control-theoretic steering model (from Weir and Chao, 2005)

A common criticism of these classical control models is that they are based on the assumption of the controller as an *optimiser*, i.e. with the goal of always minimising the difference between the target value and the actual state. However, humans rather tend to act as *satisficers*, i.e. they do not put more effort into a task than needed and we tolerate some deviation from a perfect control i.e. no deviation from the target value. In the case of driving, there is considerable evidence that drivers adopt time-based *safety margins* as the criteria for vehicle longitudinal and lateral control (Gibson and Crooks; 1938, Lee, 1976; Godthelp, Milgram and Blaauw, 1984; van Winsum, 1996; van der Hulst, 1999; Boer, 2000; Nilsson, 2001, van der Horst, 2005).

Control theory has also been applied to the modelling of higher level aspects of driving (e.g. Wilde, 1982; Vogel, 2002; Nilsson, 2001; Hollnagel, Nåbo and Lau, 2003). See also Jagacinski and Flach (2003) for a general text book on control theory applied to human performance modelling.

A general account for human control of a process or plant, based on Neisser's (1976) cognitive cycle concept, is offered by the Contextual Control Model (COCOM) as described in e.g. Hollnagel and Woods (2005). An important starting point for COCOM is that the controller and controlled system is viewed as an integrated Joint Cognitive System (JCS). A key concept in the model is the *construct*, which refers to what the controller knows or assumes about the situation in which the action takes place. This understanding is the basis for controller's selection of actions and interpretation of system state. The selected actions affect the process/application under control. Actions and external disturbances modify the system state which is perceived by the controller as a construct of the system state and determine future action selection. An important property of this model is thus that it captures both the feedback and feed forward aspects of control, i.e. action selection is a function of both direct feedback and the predictions of future events.

Another key property of COCOM is that it offers an explicit account of time. As illustrated in *Figure 2*, the model contains three basic time parameters: (1) Time to perceive, (2) time to decide and (3) time to act. As in any control system, the relations between these different time parameters strongly determine the performance of the JCS. For instance, long delays between action and feedback, such as in the control of a large ship, make feedback control difficult and thus put greater requirements on feed forward control mechanisms. Moreover, the time needed to perform one cycle must be less than the total time available. If not, the control will deteriorate and become sluggish. However, the available time may be increased by slowing down the pace of the task, e.g., in the case of driving, by reducing speed. The paradox is slow down to gain time.

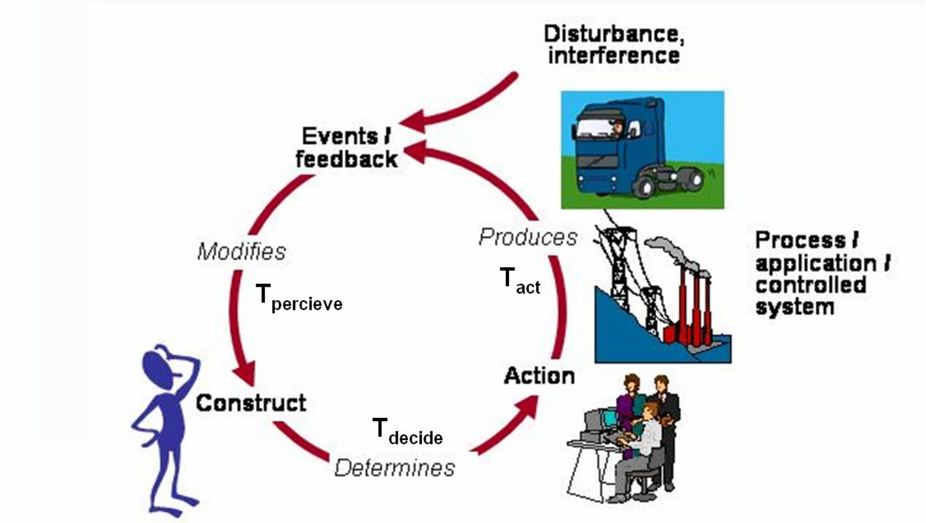


Figure 2 The Contextual Control Model (Hollnagel and Woods, 2005)

2.2.2 Hierarchical control

While COCOM only describes a single control process (i.e. pursuit of a single goal), the driving task can be viewed as the pursuit of several sub-goals with different time frames. A long-term goal could be to reach a destination in time. A medium-term scale goal would be to overtake a lead vehicle, while short-term goals include staying in lane and avoiding obstacles. A goal may also subsume goals on shorter time frames. For example, in order to reach a destination in time, it may be necessary to overtake a number of vehicles. This, in turn, requires safe vehicle handling in order to avoid collisions. Thus, the driving task can be described as a set of simultaneous, interrelated and layered control processes. In addition, drivers generally pursue many goals that are totally unrelated to driving e.g. talking to a passenger, using cell phone.

This hierarchical organisation is reflected in many models of driving, e.g. Allen, Lunefeld & Alexander (1971) and Michon's influential description of the driving task in terms of strategic, tactical and operational levels (Michon, 1985). While, Michon's model is useful as general descriptive frameworks, it does not account for driving control as such. More specifically, it does not account for the dynamical aspects of driving, nor the relations *between* control processes on different layers.

A general framework for describing behavioural dynamics in terms of hierarchical control has been offered by Perceptual Control Theory (e.g. Powers, 1998; Marken, 1986). Hierarchical control models of behaviour have also recently been very influential in the field of mobile robotics (e.g. Brooks, 1986). In the automotive field, a hierarchical control model of driving has been proposed by Hollnagel in terms of the ECOM (Extended Control Model) (e.g. Hollnagel, Nåbo and Lau, 2003; Hollnagel and Woods, 2005; Engström and Hollnagel, 2005). The basic structure of ECOM is illustrated in *Figure 3*.

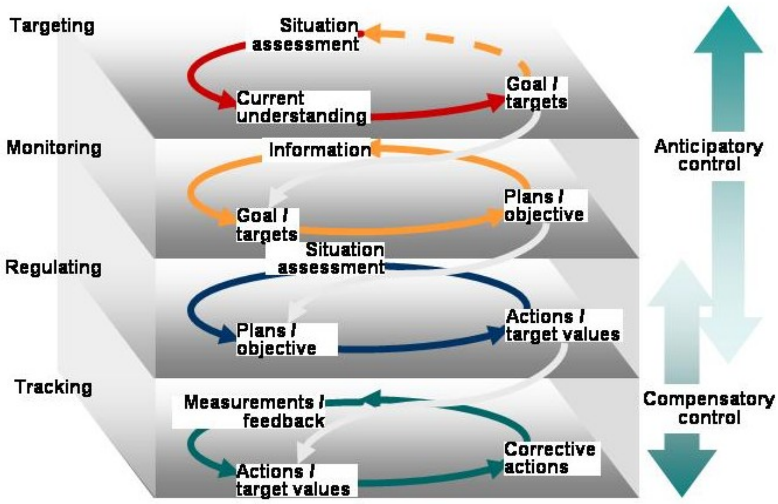


Figure 3 The Extended Control Model (ECOM)

A basic assumption behind the ECOM model is that goals corresponding to different layers can be pursued simultaneously, and that these goals and their associated control processes interact in a non-trivial way. In the current version, four control layers are postulated: tracking, regulating, monitoring and targeting. A key property of the model is that, during normal (controlled) task performance, the goals/targets for the control process on a given layer are determined by the control process one layer up. In the driving domain, the *tracking* control refers to the momentary, automated, corrections to disturbances, e.g. wind gusts. *Regulating* refers to more conscious processes of keeping desired safety margins to other traffic elements. This determines the target values for the tracking loop. *Monitoring* refers to the control of the state of the joint vehicle-driver system relative to the driving environment. It involves monitoring the location and condition of the vehicle, as well as different properties of the traffic environment, e.g. speed limits. This generates the situation assessment that determines the objectives for the regulating layer. Finally, the *targeting* control level sets the general goals of the driving task, which determines the objectives for the monitoring layer. As illustrated in Figure 3, tracking control is typically feedback while monitoring and targeting are generally feed-forward. The regulating layer may involve a mix of feedback and feed forward control. The ECOM model provides a neat account for how goals at different layers interact and how higher goals propagate all the way down to moment-to-moment vehicle handling. Control tasks on different layers may also interfere. For example, looking for directions (monitoring) may disrupt visual feedback which may affect regulating and tracking control. Driving means that control has to be carried out on different layers simultaneously and timing (synchronisation of goals and feedback) becomes critical.

Importantly for present purposes, driving performance can be defined with respect to any of these layers. For example, the ability to find a destination would be related to performance on the monitoring layer, the ability to keep safe distance related to the regulating layer and the ability to stay in lane related to the tracking layer and so on.

In addition to driving-related activities, drivers may also engage in other types of behaviour directed towards non-driving-related goals, e.g. tuning the radio or talking to passengers.

2.3 Functional mapping of ADAS/IVIS to ECOM

Since the objective of T2.2.5 is to define methods and metrics aimed to assess the effects of different types of in-vehicle functions on driving performance, it is important to find a suitable way to describe these functions. The hierarchical control framework outlined above enables a human-centred characterisation of different systems/functions, based on what aspect of driving (i.e. which control process) they support (if any).

A general distinction is often made between ADAS (Advanced Driver Assistance Systems) and IVIS (In-vehicle Information Systems). In general, there seems to be a consensus that ADAS refer to systems mainly intended to support the (primary) driving task, while IVIS are systems mainly supporting other (secondary) tasks. Many projects define their scopes in terms of these categories (e.g. in HASTE, the developed evaluation methodology is only applicable to IVIS). However, as discussed in the AIDE Deliverable D2.2.1 (Johansson et al., 2004), this distinction is somewhat problematic, since many functions that support the driving task on higher levels (e.g. route guidance, traffic information and speed alert) may interfere with driving at lower layers in the same way as "typical" IVIS such as the radio, mobile phones etc.. Thus, these systems/functions are often considered as being in a "grayzone" between ADAS and IVIS.

This issue can be resolved based on the present framework, where functions can be described in terms of the driving sub-task (i.e. which goal) that they support (where some systems may not support driving goals at all). In this context, it is more useful to talk about "functions" than "systems", since a system may contain many functions of very different types. A tentative mapping of some example IVIS and ADAS functions onto the ECOM layers is illustrated in *Figure 4*. This taxonomy is intended as a complement, rather than as an alternative, to the more traditional taxonomy of IVIS and ADAS defined in the AIDE deliverable D2.1.2. (Floudas et al., 2004)

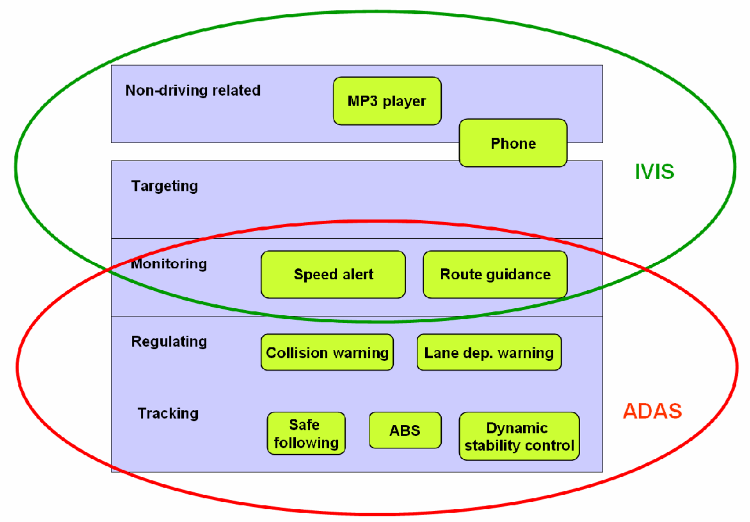


Figure 4 Proposed mapping of the ADAS/IVIS categories and some example functions onto the ECOM layers.

2.4 Characterising effects of ADAS/IVIS on driving performance

The main objective of the present work is to develop methods to quantify the effects that ADAS/IVIS may have on driving performance. This section identifies some key concepts and distinctions, and reviews the most prevalent effects of IVIS and ADAS found in the literature.

2.4.1 Direct vs. indirect effects

In this context, it is useful to make a general distinction between *direct* and *indirect* (side) effects on driving performance. *Direct* effects are those that are intended by the system designers, and implied by the system's functional specification. For ADAS, direct effects are normally the intended performance enhancements (e.g. better lane keeping performance for LDW and increased route-finding ability for navigation support). By contrast, *indirect* effects are not intended by the designers (and thus not implied by the functional specification). It should be noted that indirect effects may have both positive and negative effects on safety. For example, a potential positive effect of a lane departure warning system is the increased use of the turn signal.

2.4.2 Adaptation to safety margins

A central issue when discussing driving performance is adaptive/compensatory behaviour. The ability to adapt to, and compensate for, changing circumstances is critical for driving as well as most other day-to-day activities that we encounter. As mentioned above, there seems to be a general consensus that driving involves adaptation to some type of subjectively chosen safety margins, although these have been conceptualised quite differently by different authors. There is abundant evidence that *time-to-object* information is used by humans and animals for guiding locomotion (Gibson, 1979). Probably the first account of safety margins in driving was offered by Gibson and Crooks (1938), who proposed that drivers aim to stay within a "field of safe travel", which can be conceptualised as "tongues" stretching out in front of the vehicle, the size and form of which are determined by the time-to-contact to surrounding obstacles. More recently, this concept has been developed into more concrete time-based safety margins parameters. For example, Lee (1976) used the perceptual variable *tau*, representing time-to-contact in terms of optically specified parameters, to model drivers' braking behaviour. In traffic research, time-to-collision is often used as a driving performance metric (e.g. van der Horst and Godthelp, 1989). The corresponding metric for lateral control is time-to-line-crossing (TLC). For example, Godthelp, Milgram and Blaauw (1984) demonstrated that the TLC correlates strongly to driver's self-chosen occlusion time. Too small TLC values are thus strong indicators of violations to the driver's subjectively chosen safety margins. However, it might be difficult to determine the safe travel boundaries actually used by the driver.

In addition to the more quantitative approaches just described, there is a large body of literature that aims to characterise adaptive driver behaviour in terms of dynamic regulation of *risk*. For example, the risk-homeostasis theory (Wilde, 1982) hypothesises that drivers (at aggregated level) strive to maintain a constant level of risk. By contrast, the zero-risk theory (Summala, 1985; 1988), states that the driver aims to keep the subjectively perceived risk at zero. A related description has been offered by Fuller, who originally proposed that drivers' behaviour is guided by threat avoidance (Fuller, 1984) However, more recently he proposes that the driver attempts to maintain a certain level of task difficulty, rather than a level of risk (Fuller, 2005). Finally, based on Damasio's (1994) concept of somatic markers, Vaa (2005) proposes that driver behaviour is largely driven by emotional responses to risky situations. In terms of the COCOM/ECOM framework, the safety margins are determined in the regulating layer, based on the general situation assessment performed on the monitoring layer.

To summarise, the research cited above indicates a strong convergence to the general idea that driver performance should be understood in terms of adaptation to some type of safety margins, although these are conceptualised differently by different authors, e.g. in terms of objective quantitative parameters such as TTC and TLC, or more general concepts such as perceived risk or task difficulty. This concept has also been specifically advocated by e.g. Summala (1988) and Rumar (1988).

The effects of a specific ADAS/IVIS on driving performance is thus the result of a complex dynamic interaction between driver characteristics (subjectively chosen safety margins, driving skills, effort, etc.), vehicle parameters (vehicle dynamics) and environment (road type, curvature, lane width, traffic density etc.)[1]. In the following sections, some of the main types of effects found in the literature are discussed from the perspective of the adopted framework.

2.4.3 Behavioural adaptation

Behavioural adaptation generally refers to "the whole set of behaviour changes that are designed to ensure a balance in relations between the (human) organism and his surroundings, and at the same time the mechanisms and processes that underlie this phenomenon" (Grand Dictionnaire de la Psychologie; Bloch et al. 1999). The same general theory was applied in control theory or cybernetics (Wiener, 1948), General Systems Theory (Bertalanffy, 1968) and self-regulating systems (Carver and Schreier, 1982). However, in traffic research, *behavioural adaptation* (BA) is often used to refer to a more specific type of adaptation, as defined by OECD (1990): "those behaviours which may occur following the introduction of changes to the road-vehicle-user system and which were not intended by the initiators of the change." (p. 23). Thus, according to this definition BA only refers to indirect effects, e.g. compensatory behaviours that may reduce, or even cancel out, the expected benefits of a safety measure. Behavioural adaptation has been demonstrated to occur for many different types of ADAS (see Smiley, 2000 and Saad et al, 2004 (AIDE D1.2.1) for examples). According to the proposed framework, such changes can be more precisely described in terms of the ECOM control layers affected, and the relation between them. For example, in one of the most cited studies on BA, it was found that drivers of ABS-equipped-vehicles tend to drive with higher speed and adopt shorter headways than drivers without ABS (Fosser, Saetermo and Sagberg 1997). According to the present framework, this can be described as an adaptation on the regulating layer, where drivers exploit the improved braking performance (i.e. improved longitudinal tracking control) provided by the ABS, to reduce the margins to other vehicles. It has been hypothesised that such compensatory effects are more likely to occur if the change (in the traffic system or the vehicle) is mediated through direct sensory feedback (Evans, 1983). In terms of the present framework, this could be conceptualised as to what extent the change affects tracking control.

2.4.4 Driver distraction and workload

Driver distraction has been defined as "attention given to a non-driving related activity, typically to the detriment of driving performance" (ISO TC22/SC13/WG8 CD 16673). This

[1] The development of a DVE model and simulation of this interaction is addressed in AIDE SP1 with strong links to the present work.

definition is somewhat unclear due to the fact that “driving-related activity” and “driving performance” are not further defined.

As suggested above, the driving task should not be seen as a single activity, but rather as a set of multiple simultaneous, and layered, control tasks. Thus, driving performance could be defined with respect to any of these control tasks. For example, driving performance on the tracking level is associated with the ability to keep the vehicle within acceptable safety margins. Similarly, performance on the regulating level would be related to the ability to select these safety margins based on the general situation assessment at the monitoring level.

Distraction with respect to a given control process (e.g. tracking) could thus be viewed in terms of interference by another (driving- or non-driving related) control process, typically resulting in degraded performance on the given control task. In practice, “driver distraction” is normally used to refer to interference with the tracking and/or regulating control tasks (and this is probably the intended meaning of the ISO definition). However, the present framework enables a more precise conceptualization and makes it possible to describe distraction in more detail which are the tasks/control processes that are affected in a given distraction scenario. The main focus of the work in T2.2.5 is expected to be on distraction related to the tracking and regulating layers, although distraction on the monitoring layer is also relevant (distraction to monitoring control could have the effect that the driver misses important information, e.g. traffic signs).

Another commonly used concept is *mental workload*. While distraction is defined on the basis of attention allocation, driver workload refers to the amount of resources that the driver needs to perform one or more tasks, relative to a limited subjective resource pool. *Mental workload* has been defined as “the specification of the amount of information processing capacity that is used for task performance” (de Ward, 1996). If the total workload (resulting from performing the primary and the secondary tasks) exceeds the total available resources, the driver either needs to increase effort or reduce the task demand (e.g. by slowing down), in order to maintain task performance. Dual task studies have shown that some tasks combinations are easier to perform simultaneously than others, a phenomenon that has been modelled in terms of multiple resource pools (e.g. Wickens, 1992).

Distraction (to the driving task) is naturally related to the workload imposed by a specific secondary task. However, high workload is not a necessary precondition for distraction, since inappropriate attention allocation may be caused by low-workload tasks as well (e.g. daydreaming, looking at road signs). Based on the COCOM model, a more dynamic view of workload can be outlined; where the spare resources for a control process can be viewed in terms of the difference between total time available and the total time needed to perform the control loop (see Figure 2).

It has been shown that different types of secondary tasks induce qualitatively different effects on driving performance. For example, results from the HASTE project clearly demonstrate the radically different effects of visual and cognitive workload/distraction on driving performance (Engström, Johansson and Östlund, 2005; Victor, Harbluk and Engström, 2005). Visual distraction typically induces a visual time sharing between the road ahead and the system display. During glances to the system, the visual input needed for lateral control is reduced (or entirely inhibited) which temporarily inhibits the drivers’ steering response, leading to a steering hold (i.e. fixed steering angle). This, in turn, causes lane drifts which are compensated for by large and disruptive steering manoeuvres when the gaze returns to the road. These effects, which are mainly related to the tracking control layer, can be quantified by lane keeping variation metrics (e.g. standard deviation of lateral position) and steering performance metrics (e.g. reversal rate and steering entropy). Moreover, visual distraction generally induces compensatory behaviour, e.g. in terms of speed reduction and increased headway. Given the present framework, this can be viewed as an adaptation on the regulating level, where the safety margins are increased in order to increase the total time available to perform the tracking task.

By contrast cognitive workload/distraction does not appear to interfere with tracking control. Rather, in the HASTE studies, performance of purely cognitive tasks led to significantly *reduced* lane keeping variation compared to baseline driving (Engström et al., 2005), an effect that has also been documented in other studies (e.g. Brookhuis, de Vries and de Ward, 1991). This better lane keeping performance appears to be coupled to a concentration of gazes towards the road centre (Victor et al, 2005; Harbluk and Noy, 2002; Recarte and Nunes, 2003). However, cognitive load has also been found to strongly increase response times to sudden obstacles, as well as generally reduced situation monitoring ability (e.g. Greenberg et al., 2003). This can be viewed as a degradation of regulation control, since responding to obstacles involves changing the target values for tracking control. The results on the effects of cognitive load on speed found in the literature are rather inconsistent. For example, in the HASTE studies, significant speed increase, significant decrease as well as null effects were found (Östlund et al., 2004). In a Swedish mobile phone study, speed was reduced in response to phone conversation with a hand-held phone, but not for the hands-free solution (Patten, Kircher, Östlund, & Nilsson, 2003). The lack of speed adaptation, found at least in some studies, can also be interpreted as an impairment of the ability to adapt the safety margins to the increased response time, i.e. as a degradation of regulatory control.

A summary of documented and potential effects of different example IVIS and ADAS are presented in *Table 1*.

Table 1 Examples of ADAS/IVIS and their potential effects on driving performance

Function	Supported control layer	Direct (intended) effects on driving performance	Examples of potential indirect effects on driving performance	References
ABS	Tracking	Enhanced longitudinal (braking) control	Reduced headway/increased speed	Fosser, Saetermo and Sagberg (1997)
Lane departure warning	Regulating	Enhanced lateral control	Over-reliance -> reduced control when system malfunctions	Burns (2001)
Speed alert	Monitoring	Better speed keeping	Over-reliance -> reduced control when system malfunctions	Hjälmdahl and Varhelyi (2004)
			Visual distraction -> reduced lateral and longitudinal control, reduced event detection performance	No reference found
Navigation support	Monitoring	Improved route finding	Over-reliance -> get lost when the system gives errant guidance	No reference found
			Visual distraction -> reduced lateral and longitudinal tracking control, reduced event detection performance	Antin, Dingus, Hulse and Wierwille (1990)
Phone	Non-driving related	As small as possible	Dialling -> visual distraction -> reduced lateral and longitudinal tracking control, reduced event detection performance Conversation -> cognitive distraction -> More focused tracking control, reduced event detection performance	Engström, Johansson and Östlund (2005); Horray and Wickens (2004); Patten et al. (2004)
Action scheduling (AIDE metafunction)	N/A	Improvements on all layers	Indirect effects have not been studied – difficult to predict.	No reference found

2.4.5 Driving performance and risk

In the previous section, different types of effects of ADAS and IVIS on driving performance were reviewed. Understanding the relation between such performance effects and actual accident risk is one of the most important (and difficult) issues in current traffic safety research. The difficulties are due to a lack of sufficiently detailed behavioural data in existing accident databases as well as the lack of a basic understanding of the behavioural factors that cause accidents.

Viewed from the perspective of the proposed framework, the relation between driving performance and risk for an individual driver can be described as a complex function involving (at least) the following factors:

1. The current complexity/difficulty of the driving task
2. The driver’s vehicle handling skills (performance on the tracking layer)
3. The ability to adopt safety margins that are appropriate to (1) and (2), based on (4) (determines performance on the regulation layer)
4. The ability to make a correct situation assessment (performance on the monitoring layer)
5. The effort spent on the control tasks on the different layers (2-4)

Based on this view, it is clear that performance degradation on one control layer does not automatically imply increased accident risk. For example, reduced tracking performance (e.g. due to visual distraction) is risky if not properly compensated for on the regulating layer. Thus, risk must be understood in terms of the *relation* between performance on the different layers, where *inadequate adaptation* to the current driving condition and own driving skills could be hypothesised to be a critical factor. One typical example of this is drunk driving, where alcohol is well known to induce overestimation of own performance capability. Thus, in this case, the erroneous safety margin setting is due to the cognitive impairment induced by the drug. Another example is run-off-road accidents due to slippery roads. In this case, the erroneous safety margin setting (reflected e.g. in too high speed in a curve) is due to an erroneous situation assessment on the monitoring layer, which propagates down to the regulatory level, and finally induces instability and breakdown of the tracking control.

However, it is very difficult to determine whether adequate adaptation has been achieved in a particular situation, especially in terms of quantitative performance metrics. One potential approach is to look for *violations to safety margins*. Possible metrics of such violations include the amount of involuntary lane departures and/or the amount of TLC/TTC values below some critical value. However, a basic problem with this is that the adopted safety margins generally differ substantially between drivers (and possibly also vary over time for an individual driver). One metric that explicitly addresses this problem is *steering entropy* (Nakayama, Futami, Nakamura, and Boer, 1999; Boer, 2000; Boer, Rakauskas, Ward and Goodrich, 2005).

As stated above, estimating accident risk based on driving performance metrics requires a holistic assessment of the relation between the different control tasks involved. It is clearly beyond the scope of T2.2.5 to develop a detailed model of these relations. Rather, the goal here is to identify and specify a set of metrics that can be used to quantify performance on relevant control processes on different control layers. These may then be used as potential inputs to behavioural driving risk assessment models. Work on risk assessment is being performed in AIDE WP2.3, although the focus there is more on risk at the collective- (or aggregate-), as opposed to the individual, level.

2.5 Initial selection of driving performance metrics for ADAS/IVIS evaluation

The central goal of WP2.2 is to develop methods and tools that can be applied to the AIDE prototype evaluation in SP3 (formative evaluation) and WP2.4 (final, summative, evaluation). Thus, the basic requirements on the methods need to be derived based on how they are intended to be used in SP3. This section provides a brief summary of the intended AIDE functionalities and how they are planned to be evaluated. On this basis, two general evaluation scenarios are specified from which a number of more detailed requirements on the performance metrics are derived. Based, on these scenarios, an initial set of driving performance metrics was selected.

2.5.1 AIDE functionalities to be evaluated

The AIDE system should be considered as an integrated interface between the driver and multiple in-vehicle systems, including ADAS as well as IVIS. The AIDE prototypes will contain a relatively large number of applications that share a common human-machine interface, including, novel types of input/output devices such as haptic/input output, voice interaction and head-up and configurable displays. The AIDE evaluation methodology needs to cover human-factors issues related both to individual systems/functions/HMI solutions as well as their integration.

A central aspect of the AIDE concept is the AIDE metafunctions, intended to manage other (individual) functions, e.g. prioritisation, message scheduling, I/O allocation etc. The so-called AIDE Design Scenarios developed in SP3 (D3.2.1) focus exclusively on the AIDE metafunctions (and generalise over individual functions). Evaluation of metafunctions has previously proved to be quite difficult (see e.g. the result from the evaluation of the COMUNICAR system, Schindhem et al., 2003) and is rather different from evaluation of individual functions. As mentioned above, the AIDE evaluation methodology should cover both cases.

2.5.2 Formative and summative evaluation

A key distinction for present purposes is that between *formative* and *summative* evaluation.

Formative evaluation is performed early in the design process with the main goal to improve design. It is thus a key part of an iterative human-centered design process with repeated evaluations and re-designs. Formative evaluation requires methods that are cost/time-efficient and easy to use. In AIDE SP3, formative evaluation will start around M20 on the first virtual prototypes.

By contrast, the purpose of summative evaluation is to verify an almost finalized system according to certain criteria, mainly those specified in the requirements phase. In AIDE, this refers to the final prototype evaluation in WP2.4.

Table 2 and *Table 3* provide a description of two types of evaluation scenarios in AIDE. The scenario specified in *Table 2*, represents an instance of formative evaluation in a simulator, while the one given in *Table* represents summative evaluation of the final system in the field. Finally, *Table 4* lists some candidate performance metrics, based on these scenarios and the review in AIDE Deliverable 2.2.1 (Johansson et al., 2004). It should be noted that these tables only include the types of metrics relevant for the present deliverable (performance metrics) and, thus, they could be complemented by other types of metrics developed in AIDE WP2.2.

Table 2 Scenario for AIDE formative evaluation

Formative evaluation of virtual prototypes		
Purpose	Compare different design solutions -> improve design	
Test setting	Driving simulators (of different fidelity)	
General requirements on metrics	1. Should be possible to measure in the simulator 2. Should have a clear safety interpretation	
Hypothesis addressed	*Specific requirements on performance metrics*	*Candidate types of performance metrics*
I. Is the HMI solution X safer than solution Y? *Example: X=traffic info displayed in HUD,* *Y= traffic info displayed in instrument cluster*	1. Shall quantify safety-reducing effects on tracking performance (longitudinal and lateral control	Steering performance metrics Lane keeping metrics LCT (deviation to normative path during commanded lane change) Response time to critical events
	2. Shall quantify safety-reducing effects on regulating control	Headway metrics Lateral safety margin metrics (e.g. TLC)
	3. Shall quantify safety-reducing effects on monitoring control	Speed relative to posted speed limits
II. Is the HMI solution X more safety efficient than solution Y? *Example: X= visual forward collision warning, Y= auditory forward collision warning*	1. Shall quantify enhanced performance of specific functions (specific requirements for each ADAS) 2. Shall be possible to measure in the function's working environment	ISA: Mean speed, speed variation LDW/LKA: Lane exceedances, standard deviation of lateral position FCW: Response time to sudden obstacles
III. Is the metafunction solution X more safety efficient than solution Y? *Example: Resolving conflicts between simultaneous actions* *X= prioritisation (delay one of the actions)* *Y= modality adaptation (present both actions simultaneously in different modalities)*	Shall work in the functions' working environment	Speed Response time

Table 3 Evaluation use case for AIDE summative evaluation

Summative evaluation of demonstrator vehicles		
Purpose	Validate the AIDE system and HMI design as a whole, mainly with respect to safety and user acceptance.	
Test setting	Field trials in demonstrator vehicles.	
General requirements on metrics	1. Should be possible to measure in the field 2. Should have a clear safety interpretation	
Hypothesis addressed	*Specific requirements on performance metrics*	*Candidate types of performance metrics*
I. Is function X safe to use while driving? *Example: X= speech control of mp3 player*	1. Shall quantify safety-reducing effects on tracking performance (longitudinal and lateral control)	Steering performance metrics Lane-keeping metrics feasible to compute based on lane-tracker data
	2. Shall quantify safety-reducing effects on regulating control	Conflict measures (observation protocol) Vehicle following metrics
	3. Shall quantify safety-reducing effects on monitoring control	Speed relative to posted speed limit
II. Does function X enhance safety as intended? *Example: X= Lane departure warning (LDW)*	1. Shall quantify enhanced performance of specific functions (specific requirements for each ADAS). 2. Shall be possible to measure in the function/system's working environment	ISA: Speed metrics LDW/LKA: Lane keeping metrics FCW: Response time to sudden obstacles
III. Does AIDE meta-function X enhance safety as intended? *Example: X= action re-scheduling in demanding situations*	Shall work in the functions' working environment	Mean speed (in free driving)

Table 4 Candidate performance metrics (based on the review in D2.2.1)

Metric type	Metric	Key references
Steering performance metrics	High frequency component (HFC) of steering wheel angle	McLean and Hoffman (1975); Östlund et al. (2004)
	Steering wheel reversal rate	McLean and Hoffman (1975); Östlund et al. (2004)
	Steering wheel action rate	Verwey (2000)
	Rapid Steering Wheel Turns	Östlund et al. (2004)
	Steering entropy	Nakayama et al. (1999), Boer (2000); Boer, Rakauskas, Ward and Goodrich (2005)
	Standard deviation of steering wheel angle	Liu, Schreiner and Dingus (1999)
Lane keeping metrics	Mean lateral position	Östlund et al. (2004)
	Standard deviation/variance of lateral position	Wierwille et al. (1996); Östlund et al. (2004)
	Lane exceedences (Lanex)	(Wierwille et al. 1996); Östlund et al. (2004)
	Mean TLC	Godthelp et al., (1984)
	15% level TLC	Godthelp et al., (1984)
	Minimum TLC	Östlund et al. (2004)
	Mean value of the min TLC values	Östlund et al. (2004)
	The proportion of TLC min values < X s	Östlund et al. (2004)
Speed metrics	Mean speed	Östlund et al. (2004)
	Standard deviation/variance of speed	Östlund et al. (2004)
	Maximum speed	Östlund et al. (2004)
Vehicle following metrics	Mean time headway	Östlund et al. (2004)
	Standard deviation of time headway	Östlund et al. (2004)
	Minimum time headway	Östlund et al. (2004)
	Mean time headway	Östlund et al. (2004)
	Standard deviation of time headway	Östlund et al. (2004)
	Minimum time headway	Östlund et al. (2004)
	Mean of TTC local minima	Östlund et al. (2004)
	Minimum TTC	Östlund et al. (2004)
	Time Exposed TTC (TET)	Östlund et al. (2004)
	Coherence	Brookhuis, de Ward and Mulder (1994)
	Phase shift	Brookhuis, de Ward and Mulder (1994)
	Modulus	Brookhuis, de Ward and Mulder (1994)
Response time metrics	Brake response time	Green (1993); Wierwille et al. (1996)
	Steering avoidance response time	Wierwille et al. (1996)
LCT metrics	Deviation from normative path	Mattes (2003)
Steering grip metrics	Steering grip force variation	Peters et al (2005)

3 Vehicle Control Metrics

The purpose of this chapter is to provide the basis for the choice of vehicle control metrics to be proposed by this deliverable. Several metrics are described and analysed with respect to their feasibility to be used in the evaluation of IVIS and ADAS. The initial choice of metrics to be included in this chapter is based on a long European transportation research tradition, reflected primarily in the European research projects such as HASTE.

In this chapter, existing and new analyses of previously collected data as well as new small scale experiments, lay ground for the final recommendation of metrics, and also the development of some new metrics. The most relevant factors in the analyses of metrics are their sensitivity, reliability and applicability in different road environments. It is of high priority that the recommended metrics provide valid and reliable results in a representative range of scenarios where IVIS and ADAS are likely to affect driving performance. Since most of the work was based on re-analysis of data from the HASTE project, a brief description of this data set is provided in the next section.

3.1 The HASTE data set

Most of the analyses in this chapter were based on data originally collected in work package (WP) 2 of the HASTE FP5 EU-funded project. Thus, this section provides a general description of this data set. The work in HASTE WP2 comprised 10 parallel experiments, conducted at 8 different sites across Europe and Canada during 2003-2004. The main purpose of these experiments was to investigate systematically the effects of visual and cognitive secondary task load on driver behaviour and performance. Thus, all experiments had the same general design and differed mainly in terms of the test setting, which included desktop simulators, medium-to-high fidelity simulators as well as field trials. The results are reported in Östlund et al. (2004) and in a series of papers contained in a special issue of Transportation Research Part F (No. 8 2005). The results from the HASTE analysis of the data used for the present work are also reported in Engström et al. (2005) and Victor et al. (2005).

Visual and cognitive secondary task load was varied systematically by means of two so-called surrogate (S-) IVIS, one visual and one auditory/cognitive. The visual S-IVIS, known as the Arrows Task, was to detect the presence of an arrow with upward direction in a display of arrows. The difficulty of the task was varied by changing the orientation and number of the non-target arrows. If the arrow was present, the subject was instructed to press “yes” on a touch display and “no” otherwise. Each task consisted of 6 arrows displays, presented every 5 s. The set-up for the Arrows task is illustrated in Figure 5. For a more elaborate description of this task, se Östlund et al (2004).

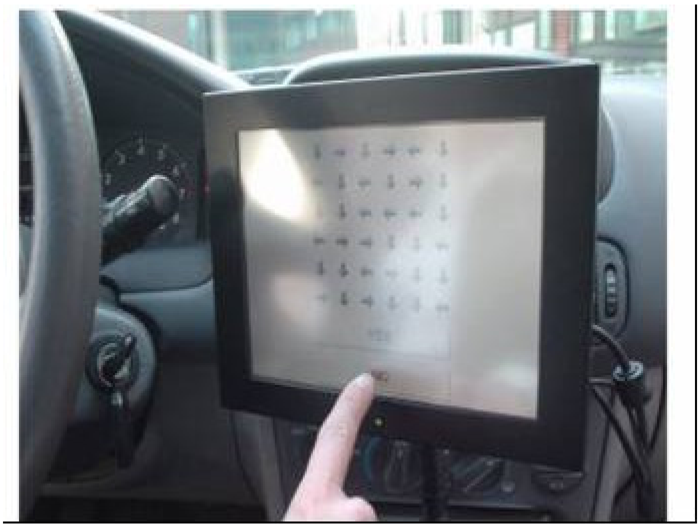


Figure 5 Set-up for the Arrows Task

The cognitive task, called the Auditory Continuous Memory Task (aCMT), involved the presentation of fifteen sounds at a rate of 2 seconds. The task was to keep track of certain target sounds identified before the trial. The difficulty was varied by changing the number of target sounds. The task lasted about 30-40 seconds. For both the Arrows Task and the aCMT task there were three difficulty levels, with level three being the most difficult setting. See Östlund et al. (2004) for more detailed descriptions of the S-IVIS tasks.

The data used for the present analyses were collected in three settings: (1) the Volvo Technology fixed-base simulator, (2) the VTI moving-base simulator and (3) a field study using a Volvo S80. In the simulators, data from two road types, motorway and rural road, were used. These simulated roads were identical for the two simulators. The simulated motorway had two lanes in each direction with a separating rail between them. The lane width was 3.75 m and the speed limit was 110 km/h. The simulated rural road had one lane in each direction, a lane width of 3.65 meters and a speed limit of 90 km/h. For the simulated rural road, straight and curved sections were included. The latter were gentle s-shaped curves, which required some negotiation by the driver. Finally, the field study was conducted on a motorway outside Linköping, Sweden. The road was similar, but not identical, to the simulated motorway. A wide range of behavioural data was collected, including driving performance data, eye-movements, physiological data and subjective ratings.

3.2 Speed Metrics

3.2.1 Introduction

Due to the simple measurement, speed metrics are the most commonly used metrics in driving behaviour studies. Mean speed, speed variation and maximum speed are frequently used speed metrics. A typical application of speed metrics is evaluating speed reducing effects of road design and road signs in accident prone road sections. Similarly, the speed reducing effects of e.g. ISA (Intelligent Speed Adaptor) can be evaluated using speed metrics.

The effects of mental workload and distraction on speed are however not as easily interpreted. In several studies (e.g. Patten, Kircher, Östlund & Nilsson, 2004, and all experiments including visual distraction in the HASTE-project, Östlund et al., 2004) it has been found that visual distraction leads to decreased travel speed, which has been attributed to the idea that the driver reduces speed in order to cope with the reduced visual input. This can also be described in control theoretical terms; the driver reduces speed in order to gain time for the tracking control loop (see the Theoretical Framework in chapter 2). Cognitive load, however, has been found to influence speed less, and also inconsistently. The results from the HASTE project showed that speed was unaffected by the cognitive task (Östlund et al, 2004), although there were some indications in one study that the speed increased for the highest cognitive load levels. The same was found in a study on mobile phones (Patten et al, 2004), where speed tended to increase while the drivers were engaged in a mobile phone conversation with a hands free unit. This effect may indicate a loss of speed monitoring, which would be a dangerous driver state.

Speed variation is influenced by voluntary speed changes due to the road environment, by the interaction with other road users and by involuntary changes due to loss of speed control. Speed variation can e.g. be used as a metric of driver response in situations requiring the driver to change the speed. However, it has also been used in rather unrestricted driving conditions, where the relation between speed variation and risk of accident is difficult to comprehend and explain. In such scenarios there are more useful alternatives: e.g. average speed and headway metrics, which are more easily interpreted in terms of driving performance and risk.

As for lateral position variation, speed variation is influenced by data duration (time window used in the analysis) since speed variations can vary from very fast (e.g. braking before a narrow curve) to very slow (e.g. travelling slightly faster in areas where the visibility is good and the road is straight). A consequence of this is that any comparisons of data of different lengths mainly reveal effects of data length (time window), other effects might be obscured. Thus, it is recommended that the same time window is used when comparing results.

3.2.2 Objectives

The objective of this study was to determine the characteristics of speed variation in order to possibly develop a different method for calculating speed variation that is independent of data length (time window). The new method is compared to conventional speed variation with respect to applicability in different scenarios. The new method is based on high-pass filtering speed data before the standard deviation is calculated. Finally, the cut-off frequency is specified.

3.2.3 Method

Motorway field data from a HASTE WP2 experiment and rural road simulator data from a VTI road design study (not yet published) were used to calculate examples of effects of data length on speed variation. The data set was derived from an experiment with average non professional drivers (age 25-50), no secondary task, just driving, low traffic density, no critical events and no planned car following situations. A total of 60 hours of driving data was used for the analysis, including 84 participants.

Data length (time window) dependencies were investigated by taking partly overlapping speed data sections of one second to ten minutes duration and for each section calculating standard deviation of speed. All values based on the same data length were averaged. All mean values were stored in an array, representing 1 to 600 seconds speed data lengths.

3.2.4 Results

It was found that even after ten minutes speed variation was still influenced by data length (time window) (see *Figure 6*). Motorway real car driving resulted in less variation, but with the same progression of the variation over time. Obviously, data of different lengths are thus not comparable due to very slow variations in speed.

In order to make it possible to compare data of different lengths, the slow variations have to be eliminated. For this, a standard high pass filter was used on speed data, resulting in speed variation, calculated as the standard deviation of the filtered speed data, being uninfluenced by data lengths *over the filter time period*. Below this value, speed variation is not valid. The cut-off frequency influences the variations that can be detected. Since AIDE focuses on interaction with IVIS and the use of ADAS, data to be analysed will be rather short, typically from ten seconds to a minute. Thus, a cut-off frequency of 0.1 Hz is applied, resulting in a minimal valid data length of 10 seconds (see *Figure 6*).

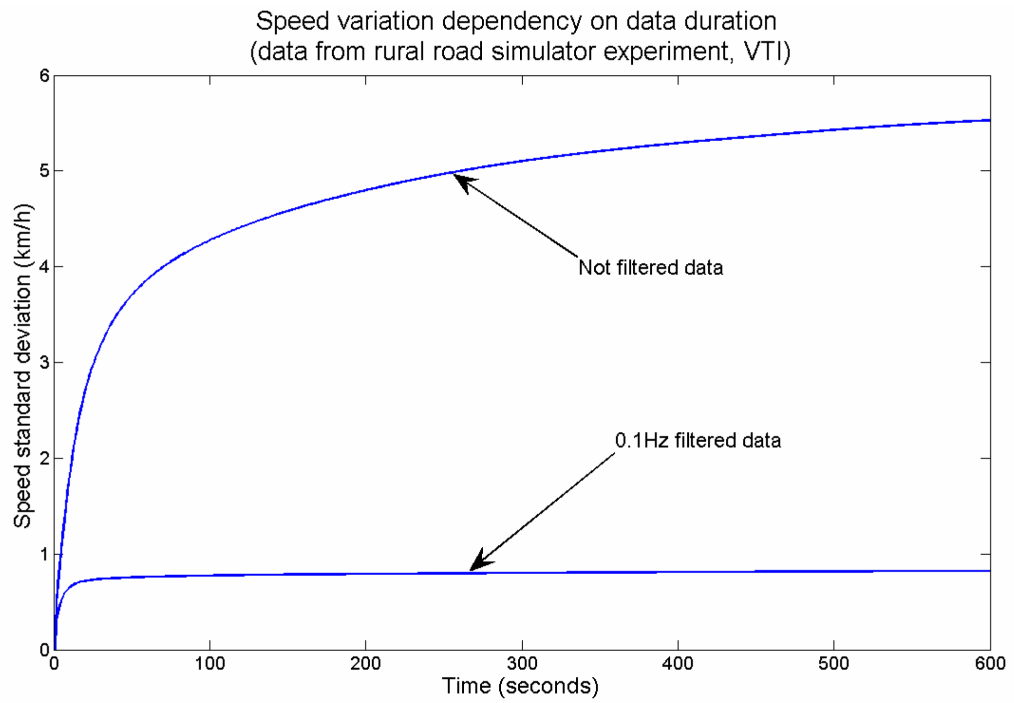


Figure 6 Speed variation progression with data length. Data were taken from the VTI & VTEC WP2 motorway field experiment in the HASTE project (Östlund et al, 2004). The condition was low traffic density.

3.2.5 Discussion and conclusions

As can be seen in the results, speed variation increases over long time and a 0.1 Hz filter reduces a stationary variation value (after very long time) to less than 20%. Due to this large effect of the filter on the variation and due to the unspecific nature of speed variation, this metric is not recommended to be included in IVIS and ADAS evaluations. However, mean speed is recommended for inclusion in IVIS and ADAS evaluations.

3.3 Lateral position metrics

3.3.1 Introduction

Next to speed, lateral position is one of the most commonly used driving behaviour metric. Mean lateral position is used as a metric of driving strategy, a measure of the driver choice to drive within a safe path of travel (compare to Gibson and Crooks, 1938). A distracted driver may move away from what is considered less safe, e.g. oncoming traffic or a steep road side. This is also how lateral position is recommended to be interpreted, requiring the analysis of lateral position data to consider the current driving scenarios.

Lateral position variation is influenced by unintentional variations of lateral position caused by tracking error – the difficulty to drive completely straight. Further, the road curvature influences the choice of path of travel along the road. In distraction studies, mainly the tracking error is of interest. These variations are faster than other variations (see e.g. the ECOM model described in Chapter 2. and in (Hollnagel and Woods, 2005). They are however still difficult to accurately and precisely separate from other variations in lateral control.

Increased lateral position variation is commonly related to a reduced lateral control. *Reduced* lateral position variation has been found when drivers are under cognitive load (Brookhuis, de Vries and de Ward, 1991; Engström et al., 2005). However, the interpretations of this behaviour differ. One explanation is that the driver is aware of being mentally distracted and thus having less capacity for making fast and correct decisions e.g. in the interaction with other road users. Thus, the driver compensates by investing more effort in the lateral control in order to increase the safety margins. Another explanation is that cognitive load results in the driver focusing the gaze on the road ahead (see e.g. HASTE Deliverable 2) and thus becomes able to control lateral position more precisely. The reason for choosing not to scan the peripheral road scene may be caused by the driver being on the verge to being cognitively overloaded and thus not having more capacity for information processing (see Wickens resource model in Wickens (1992)). Anyway, reduced variation in lateral position when engaged with a cognitive task should not be interpreted as beneficial for driving performance. Instead, this could be interpreted as a symptom of driver overload and increased risk of incorrect decisions due to being engaged in a distracting task.

Lateral position is easily measured in simulators. In field experiments lateral position requires a reliable and accurate sensor. Also, the lane width should be constant not to give irrelevant data variations. A wide hard shoulder might make the driver choose to drive closer to the road edge compared to a narrow hard shoulder (Peters & Östlund, 2005). Such variations might be difficult to cope with, but not impossible problems to solve. Attention has to be given this when choosing experimental route and vehicle sensors!

Lateral position variation is commonly calculated as the lateral position standard deviation (SDLP). But since the variations in LP are rather slow, SDLP becomes highly correlated to data duration – as is also the case for speed variation. In *Figure 7* SDLP dependency on data duration is displayed. This problem has not been subject to any previous study, except to a small scale in the HASTE project.

It can be, and is by some, argued that time dependence is not an unwelcome effect, but rather reflects that within a short time duration, the driver cannot vary the lateral position very much, indicating a low risk of running off the road. We however already know that long exposure (time on task) to distraction is hazardous. Here, however, we focus on the effects on driving *when* the driver is distracted. If exposure is assessed, task duration should be measured separately as a risk metric. Finally, if data length (time window) contaminates SDLP, all other effects are likely to disappear since data length so heavily influences SDLP. Therefore, SDLP should be made independent of data length i.e. which time window is used in the analysis.

3.3.2 Objective

The objective of this study was to determine the characteristics of lateral position variation in order to possibly develop an alternative method to calculate lateral position variation that is independent of data length. The new method is based on high-pass filtering LP data before standard deviation is calculated. The cut-off frequency will be specified, but the first approach will be 0.1 Hz. This frequency is found as the break point in steering spectral density, over which steering variation most reliably is influenced by visual distraction (see *3.8 Steering wheel metrics*). Since steering characteristics are transmitted to lateral position characteristics, this frequency is a relevant starting point.

3.3.3 Method

HASTE WP2 rural road simulator data were used for analysing the frequency characteristics of lateral position data. Data were collected in the high fidelity VTI Driving Simulator II (see www.vti.se for description). The following conditions were included in the analysis: secondary task (visual, cognitive, no task) and road curvature (straight, curved). A total of 16 hours of data was used. Data were derived from an experiment with 48 participants. The frequency characteristics of LP data were analysed using spectral density calculations. Also, the effect of data duration on LP-variation was assessed using the same method as used on speed variation (see previous chapter), although the maximum data duration was set to 100 seconds (600 for speed data).

3.3.4 Results

See *Figure 8* for a plot on lateral position spectral density. Using t-test for comparing differences between the spectral plot densities, it was found that visual distraction resulted in an increased spectral density compared to baseline (no secondary task), and that cognitive load resulted in less spectral density than baseline ($p<0.05$). It was however not possible to identify any specific area in the spectral density that was more influenced by visual/cognitive distraction than another. Rather, the whole spectral density was influenced. Although not illustrated here, it was also found that curved road resulted in higher spectral density compared to straight road, but the effects of visual and cognitive load were the same as for straight road sections. The effect size of road curvature is highly dependent of the frequency and radii of the curves.

The slow variations in LP data were also reflected in the data duration dependency plot, Figure 7. A high pass filter with 0.1 Hz cut off frequency was applied on LP data and a new data length dependency curve was calculated. As can be seen in Figure 7, this makes the variation constant after approximately 10 seconds. As can also be seen, a large part of the variation still remains after the filter. 0.5 Hz cut-off frequency was also tested, but this filter reduced the total variation only a fraction of the total variation found for the 0.1 Hz filter.

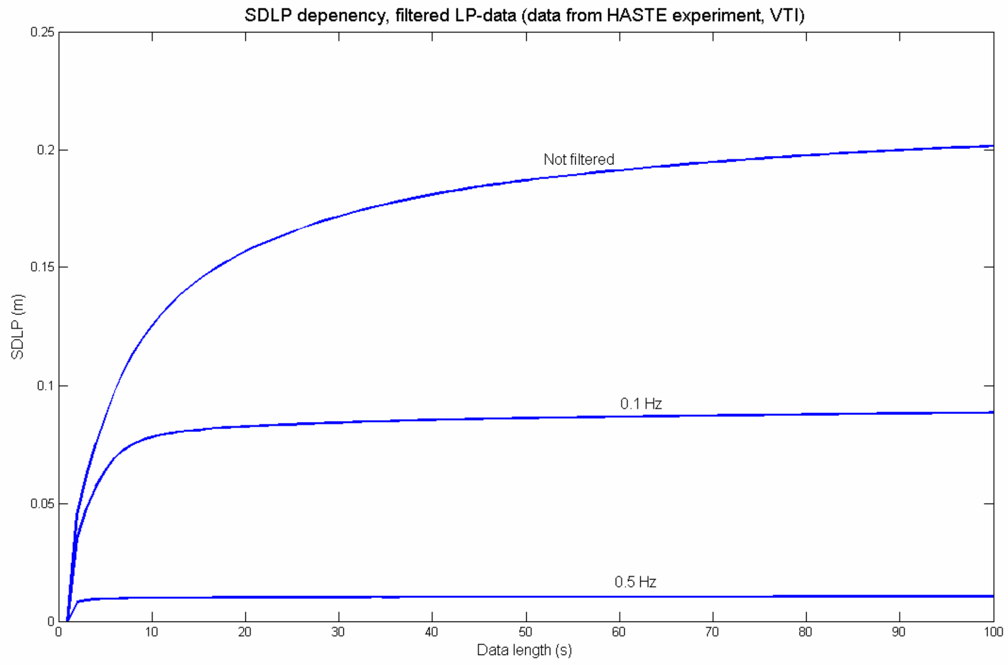


Figure 7 Lateral position variation progression with data length. Data were taken from the VTI WP2 Rural Road experiment in the HASTE project (Östlund et al, 2004). Original data and data filtered with 0.1 Hz and 0.5 Hz are included in the figure. Straight and curved road data combined.

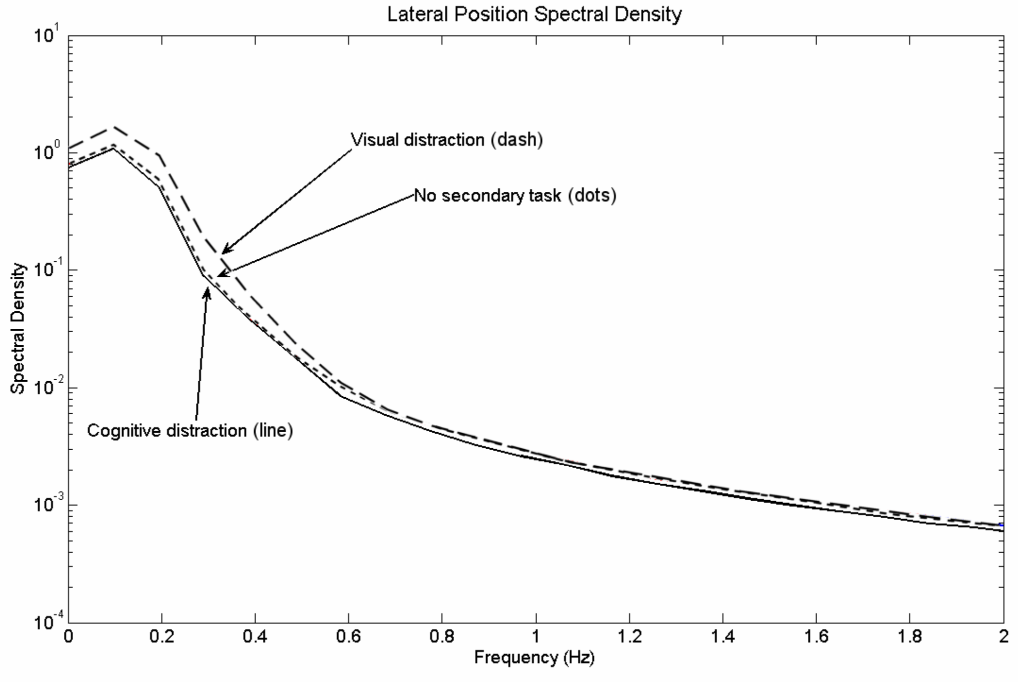


Figure 8 Lateral position spectral density. Data were taken from the VTI WP2 Rural Road experiment in the HASTE project (Östlund et al., 2004). Data for visually distracted (dashed), cognitively loaded (line) and non distracted drivers (dots) are displayed. Straight and curved road data combined.

3.3.5 Discussion and conclusions

Visual distraction was assumed to cause tracking error but not any effect in voluntary lateral position control variation. It was however found that visual distraction caused a general increase in lateral position spectral density, indicating that tracking error is not easily discriminated from other variations. An analogy to these results was found in steering activity data, where it was found that visual distraction caused a general steering activity spectral density increase (see section *3.8 Steering wheel metrics*). Cognitive load was assumed to cause a decreased tracking error, but a maintained voluntary variation in lateral position. However, cognitive load caused a general decrease in lateral position spectral density. Thus, even in this case we find that tracking effects are not easily discriminated from other effects.

Although a definitive cut-off point between tracking error and voluntary lateral position variations is difficult to define, a filter has to be applied to lateral position data in order to eliminate the effect of data duration when calculating lateral position standard deviation. The filter that was applied resulted in SDLP being uninfluenced by data lengths *over the filter time period*. Below this value, SDLP will not be valid, which is reasonable since slow variations cannot be represented within very short time frames. The cut-off frequency, and consequently the shortest data length that is possible to analyse, is a trade-off between sensitivity to tracking error, exclusion of voluntary steering actions and normal data lengths that are to be analysed in the experiments. Since AIDE focuses on interaction with IVIS and the use of ADAS, data to be analysed will be rather short, typically from ten seconds to a minute. Thus, the cut-off frequency of 0.1 Hz is recommended, resulting in a minimal valid data length of 10 seconds. Also 0.5 Hz was tested but was found to eliminate too much of the total variation. See Figure 7. This is also found when studying the spectral density; over 0.5 Hz the spectral density is less than 2% of what is found for 0.1 Hz. The major part of lateral position variation is found in the 0 - 0.5 Hz frequency band. 0.5 Hz cut-off frequency was thus too high. 0.1 Hz is a compromise between feasibly short required data length (10 seconds) and including sources of lateral position variation.

Since road curvature was reflected over the entire lateral position spectrum, this effect remains also after the proposed filter. Similar curvature characteristics should therefore be included for different experimental conditions in driving experiments in order to control for the curvature effects on SDLP.

A different technique was tested in the HASTE-project, using a sliding window averaging method for eliminating slow lateral position variations. This method was found better than conventional lateral position variation since the dependency of data length was eliminated. Using a square sliding window is however far from an optional filtering technique. The problem with using a square or even hanning-shaped window for averaging and subtracting from original data lateral position data is that relevant frequencies are damped, and that the damping effect on frequencies slower than 10 seconds is quite poor, except for 0 Hz level of course. A static level plus a variation of a frequency that is a multiple of 1/T Hz (0.1, 0.2, 0.3 etc), where T is the window length, is always perfectly brought to zero level with maintained variation. But any other frequency, and especially the frequencies in between (i.e. 0.15, 0.25, 0.35 Hz etc) is also damped by the filter. See Figure 9.

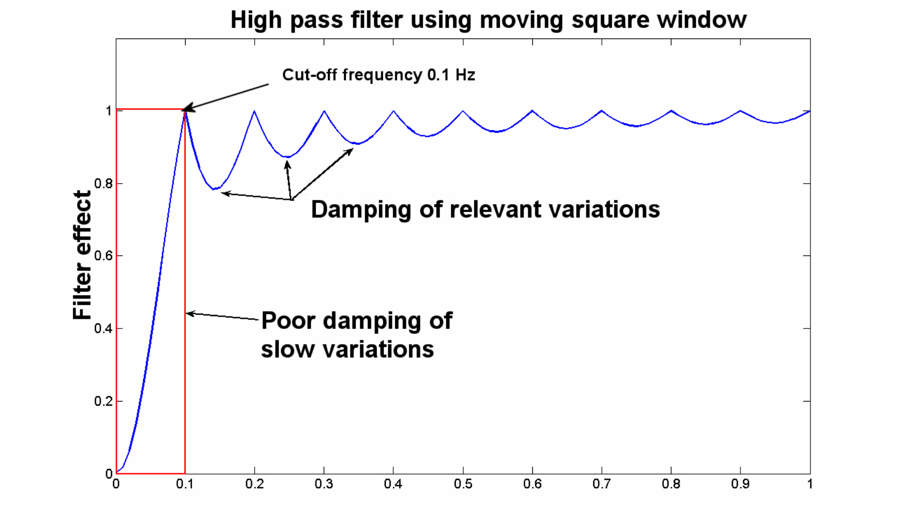


Figure 9 High pass filter effect of using a moving square window in time domain

The recommended lateral position measures are thus mean lateral position and modified lateral position variation. Conventional lateral position variation (std of LP-data) is considered too dependent of data duration.

3.4 Time to Line-crossing metrics

Time-to-line crossing (TLC) was first developed by Godthelp and Konings (1981). TLC is defined as the time to reach the lane marking assuming fixed heading angle and constant speed. Approximations are however conventionally used (van Winsum & Godthelp, 1996), since they require fewer parameters than true TLC and also can be applied on field experiment data. Calculation of approximate TLC is done with lateral position data. A description of the calculation of TLC can be found in Appendix 2.

TLC data has typical waveforms with local minima. Certain rules of thumb are used to identify TLC minima. The local minima are used to calculate several TLC-metrics. These are e.g. mean, median and 15% percentile of TLC minima. For short data durations, however, only few minima will be found – if any. In the HASTE project, task lengths of approximately 5-50 seconds were used. Significant effects of visual or cognitive distraction were found on mean TLC for task lengths longer than approximately 10 seconds. For shorter IVIS/ADAS tasks there may thus be a risk that no effects are found due to the fact that there is too few TLC-minima. Complex metrics, such as 15% percentile, must be considered very unfeasible for data less than several minutes. In AIDE we can expect short data durations since we will evaluate data only during ADAS activity and IVIS interaction. Thus, the included TLC metrics must rely on very few minima, which is the case for mean TLC.

One big problem with TLC is that if the lane markings do not represent the safe travel path as perceived by the driver (Peters & Östlund, 2005), either very large TLC values will be found, or there will be several line crossings, resulting in unreliable TLC metrics which are very difficult to interpret.

Two TLC metrics are recommended for IVIS and ADAS evaluations: Mean TLC and number of line crossings. These metrics reflect the drivers' ability to stay on the road. More complex

TLC metrics are not recommended since these require several minutes of data to provide reliable results.

3.5 Headway metrics

Headway is either measured in seconds or in metres and represents the gap or distance to a lead vehicle. Based on this data, several metrics can be calculated. Minimum headway reflects the least margin to a lead vehicle. Mean headway may reflect car following strategy, i.e. choice of longitudinal safety margin, and situation awareness. But this requires a steady state car following situation. Comparable, steady state car following situations should not be expected to occur often in field experiments, only in simulator experiments. But also in simulator experiments, there are problems with e.g. mean headway. Any transient effects in the beginning of the situation should be excluded from data, as should also the end phase of the situation. Also, driver strategy depends on the experimental instructions. If these problems however can be reliably solved, mean headway is a sensitive indicator of two phenomena: (1) The driver compensating for visual or cognitive distraction, reflected in increased headway, or (2) failing in monitoring of the own travel speed and the behaviour of surrounding road users, reflected in decreased or increased headway. As with most metrics, the interpretation of any effects must be made with the scenario and other effects in consideration. Time headway was found more sensitive and reliable than distance headway in HASTE, this is why distance headway metrics are not included here.

Time-To-Collision (TTC) is closely related to headway and is defined as the time to collision with a lead vehicle in the travel path if the speeds of the vehicles are maintained (speed difference divided by the distance to a lead vehicle). The idea of TTC as a performance indicator is mainly the same as for headway, it measures the longitudinal margin to lead vehicles or objects. But TTC has the advantage of taking the speed difference between the vehicles into account, which indeed is a safety related factor. A disadvantage of TTC is that only the TTC minima are analysed, which are single data points rather than averaged driving behaviour during a prolonged time as is the case for mean headway. Therefore, TTC measures vary more than headway measures, resulting in less statistical power. Further, headway generates continuous data ranging from e.g. 0.2 to 3 seconds in car following situations, but TTC generates data ranging from e.g. 1 second to infinity or undefined when the speed difference between the vehicles is zero, which also is a source of large variation in TTC metrics. In order to TTC metrics to be feasible, a large amount of TTC minima have to be collected, corresponding to a large amount of lead vehicle or object interaction situations. In HASTE, it was found that time headway metrics were more sensitive than TTC metrics. Based on this discussion and the HASTE findings, headway is prioritised rather than e.g. mean of TTC minima.

Two headway metrics are recommended for IVIS and ADAS evaluations: Mean time headway, which is easy to calculate and reliably reflects safety margin to a lead vehicle or object. Minimum time headway, which reflects rather mistakes in longitudinal control if found very small.

3.6 Brake metrics

Brake reaction time is the most commonly used brake related driving performance metric. It requires a critical event to which the driver has to react quickly in order to avoid a collision. The event has to have a well defined onset in order to achieve a reliable brake reaction time.

This is the reason why brake reaction time is very complicated to use in field experiments – critical events cannot be planned (well, they can if the events are set up, but it is very risky) and it is difficult to define and metric the event onset with sufficient accuracy and precision. Brake reaction time is almost exclusively used in simulator experiments. Still, however, it is difficult to use brake reaction time, because small variations in scenario may cause large effects in brake reaction time (Östlund et al., 2004). Another problem is the risk of learning and anticipation effects in brake reaction.

Nygård (1999) has investigated braking behaviour in conflict situations in field experiments and have found that the criticality of the situations can be measured by change in deceleration (second derivative of speed). Based on these findings, a brake-jerk metric was constructed in the HASTE project (Östlund et al., 2004), and slightly refined here. In HASTE, it was defined as the number of decelerations larger than 10m^2 per second. Here, however, it is a binary metric, which is *yes* or =1 if there was one or several abrupt onsets of the brakes during driving, and *no* or=0 otherwise. The reason is that this metric is preferably used for identifying critical situations rather than statistically comparing decerlation values of critical situations.

Of course, brake jerks will not occur often, only in very hazardous situations. It however indicates deteriorated driving performance – if e.g. video based assessment of the current situation revels that the hazardous situation was caused by the driver being distracted. The strength of the brake jerks metric is that it in contrast to brake reaction time can be used in field experiments.

3.7 Steering grip metrics

3.7.1 Introduction

Measuring steering grip pressure gives good opportunities to directly assess steering control efforts early in the chain of driver-vehicle reactions. Behind this statement lies the assumption that steering grip force reflects driver's efforts put into steering control. Both hands on the steering wheel could indicate that the driver is better prepared to cope with an unexpected event. Also a firmer grip or more active grip on the steering wheel could be an indication of the driver's urge to be in better control of the steering.

A steering grip pressure sensor was developed within the AWAKE project by Autoliv and Hök Instrument (see e.g. Muzet et al, 2004). The sensor consisted of an air-filled elastic tubing moulded in the steering wheel along the perimeter connected to pressure sensors, mounted inside a steering wheel. The grip sensor was divided into two sections, making it possible to discriminate between right and left hand pressure. In the first version of the sensor, absolute pressure was measured. The sensor was sensitive to temperature changes which caused the signal to drift. The prototype was partly reconstructed in order to overcome this flaw. The signal was high pass filtered, rectified and further filtered. Thus, the sensor could not be used to measure absolute pressure. Rather, the final signal reflected changes in the pressure applied, increase and decrease, both changes gave a positive signal change. Grip onsets and releases were reflected as fast and high signal increases. It was still possible to discriminate between left and right side steering actions.

The AIDE internal deliverable *Analysis of steering wheel grip activity in a driving simulator experiment* (Peters, Thorslund, & Östlund, 2005) reports the results of data from the second version sensor was tested. Data collected in the HASTE project were used. It was indicated that grip activity was higher during cognitive secondary tasks compared to without a secondary task (baseline). It was also indicated that visual/manual tasks requiring one hand leaving the steering wheel still resulted in the same total grip activity as during baseline. This indicates that the hand remaining on the steering wheel was twice as active as during baseline (both hands on wheel, no secondary task). Finally, steering grip activity was higher in curves than on straight road sections. The authors concluded that the collected steering grip data reflects steering grip activity.

3.7.2 Objective

The objective of this short study was to analyse the grip response characteristics of the grip sensor prototype, and also propose how the sensor should be refined for being able to provide more elaborate measures of steering activity.

3.7.3 Method

A small set of sensor data (approximately 20 minutes) were collected in a simple experiment. Different levels (0.1, 0.2, 0.3 and 0.4 Bar) of static and varying pressure were applied on the steering wheel. This was accomplished by holding a vigorimeter ball inside the hand and gripping the steering wheel, in which a steering grip sensor was welded. The pressure in the ball was measured in order to apply the predefined pressure levels.

Figure 10 Using the vigorimeter for applying pressure on the steering wheel

3.7.4 Results

It was found that static pressure caused the sensor output to rise quickly and then return to zero within ten seconds. It was accordingly also found that rather quickly varying grip pressure generated a static level output signal. See Figure 11 and Figure 12.

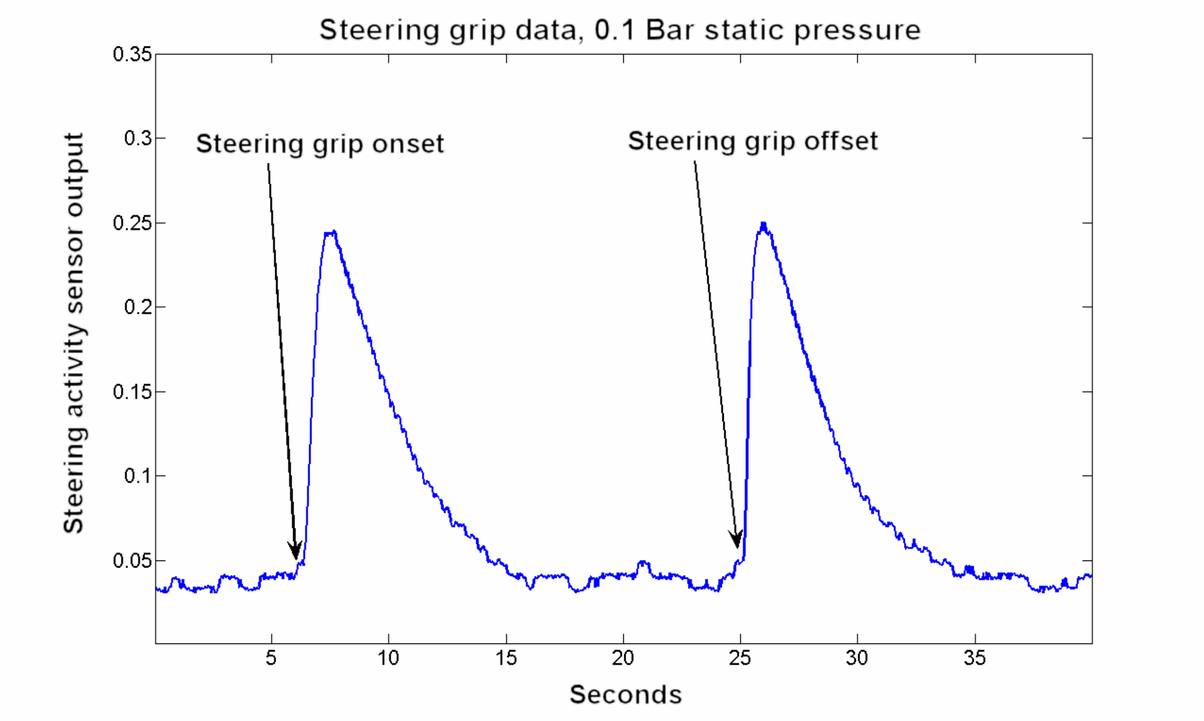


Figure 11 Effect of static steering grip pressure on the steering grip sensor output signal

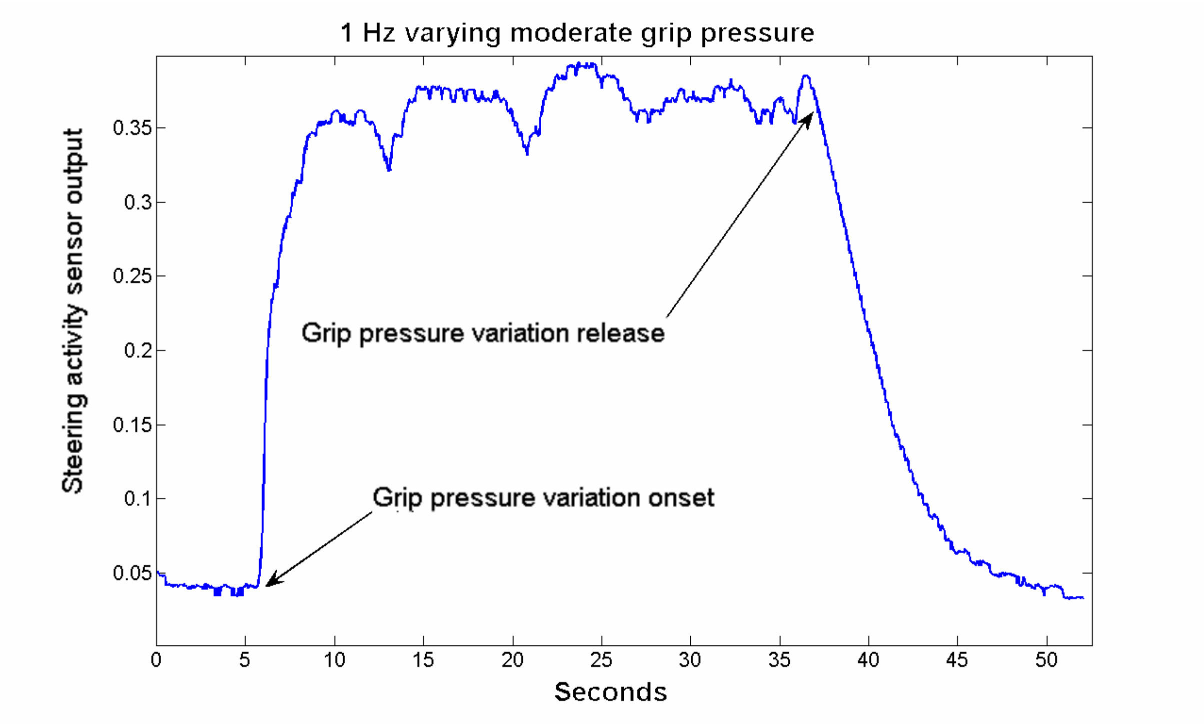


Figure 12 Effect of varying steering grip pressure on the steering grip sensor output signal

3.7.5 Discussion and conclusions

Mean value of the measured signal has been used as the indicator of grip activity. However, a firm but static grip would not be reflected, although it can be assumed that high grip pressure often is related to high grip force variation. In order to be able to measure static pressure, the sensor must be refined or even redesigned. The original problem of temperature induced signal drift may be compensated, perhaps by measuring and including temperature changes in the algorithms.

Steering grip has, during the finalisation of this deliverable, been excluded from the specification of AIDE driving control metrics since the current steering grip data is based on a unique sensor which is unlikely to be available in the prototype evaluation pilots. If included in any pilots, however, the following metrics should be used:

Mean Steering Grip, defined as the mean value of steering grip pressure.

Steering grip variation, defined as the standard deviation of a 0.5-3.0 Hz steering grip pressure signal, measured in the steering wheel. This is thus a metric of steering grip pressure activity.

3.8 Steering wheel metrics

3.8.1 Introduction

Steering is naturally an essential part of the driving control task and performance changes induced by ADAS & IVIS are often manifested in changed steering patterns. A wide variety of steering wheel (SW) metrics have been proposed in the literature, from the standard deviation of steering wheel angle (e.g. Liu et al., 1999), to more advanced metrics such as steering wheel action rate (Verwey, 2000), reversal rate (e.g. MacDonald and Hoffman, 1980; Östlund et al., 2004), high frequency component of steering wheel angle (Östlund et al., 2004) and steering entropy (e.g. Nakayama et al., 1999; Boer, 2000; Boer et al., 2005). See the AIDE deliverable D2.2.1 (Johansson et al., 2005) for a review of these and other existing steering wheel metrics.

For present purposes, the main application of steering wheel metrics is the quantification of effects of secondary task load on the driving task. It has been repeatedly shown that secondary task load yields increased steering wheel activity, both for visual and purely cognitive tasks (e.g. Boer et al., 2005; Engström et al., 2005). A common interpretation of this effect is that SW metrics represent control effort needed to cope with the additional load imposed by the secondary task (e.g. MacDonald and Hoffman, 1980). However, as mentioned in Chapter 2, it has been shown that visual and cognitive secondary tasks have rather different effects on lateral control performance. In a number of parallel experiments conducted within the HASTE project, it was shown that visual secondary tasks increased lateral position variation, while cognitive tasks led to the opposite effect (Östlund et al., 2004; Engström et al., 2005). However, steering wheel activity generally increased in both cases, although there were indications that the two types of secondary tasks led to different types of steering patterns. The visual tasks yielded increased steering wheel movements in a wide range of amplitudes and frequencies, while the cognitive tasks tended to generate small corrective movements only. However, these differences were not further analysed in the HASTE project. While the increased steering wheel activity during visual load can be given a rather straightforward interpretation as a direct result of impaired tracking input during visual time sharing, the effect of cognitive load on control performance is not as easily explained (see Victor et al, 2005 and Engström et al., 2004 for some initial attempts).

The use of steering wheel metrics for HMI evaluation purposes is complicated by the fact that they are sensitive to a number of different factors other than the secondary task itself. First, individuals generally exhibit large differences in steering behaviour. One reason for this is that the same desired outcome, e.g. staying within certain satisfactory safety margins, can be achieved by different means, e.g. different steering strategies (Godthelp et al., 1984; Boer et al., 2005). Steering patterns have also been shown to be influenced by age (Greenshields, 1963, Liu et al., 1977), driving experience (Blaauw et al. 1977), and physiological impairment, such as drowsiness (Knipling and Wierwille, 1994). Environment-related factors known to influence steering wheel movements include driving speed (Blaauw et al., 1977), traffic density (McLean and Hoffman, 1971), lane width (op. cit) and curvature (Östlund et. al., 2004). Finally, steering patterns could be expected to be influenced by vehicle dynamics, such as steering ratio and under/over steering characteristics. Also, in simulator studies, the lack of motion feedback in fixed-base simulators can be expected to lead to less accurate steering performance than in motion-base simulators or real vehicles (Hildreth, et al., 2000).

3.8.2 Objective

As mentioned above, a wide variety of steering wheel metrics have been proposed in the literature. However, the choice of a particular metric is seldom motivated, and the metrics are seldom specified in enough detail to ensure repeatability of results. The general objective of the present study was to identify one or more SW metrics suitable for inclusion in the AIDE evaluation, based on detailed analysis of empirical data. A further goal was to derive detailed specifications, interpretation guidelines and descriptions for these metrics.

3.8.3 Method

The derivation of suitable SW metrics for the AIDE evaluation methodology involved four main steps: (1) obtaining a more detailed understanding of the effects of different types of secondary task load and environmental factors on steering performance, (2) selection of suitable candidate metrics to quantify these effects, (3) identifying the best metrics based on criteria of sensitivity, interpretability and ease of implementation and (4) finding the best parameter settings for the selected metrics.

The analysis was based on the HASTE data set, further described in section 3.1. The different test settings and road types in this data set combine into the following six environment conditions that were compared in the present analysis:

1. Fixed-base simulator, motorway
2. Fixed-base simulator, rural road straight
3. Fixed-base simulator, rural road, curve
4. Moving base simulator, rural road, straight
5. Moving base simulator, rural road, curve
6. Field, motorway

48 subjects participated in each of the two simulator studies. Half of these subjects performed the visual- surrogate IVIS (S-IVIS) task and the other half did the cognitive task. S-IVIS tasks represent IVIS in terms of cognitive and visual load, but are not real IVIS. In the simulators, each subject drove both the motorway and the rural road. In the field study, 24 subjects participated, all performing both the visual and the cognitive task.

The steering wheel angle data, which is of main interest here, were recorded by inductive sensors mounted on the steering column. The rate of steering wheel data sampling was 30 Hz, and the angular resolution of the acquired steering wheel angle signal was 0.1°.

While a thorough statistical analysis of the presently used data can be found in Östlund et al., (2004), Engström et al., (2005) and Victor et al., (2005), the present work focused on a more detailed analysis of the steering wheel data. The analysis was done in a series of steps:

1. Qualitative analysis, with the main purpose to obtain a general understanding of the effects of secondary task load, and environment factors on steering performance. This included time series plots of steering wheel data synchronised with other behavioural data (eye movements, lateral position and speed) as well as spectral frequency plots.
2. Identification of main candidate SW metrics, based on the qualitative analysis in (1) as well as the literature review in AIDE Deliverable 2.2.1 (Johansson et al., 2004).
3. Quantitative comparison of the candidate metrics with respect to their sensitivity to visual and cognitive load
4. Selection of a subset of the metrics best suited for AIDE purposes
5. Identification of the best parameter settings for these metrics.

The work done in each step are further described in the following sections.

3.8.4 Qualitative analysis

The purpose of this analysis was to obtain a better general understanding of how steering behaviour is affected by visual and cognitive secondary task load as well as other factors related to the driving environment and test setting. Two types of qualitative analysis were performed:

1. Time series plots of steering wheel movement data synchronised with eye-movements, lane keeping and speed data.
2. Analysis of frequency content of the steering wheel signals by means of Fast Fourier Transform (FFT)

3.8.4.1 *Time series analysis*

The purpose of this analysis was to make a qualitative comparison of steering wheel movements in baseline and visual/cognitive secondary task conditions. In order to get a better understanding of the mechanisms in play, eye-movements, lateral position and speed data were also included in the plots.

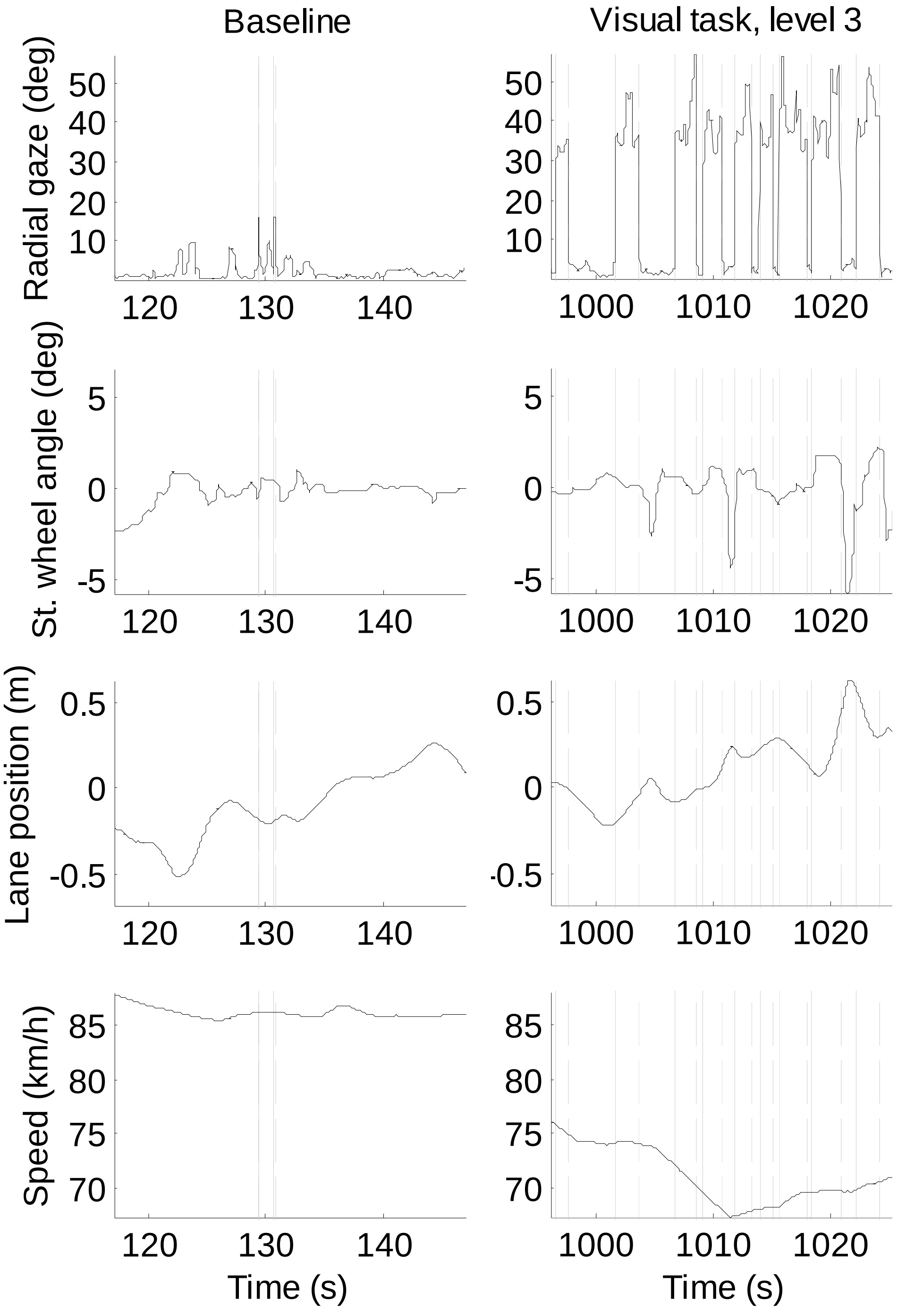
Figure 13 gives an example of synchronised behavioural data for baseline driving and the visual secondary task in straight road driving in the fixed-base simulator[2]. This example clearly shows how visual time sharing between the secondary task and the road results in an intermittent steering strategy, where lane keeping errors build up during glances away from the road and large steering corrections (in the range of 2-6 degrees) are performed during glances back to the road. Moreover, a speed reduction can be observed at the onset of the task. *Figure 14* illustrates the corresponding data (for the same driver) for curve-taking. The same

[2] The eye movement data is represented in terms of radial gaze which is the Euclidean distance from the mode (peak) of the gazes cluster towards the road centre.

general pattern occurs although the steering corrections are somewhat larger (in the range of 5-10 degrees) and the gazes away from the road are less frequent (this effect of curvature on gaze was shown to be statistically significant in the analysis (see Östlund et al., 2004; Victor, Harbluk and Engström, 2005). This general effect of the visual task data was fairly consistent across subjects and test settings, and in line with the statistical results reported in Östlund at al. (2004), Engström et al., (2005) and Victor et al., (2005).

Corresponding data for the cognitive task is shown in *Figure 15* (straight road) and *Figure 16* (curve). It can be observed that the effects are quite different compared to the visual secondary task. During the task there are almost no glances away from the road, while the baseline data contains several such glances. This is an instance of the well-documented gaze-concentration effect of cognitive load (e.g. Recarte and Nunes, 2000, 2003; Harbluk and Noy, 2002; Victor, Harbluk and Engström, 2005). It can also be observed that the lane keeping is more precise during the cognitive task than baseline, and there is a tendency for a greater amount of more small steering corrections (with a magnitude of about 1 degree). These steering micro-corrections are also evident in the data of Boer et al. (2005).

Finally, *Figure 17* illustrates examples of steering wheel movement time series from the field study. The general patterns are consistent with the simulator data presented in the previous figures, although the patterns are not as "clean" as in the simulator data, probably due to "contamination" from real-world factors such as traffic events, wind gusts, uneven road surface etc.



2

Figure 13 Synchronised behavioural data for baseline and visual task, straight road driving in the fixed-base simulator.

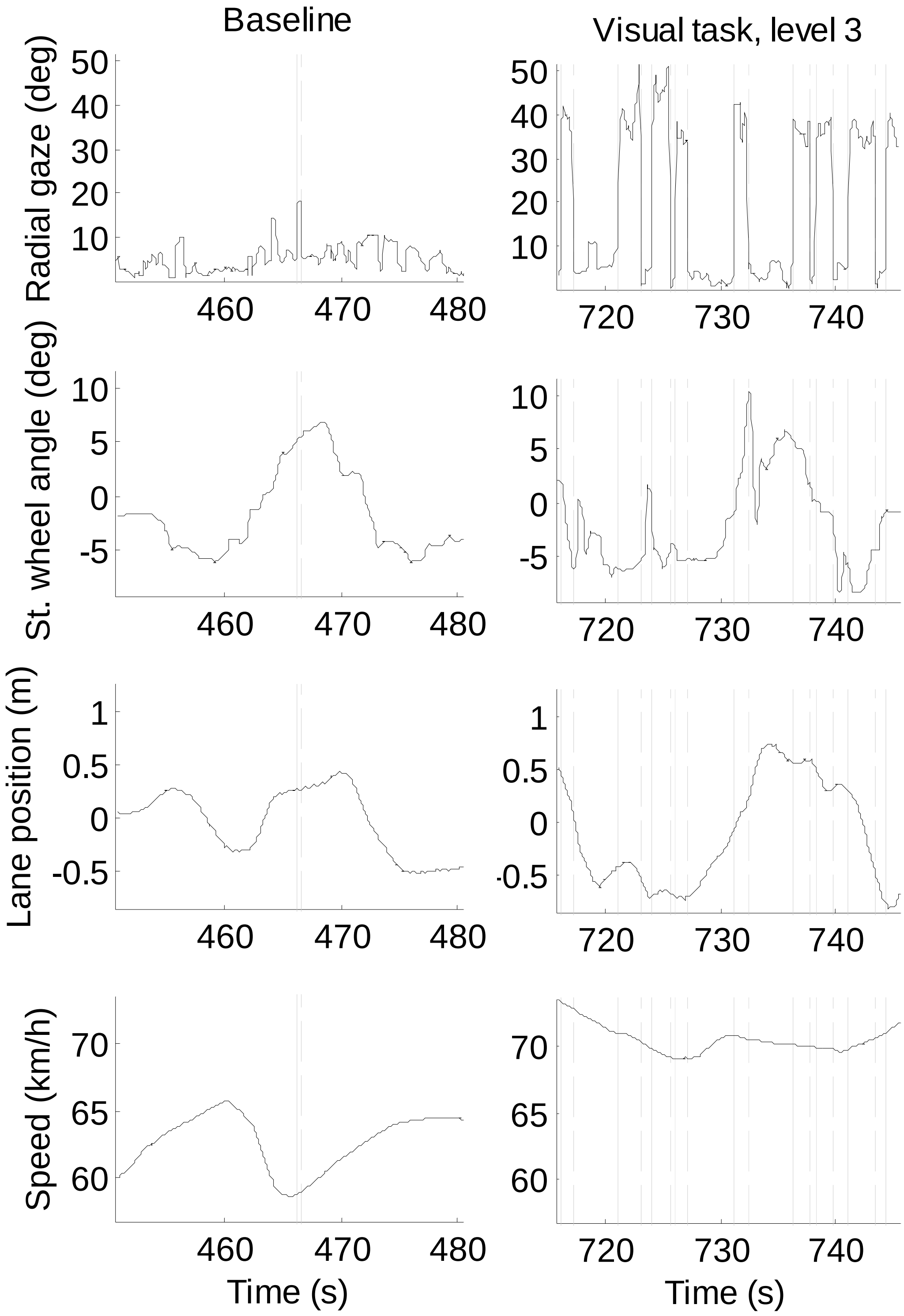


Figure 14 Synchronised behavioural data for baseline and visual task, curve driving in the fixed-base simulator

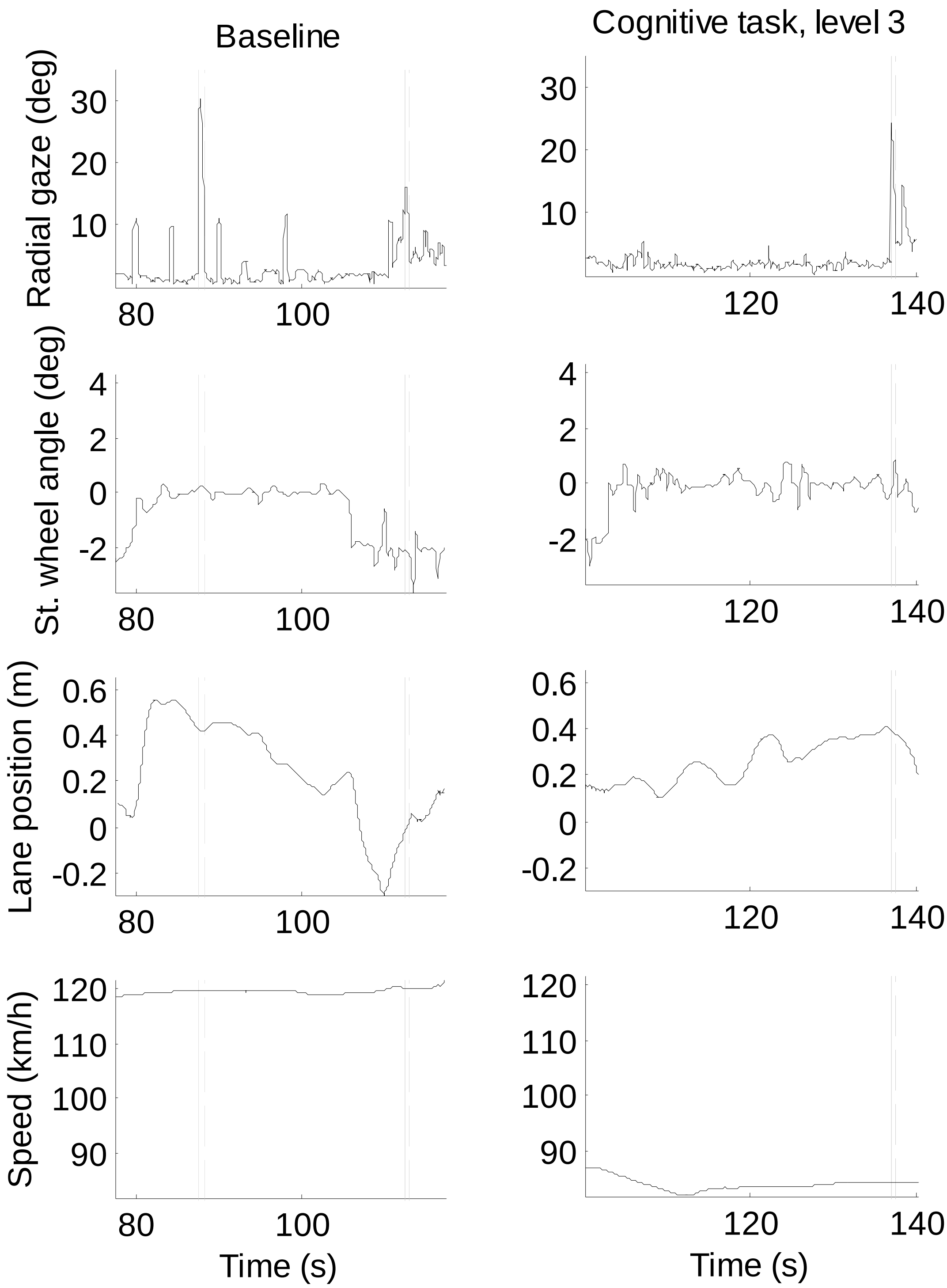


Figure 15 Synchronised behavioural data for baseline and cognitive task, straight driving in the fixed-base simulator.

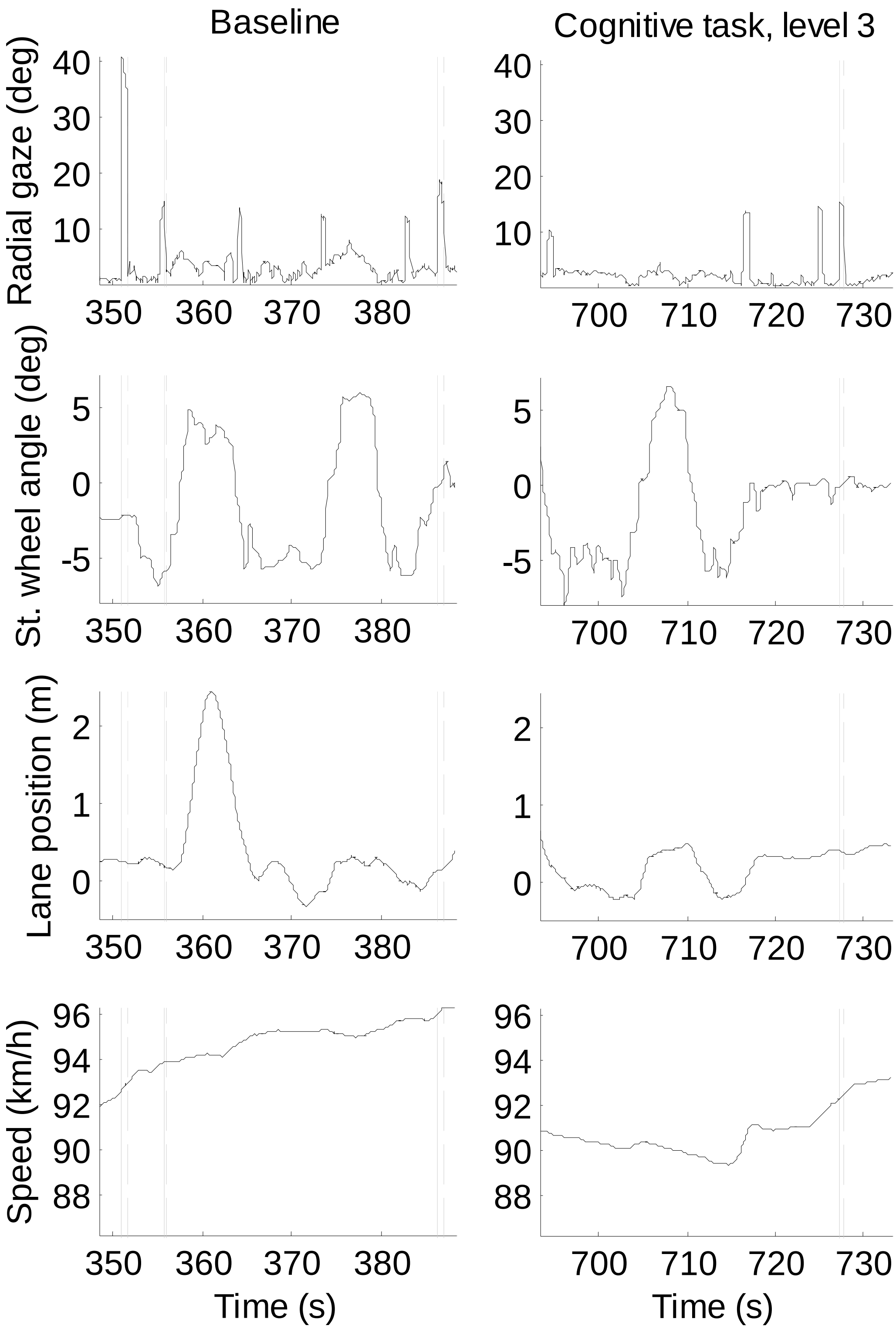


Figure 16 Synchronised behavioural data for baseline and cognitive task, curve driving in the fixed-base simulator

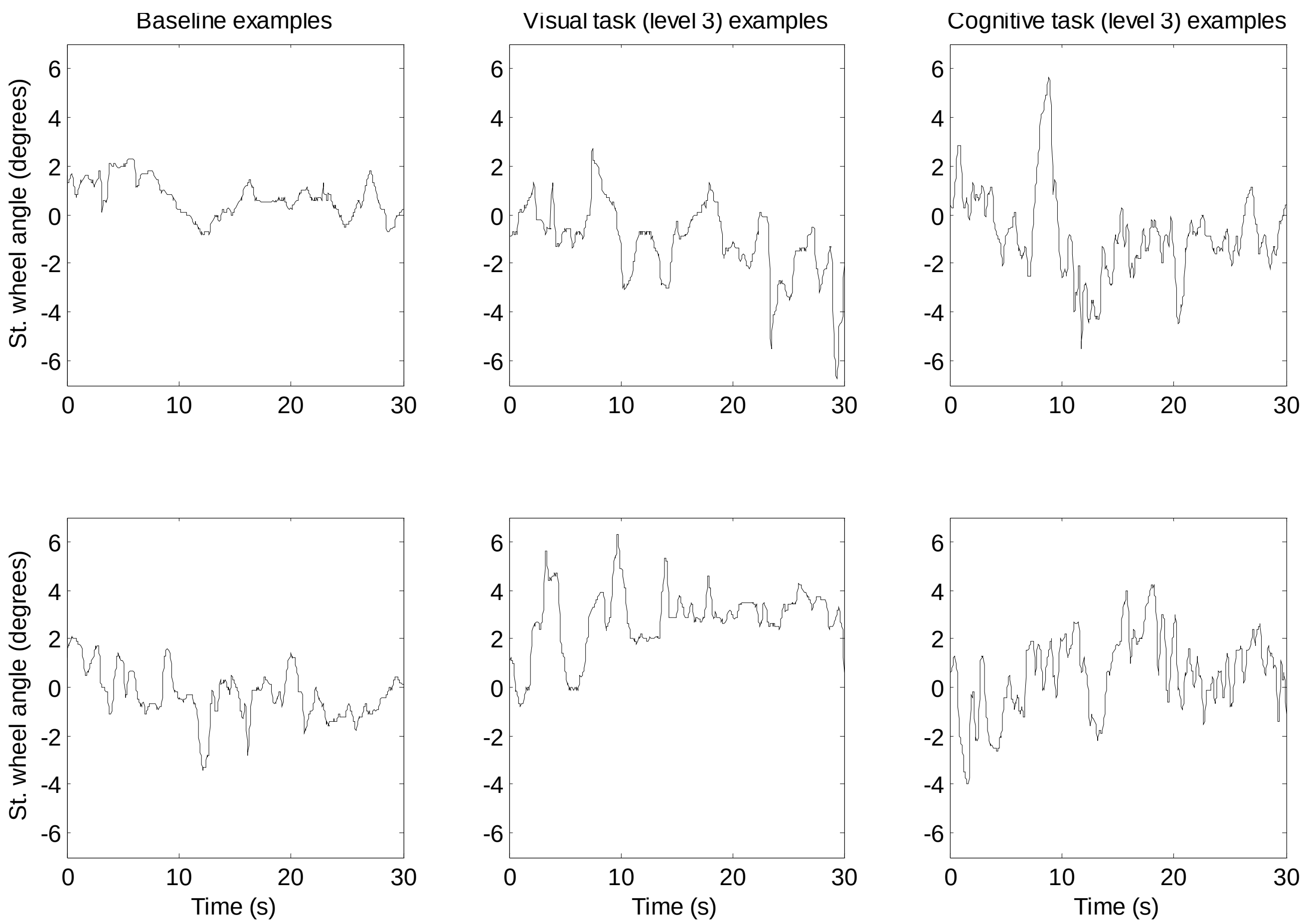


Figure 17 Steering wheel movement time series from the field data – baseline, visual and cognitive tasks.

Based on this qualitative analysis, some general conclusions may be drawn:

- Visual and cognitive load induce markedly different behavioural patterns. This is in line with previous research (Engström et al., 2005; Östlund et al., 2004), but the present analysis shed further light on the detailed nature of these patterns.
- Visual load induces abrupt steering wheel corrections, most of which are in the magnitude range of 2-6 degrees.
- Cognitive load induces an increased amount of micro steering corrections most of which are smaller than 2 degrees.
- Curves have a strong effect on the steering wheel data.
- For visual load, the magnitude of the steering corrections increases somewhat in curves. Field data look generally similar as the simulator data, but seem to contain more variance from sources other than the secondary task.

3.8.4.2 *Analysis of frequency content*

The steering data was further analysed in terms of frequency content. For this purpose, frequency spectra for the steering wheel angle data were created, using the Fast Fourier Transform (FFT).

30-second steering wheel data sequences for which a common frequency spectrum was to be calculated (e.g. all sequences for one driver in a specific condition, or all sequences for all drivers in a specific condition) were concatenated together into a single, joint data sequence θ_i, $i \in \{1, 2, \ldots, N\}$. To avoid discontinuities in the joints between the sequences, each sequence was weighted in the time domain by a Hanning window before the concatenation. The FFT, $a_i = \mathrm{FFT}(\theta_i)$, of the entire concatenated sequence was then calculated. The frequency content at frequencies $\{0, g, 2g, \ldots, (N/2)g, (N/2+1)g\}$, $g = f_s/2(N/2 + 1)$, where f_s is the sampling frequency, was taken as the first $N/2+1$ elements of the series

$$P = \sqrt{\frac{a_i \bar{a}_i}{N}}\ .$$

A Gaussian averaging filter was applied to P before plotting the spectrum to remove noise. The standard deviation of the used Gaussian was 0.01 Hz for spectra involving more than one driver, and 0.05 Hz for spectra only involving one driver (less data, yielding more noise).

Figure 18 shows a comparison of baseline steering data for joint sequences for all subjects, for all conditions. Several important observations can be made from this graph.

- Curves strongly increase the frequency content across the entire spectrum. Thus, effects of road curvature on steering wheel angle cannot simply be removed by high-pass filtering the signal.
- There are marked differences between the three test settings (fixed-base simulator, moving-base simulators and field), especially in the lower frequency region. This difference may be attributed to:
  - an identified, but to be specified, influece in the steering system of the fixed based simulator (causing steering error requiring corrections)
  - lack of motion feedback in the fixed base simulator, resulting in larger steering error during visual distraction than in moving base simulator and field
  - other differences in steering dynamics, such as steering ratio
- The two data sets from straight driving (motorway and rural road) in the fixed-base simulator are very similar in terms of frequency content. Thus, factors such as lane width or average speed did not seem to have a major influence on the steering signal in this data.

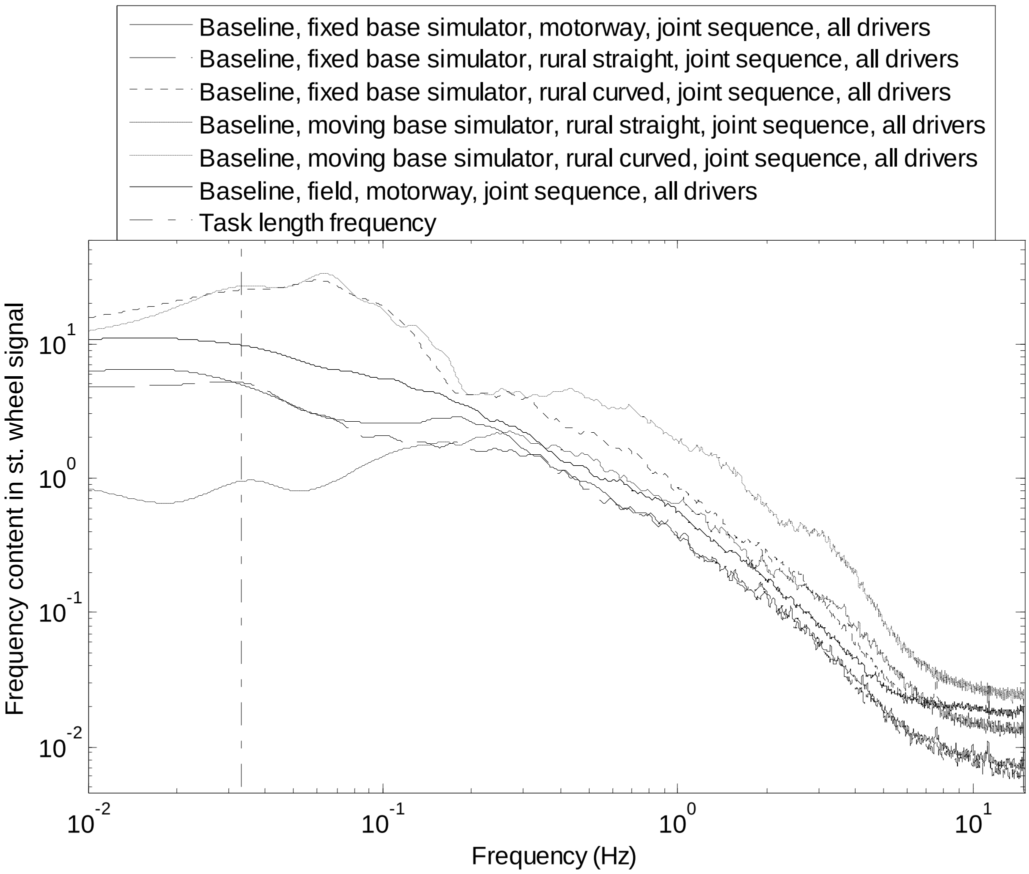


Figure 18 Baseline joint spectrum for all conditions

In order to investigate differences between subjects, individual spectral plots were computed for each driver. Spectral frequency plots for three different drivers in baseline and visual task conditions are shown in *Figure 19*, which indicate large individual differences. For driver 1, the visual task led to a general increase in power throughout the spectrum. For driver 8, there was a marked increase in the 0.1-1 Hz region, while, for driver 5, the secondary task resulted in a general reduction in frequency content while. Most drivers exhibited an increase in frequency content in the secondary task condition (driver 5 was thus an exception), although the increase generally occurred in different frequency regions. These differences are probably related to the adoption of different steering strategies, both during normal driving, and when coping with the visual task. Similar individual differences could be observed for the cognitive task.

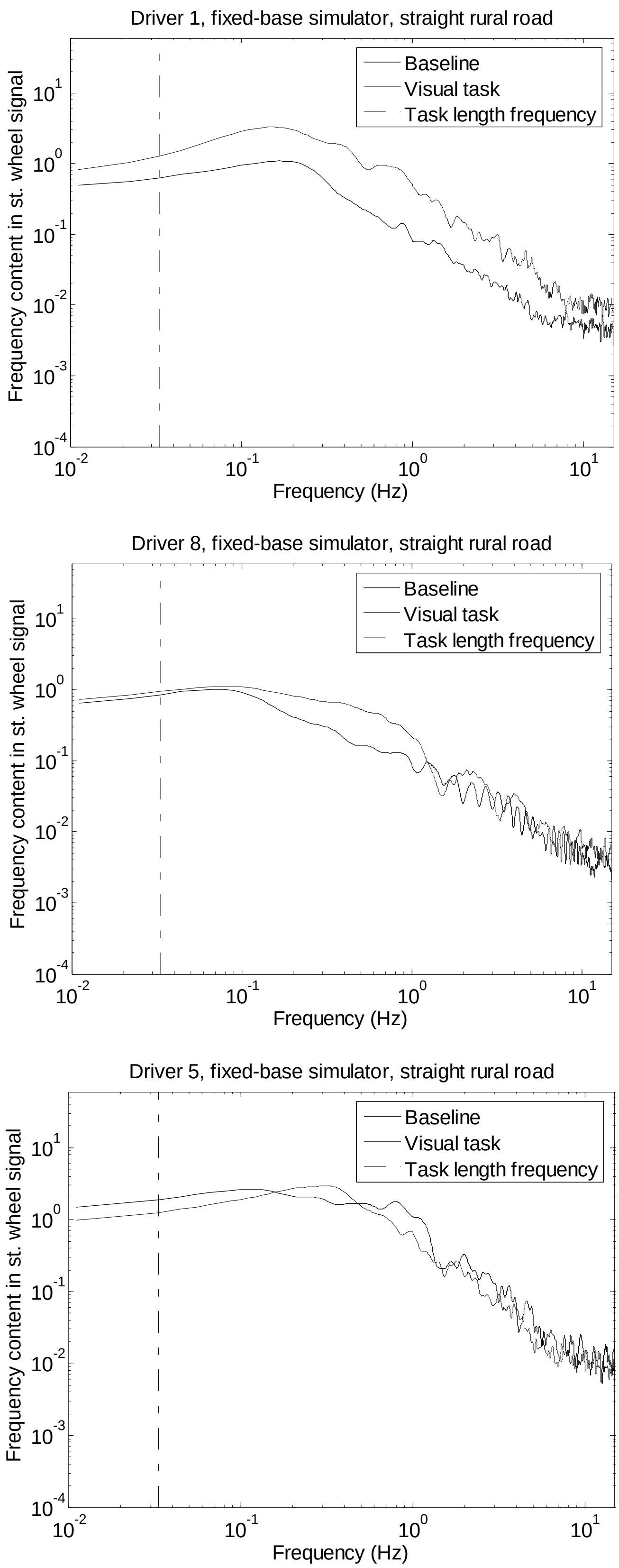


Figure 19 Spectral frequency plots for three different drivers in the fixed-base simulator, straight rural road condition

Figure 20 to *Figure 22* show spectral frequency plots for all drivers' data joined together. The three plots represent straight driving data from the three experimental settings (the two simulators and the field) for baseline, the visual task and the cognitive task. In general, the spectral density increased for both types of tasks in all data sets. However, the magnitude and nature of the effects differed between the visual and cognitive tasks, where the cognitive task mainly led to an increase in higher frequencies while the visual task increased frequency content across the spectrum.

The patterns were also quite different between the test settings. In the fixed base simulator there was a marked effect of the cognitive task in the region above 0.3 Hz, while the visual task boosted all frequencies above 0.1 Hz. In the moving-base simulator, the visual task appeared to influence the entire frequency spectrum, while the cognitive task only had a small effect in the region of 1-5 Hz. Finally, in the field data the effects were much smaller, and mainly occurred for the visual task, in the region >0.2 Hz. Taken together, these data show that the type of simulator/test vehicle affects not only the baseline SW data, but also the effects of secondary tasks on steering. In other words, the effects of test setting and secondary task demand do not simply add up but seem to interact in a non-trivial way. A general conclusion from the present analysis is that it is difficult to identify a specific frequency range that is particularly sensitive to secondary task load.

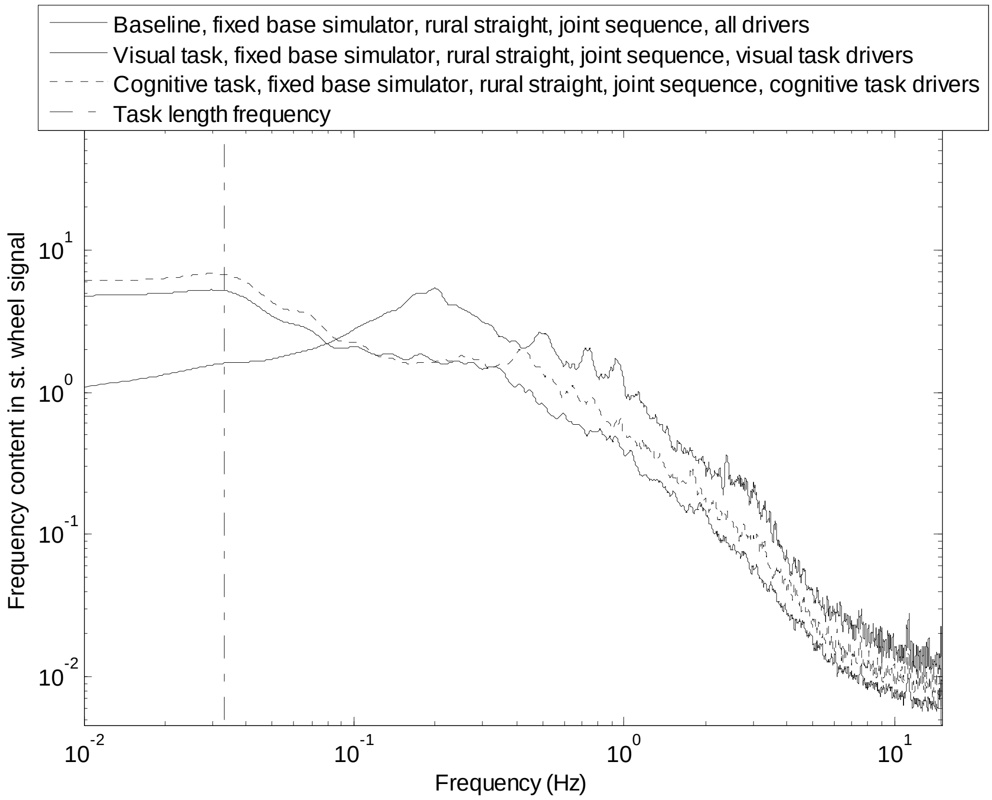


Figure 20 Joint frequency spectrum for the fixed-base simulator, rural road, straight driving

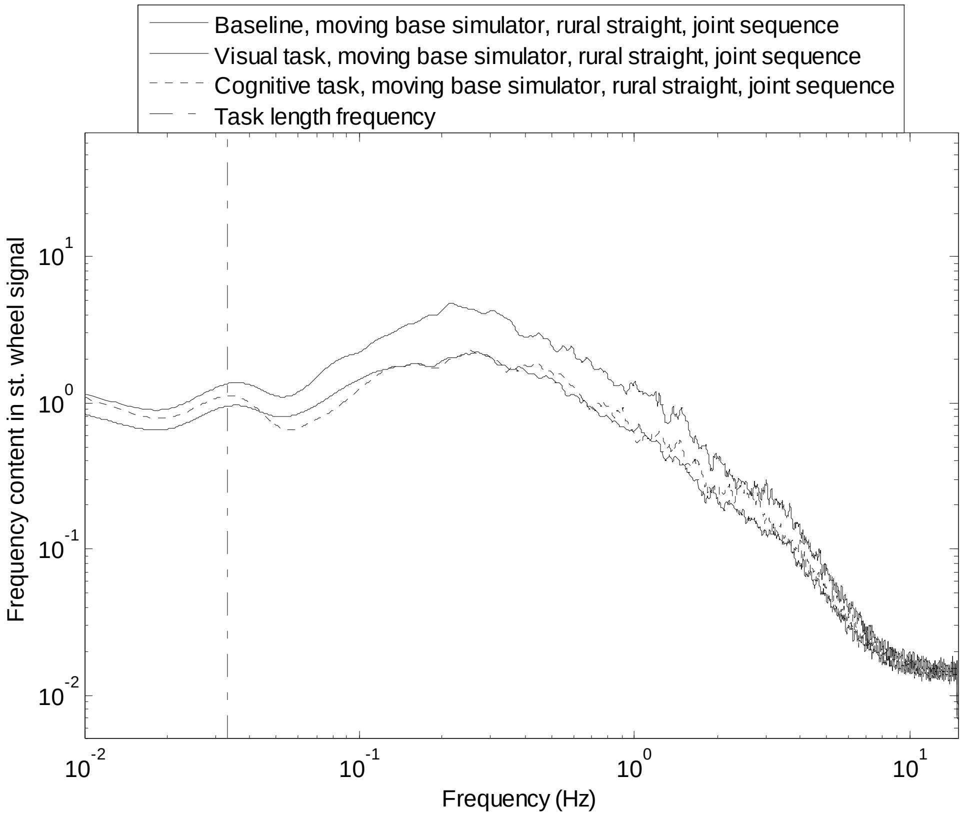


Figure 21 Joint frequency spectrum for the moving-base simulator, rural road, straight driving

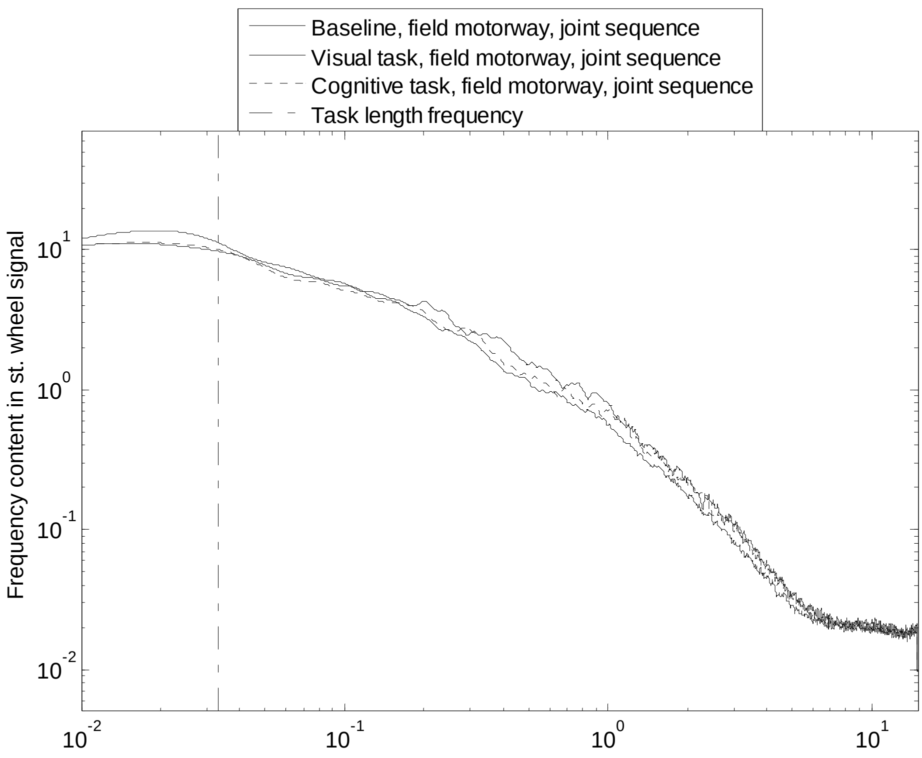


Figure 22 Joint frequency spectrum, field data

3.8.5 Candidate steering wheel metrics

Based on the literature reviewed in AIDE D2.1.2 (Johansson et al., 2005), and on the qualitative analysis reported in the previous sections, a number of candidate steering wheel metrics were selected:

3.8.5.1 *Standard deviation of steering wheel angle*

This is by far the most common steering wheel metric used for IVIS evaluation purposes (e.g. Liu et al., (1999). It is defined as the standard deviation σ of the steering wheel angle signal θ_n:

$$\sigma = \sqrt{\frac{1}{N}\sum_n \left(\theta_n - \overline{\theta}\right)^2} \ ,$$

where n enumerates the time steps, N the number of data points. This metric mainly represents the magnitude of steering wheel movements. Given the analysis in the previous section (as well as existing results in the literature, e.g. Boer, 2005), it could be expected to be too crude to capture the effects of cognitive load. Thus, it is included here mainly as a benchmark for the more sophisticated metrics.

3.8.5.2 *Reversal rate*

This metric, originally proposed by McLean and Hoffmann (1971), measures the frequency of steering wheel reversals larger than a finite angle, or *gap*. The magnitude of this gap, the *gap size*, is thus a key parameter for this metric. Gap sizes reported in the literature vary between 0.5-10 degrees. In the HASTE project, gap sizes of 1, 3, 5 and 7 degrees were tested (Engström, Johansson and Östlund, 2005). The results indicated that different gap sizes tended to yield different sensitivities to visual and cognitive loads, although this was not investigated further in HASTE[3]. In the present work, two implementations of reversal rate were tested. These are described below.

Reversal rate 1

The first reversal rate implementation was exactly the one used in the HASTE project. It proceeds as follows:

1. First, the steering wheel angle signal is low pass filtered with a second order Butterworth filter. In the HASTE project, a cut-off frequency of 0.6 Hz was used.
2. Then, local extreme points (minima and maxima) are identified within a sliding time window of size 0.8 s. The center point in a time window is a maximum (or minimum) if the steering wheel angle data in the window before the center point is monotonically increasing (decreasing) towards the center point value, and the data after the center point is monotonically decreasing (increasing).
3. A reversal is then counted each time the difference in steering wheel angle signal between a maximum and the next minimum exceeds the gap size, and each time the difference in steering wheel angle signal between a minimum and the next maximum exceeds the gap size.

[3] Rather, in HASTE, it was decided to recommend a gap size of 1 degree since this appeared to yield sensitivity for both visual and cognitive loads.

In the present analysis, the following free parameters of this metric have been varied:

- Low-pass filter cut-off frequency {0.6 Hz; 2 Hz; 5 Hz; 10Hz}
- Gap size {0.1°; 0.5°; 1°; 2°; 3°; 4°; 5°; 10°}

Reversal rate 2

In the present work, a number of potential disadvantages of the HASTE reversal rate implementation have been identified, mainly related to this implementation's time windowed definition of extreme points, and the definition of reversals as only occurring between adjacent extreme points:

- The definition of local extreme points, using the 0.8 second time window, is somewhat arbitrary. Its purpose is to avoid considering minor fluctuations in the steering wheel angle signal. By decreasing the time window's size, extreme points that are more "locally" extreme can be included. This, however, means that a greater number of extreme points will be identified, which decreases the probability of the steering wheel angle difference between two adjacent extreme points exceeding the gap size, thus resulting in a lower observed reversal rate. The combination 0.6 Hz low pass filtering and 0.8 second time window has been found to ignore small fluctuations to a reasonable extent, thus producing sensible looking reversal identification, but it is difficult to analyse what this "operational" definition of a steering wheel reversal really means.
- Further, to capture steering wheel reversals smaller than about 1°, the 0.6 Hz low pass filtering cut-off frequency may be too small, since such low pass filtering removes much of the smaller (and more high frequent) steering angle corrections. An obvious solution to this problem is to increase the cut-off value. However, to maintain successful extreme point identification on steering wheel angle sequences that have been low pass filtered with higher cut-offs, the 0.8 second time window has to be decreased in size, since the requirement of monotonic variation will be less likely to be met in this case. In practice, this means that for each new low pass filtering cut-off value, a suitable extreme point detection time window size has to be identified, which is an inconvenient process.
- Finally, consider the case of a number of extreme points in sequence, such that the difference in steering wheel angle between the first and last extreme point is larger than the gap size, but not the differences between first and second, and second and third, et cetera. See Figure 23 for an illustration. It can be argued that, in this situation, one reversal has occurred between the first and last extreme point. However, the HASTE reversal rate implementation does not make the same judgement, since reversals can only occur between adjacent extreme points.

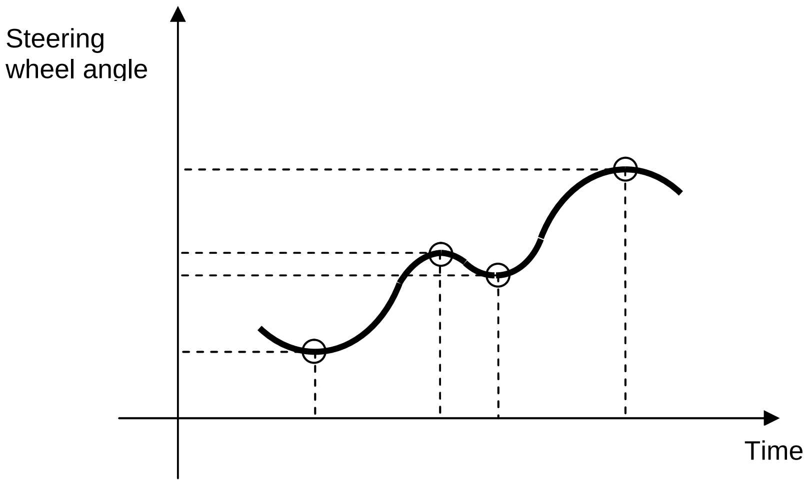


Figure 23 Example of one long reversal interrupted by a max-min pair in the middle.

In the present work, an alternative definition and implementation of steering wheel reversal rate was therefore made. A detailed description of this implementation is given in section 7.18, but the general steps of the procedure are given below:

1. "Stationary points" in the steering wheel angle data are identified:
    a. The first, optional, step is to resample the signal to a quicker sample rate (the original sample rate for the data used here was 30 Hz). This obviously cannot introduce any information into the signal that wasn't there already, but it yields better stationary point identification (step 1.c. below) when there are a lot of quick steering wheel direction changes.
    b. Then, the steering wheel angle signal is filtered with a Butterworth low pass filter. This reduces any high frequency noise in the signal, and is also needed to make the stationary point identification step (1.c.) possible.
    c. The "stationary points" in the steering wheel angle signal can now be identified. By "stationary point", we generally mean a point where the derivative of the signal is zero. Approximate such points are found by looking for zero-crossings in the steering wheel angle difference between consecutive points in the low pass filtered signal.
2. "Upwards" reversals (i.e. typically from a minimum to a maximum) are found by starting at the first stationary point and looking for another stationary point which is more than a gap size larger in steering wheel angle than the first point. If, before finding such a point, a stationary point lower in steering wheel angle than the first point is found, this point is taken as the new "first" point and the search is continued.
3. Finally, "Downwards" reversals are identified using the same algorithm as in step 3, but applied to the negative values of the low pass filtered steering wheel angle signal.

In the present analysis, the following free parameters of this metric were varied:

- Resampling frequency {30 Hz (no resampling); 100 Hz}
- Low-pass filter cut-off frequency {0.6 Hz; 2 Hz; 5 Hz; 10Hz}
- Gap size {0.1°; 0.5°; 1°; 2°; 3°; 4°; 5°; 10°}
- Butterworth filter order {2; 5}

3.8.5.3 *High frequency steering*

This metric represents the frequency content in a specified band of the steering wheel signal. The key parameters are thus the upper and lower frequency boundaries. McLean and Hoffman, 1975 found that the frequency content in the 0.35-0.6 Hz band is sensitive to variations in both primary and secondary task load, and these or similar values have been used in several subsequent studies, e.g. in HASTE (e.g. Östlund et al., 2004). The frequency content in the frequency band can be computed in different ways. In HASTE two main approaches were tested: (1) applying a band pass filter and then calculating root-mean-square (RMS) power of the remaining signal and (2) performing a Fourier transform and integrating under the selected frequency band. These two approaches were also tested in the present analysis, along with a third version, which represented an alternative implementation of the Fourier transform approach.

The three metrics were implemented:

HF steering 1

Here, a band pass filter implemented as two Butterworth filters, one low pass and one high pass, is applied. Then the high frequency steering value is calculated as the RMS power of the filtered signal x_n, i.e.

$$HF = \sqrt{\frac{1}{N}\sum_n x_n^2}$$

In the present analysis, the following free parameters of this metric were varied:

- High-pass filter cut-off frequency {no high-pass filter; 0.3 Hz; 0.5 Hz; 1 Hz; 2 Hz; 3 Hz; 4 Hz; 5 Hz; 6 Hz; 7 Hz; 8 Hz}
- Low-pass filter cut-off frequency {0.6 Hz; 1 Hz; 2 Hz; 3 Hz; 4 Hz; 5 Hz; 6 Hz; 7 Hz; 8 Hz; 9 Hz; 10 Hz; no low-pass filter}
- Butterworth filter order {2; 5}

HF steering 2

In this implementation of the high frequency steering metric, the MATLAB 6.5 command *spectrum* (using a 256 point FFT, a Hanning window, and 64 point overlap) is used to obtain a steering wheel frequency spectrum. Then a specific frequency band is extracted by (numerically) integrating under the curve from a high-pass filter cut-off frequency to a low-pass filter cut off frequency.

In the HASTE project the logarithm of the integration result was taken as the final metric value. This solution yielded good results, but no good argument for it was given. Here, we use one of the following methods of metric scaling:

- Taking the logarithm of the integration result, as in the HASTE project.
- Taking the square root of the integration result.
- Taking the logarithm of all spectrum values, before integration.
- Taking the square root of all spectrum values, before integration. Of the four methods, this one should theoretically yield the metric that is closest to the HF steering 1 metric.

The following parameters were varied:

- High-pass filter cut-off frequency {no high-pass filter; 0.3 Hz; 0.5 Hz; 1 Hz; 2 Hz; 3 Hz; 4 Hz; 5 Hz; 6 Hz; 7 Hz; 8 Hz}
- Low-pass filter cut-off frequency {0.6 Hz; 1 Hz; 2 Hz; 3 Hz; 4 Hz; 5 Hz; 6 Hz; 7 Hz; 8 Hz; 9 Hz; 10 Hz; no low-pass filter}
- Metric scaling method (see above)

HF steering 3
This implementation is identical to HF steering 2, with the single difference that the method used for obtaining the frequency spectrum is the one described in section **Fel! Hittar inte referenskälla.**. One slight variation of that method was also evaluated, using a rectangular window instead of the Hanning window. The following parameters were varied:

- High-pass filter cut-off frequency {no high-pass filter; 0.3 Hz; 0.5 Hz; 1 Hz; 2 Hz; 3 Hz; 4 Hz; 5 Hz; 6 Hz; 7 Hz; 8 Hz}
- Low-pass filter cut-off frequency {0.6 Hz; 1 Hz; 2 Hz; 3 Hz; 4 Hz; 5 Hz; 6 Hz; 7 Hz; 8 Hz; 9 Hz; 10 Hz; no low-pass filter}
- FFT window {Rectangular; Hann}

3.8.5.4 *Steering entropy*

This metric was first described in Nakayama et al. (1999), and further developed in Boer (2000). Recently, a slightly updated version has been proposed (Boer et al., 2005). The metric can be viewed as representing the predictability of steering wheel movements in terms of the entropy (disorder) of the errors of SW angle predictions generated by a linear filter. A key difference from the other metrics investigated here is that steering entropy calculation takes into account baseline data for the individual subject, thus involving a kind of normalisation with the purpose to reduce between-subject variance.

In the present study, two versions of steering entropy were tested:

Steering entropy 1
This is the implementation used in the HASTE project, based on Nakayama et al. (1999):

1. First, an initialization for the specific driver is made using one of the baseline sequences for the driver:
   a. The steering wheel angle signal of the baseline sequence is resampled down to a lower rate.
   b. Then, a linear predictive filter $x_n' = 5x_{n-1}/2 - 2x_{n-2} + x_{n-3}/2$ is applied, and the prediction errors $e_n = x_n' - x_n$ are calculated.
   c. The 90th percentile of the prediction errors, α, is calculated.
2. Then, any sequence (for baseline or task driving) for this driver can be analysed:
   a. Resampling to a lower rate, as above.
   b. Application of the linear predictive filter and calculation of the prediction errors, as above.
   c. Division of the prediction errors into bins with edges {$-\infty$; -5α; -2.5α; -0.5α; 0; 0.5α; 2.5α; 5α; ∞}, resulting in a distribution of bin "probabilities" p_i.
   d. Calculation of the entropy as $h = \Sigma p_i \log p_i$, where log denotes the 2-logarithm.

For this metric, the following parameter was varied:

- Re-sampling frequency {5 Hz; 6 Hz; 10 Hz; 15 Hz; 30 Hz}

Steering entropy 2
In addition to the HASTE implementation, an additional implementation was made based on the more recent version suggested in Boer et al (2005), with the main improvement being a more elaborate initialization phase.

1. First, an initialization for the specific driver is made using one of the baseline sequences for the driver:
    a. The steering wheel angle signal of the baseline sequence is first low pass filtered using a fifth order Butterworth low pass filter with a cut-off frequency of 3/7 of the sample rate.
    b. The filtered signal is resampled down to a lower rate.
    c. Then, a linear predictive filter is adapted to the baseline sequence, either using the Burg method for adaptation of an autoregressive model to data, or using a least squares model adaptation.
    d. This predictive filter is applied to the baseline sequence, and the prediction errors e_n are calculated.
    e. The quantity pe_α is calculated. This is a variant of the percentile, taking into account the distribution of prediction errors on both sides of the origin. Here, α is a free parameter of the algorithm, and using a given α corresponds roughly to taking the 100(1 - 2α) percentile.
    f. Division of the prediction errors into bins with edges {$-\infty$; $-6pe_\alpha$; $-5pe_\alpha$; $-4pe_\alpha$; $-3pe_\alpha$; $-2pe_\alpha$; $-1pe_\alpha$; 0; $1pe_\alpha$; $2pe_\alpha$; $3pe_\alpha$; $4pe_\alpha$; $5pe_\alpha$; $6pe_\alpha$; ∞}, resulting in a distribution of baseline bin “probabilities” $(p_i)_{BL}$.
2. Then, any sequence (for baseline or task driving) for this driver can be analysed, by:
    a. Low pass filtering, as in above.
    b. Resampling to a lower rate, as above.
    c. Application of the linear predictive filter (that was adapted to this specific driver) and calculation of the prediction errors, as above.
    d. Division of the prediction errors into bins, as above, yielding a distribution of bin “probabilities” p_i for this sequence.
    e. Calculation of the entropy as $h = \Sigma p_i \log (p_i)_{BL}$, where log denotes the 2-logarithm.

The free parameters varied here were:
- Re-sampling frequency {5 Hz; 6 Hz; 10 Hz; 15 Hz; 30 Hz}
- α {0.05; 0.1; 0.2}
- Predictive model adaptation method {Burg method; least squares method}

3.8.6 Sensitivity analysis

The next step in the analysis was to investigate the sensitivity of the selected metrics to visual and cognitive load. This followed the same approach as in HASTE, where sensitivity was quantified in terms of standardised effect sizes (Cohen’s d’), i.e. the standardised difference between baseline and the experimental conditions (secondary tasks). The standardised effect size *d* was computed as:

$$d = \frac{\mu_{task} - \mu_{baseline}}{\sigma_{pooled}}$$

where

$$\sigma_{pooled} = \sqrt{\frac{(n_{task} - 1)\sigma_{task}^2 + (n_{baseline} - 1)\sigma_{baseline}^2}{n_{task} + n_{baseline} - 2}}$$

In these formulas, μ is the mean, σ is the standard deviation and n the the number of data points.

The free parameters for each metric were varied systematically in order to find the best combination for each metric (except for standard deviation, which was calculated directly on the raw signal and thus had no free parameters). *Figure 24* shows the maximum effect sizes obtained for the candidate metrics for the visual task in the six investigated conditions ((1) fixed simulator-motorway, (2) fixed simulator rural road straight, (3) fixed simulator rural road curve, (4) moving base simulator rural straight, (5) moving base simulator rural road curve and (6) field). *Figure 25* shows the corresponding data for the cognitive task.

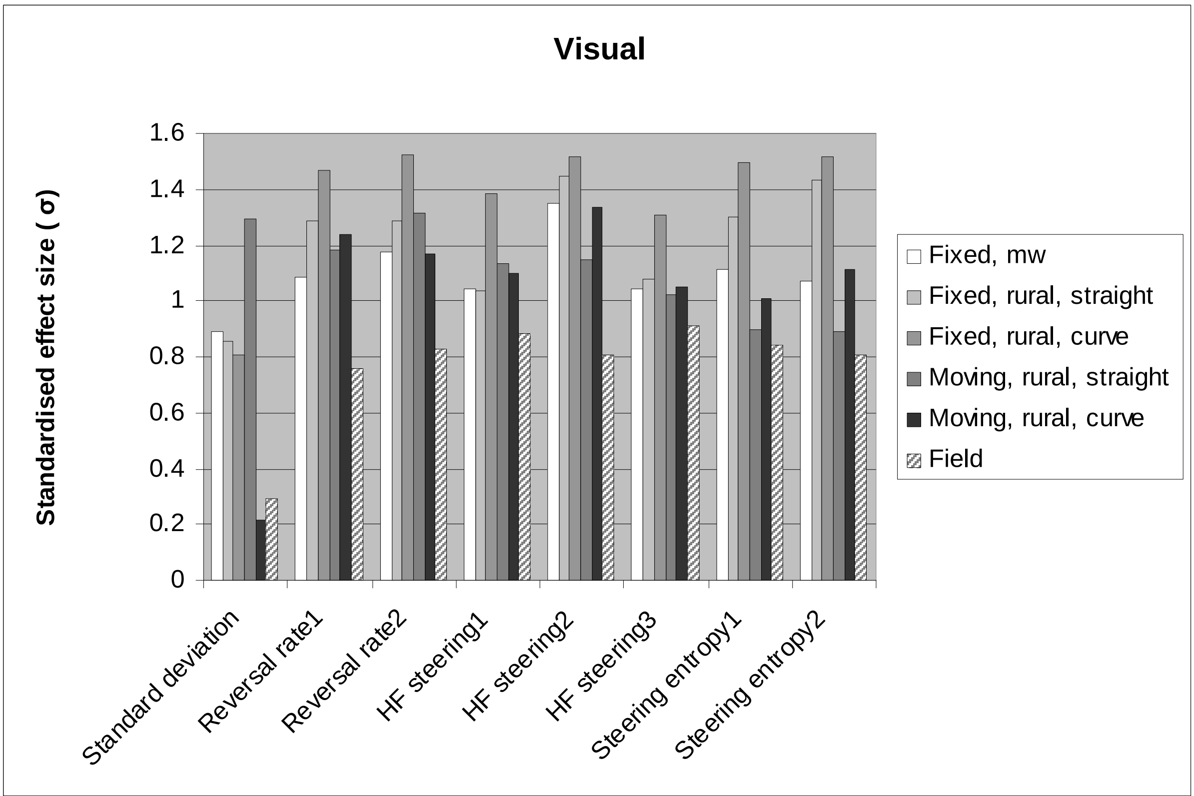


Figure 24 Maximum standardised effect sizes (i.e. those obtained for the optimal parameter setting) for the different steering wheel metrics for the visual task

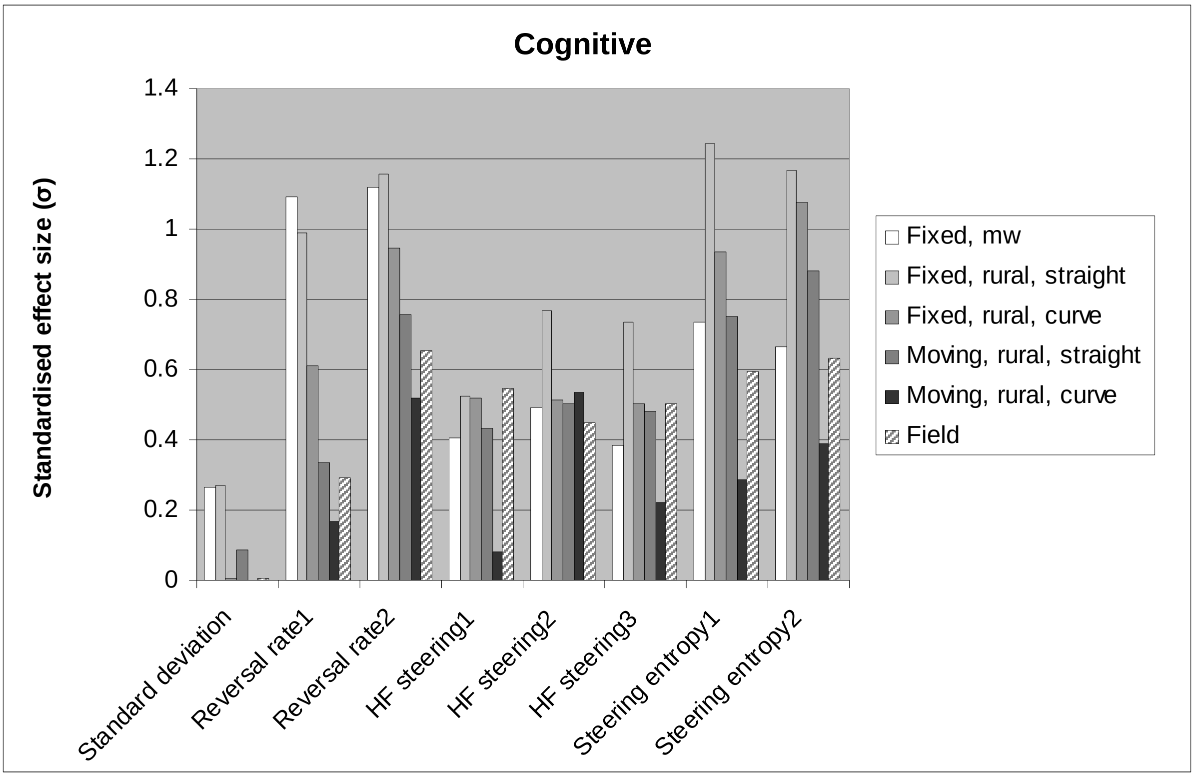


Figure 25 Maximum standardised effect sizes (i.e. those obtained for the optimal parameter setting) for the different steering wheel metrics for the cognitive task

From *Figure 24* it can be observed that, for the visual task, the effects were generally quite large for most metrics. However, as expected, the effects for the standard deviation metric were generally smaller, especially for the moving-base-curve and field conditions. This was probably due the fact that the standard deviation metric mainly represents steering magnitude and thus did not discriminate between the variation induced by the secondary task and that induced by other factors such as curvature and traffic events. It can also be observed that the effects were generally smaller in the field than in the simulators. The largest effects were obtained in the fixed-curve conditions, except for the standard deviation metric. The main conclusion from these results was that all metrics investigated (except standard deviation) seem to be sensitive to visual load in all conditions, and the differences in sensitivity between these metrics were quite small.

For cognitive load, the picture was somewhat different. From *Figure 25*, it can be observed that the standard deviation metric did hardly capture any effects of cognitive load at all. Moreover, the reversal rate and steering entropy metrics yield markedly stronger effects than the high frequency steering metrics. The lower sensitivity of the frequency metrics was quite expected, given the frequency content analysis in section **Fel! Hittar inte referenskälla.**, which showed that cognitive load did not seem to have a consistent effect on frequency content. It can also be observed that ReversalRate2 did improve performance compared to Reversal Rate 1, especially in the moving-base and field conditions. The two steering entropy metrics yielded fairly similar results, although SteeringEntropy2 (based on Boer et al., 2005) appeared to be slightly more sensitive in most conditions. Finally, by contrast to the visual task (where the strongest effects were found in curves), the effects of the cognitive task were strongest in the straight-driving conditions.

All results taken together, the ReversalRate2 metric was the most sensitive across all tasks and test settings, followed by the two Steering Entropy variants. Since the ReversalRate2 is also somewhat more straightforward to implement and to interpret, it was selected as the main candidate for further consideration. The basis for this selection is further discussed in section 3.8.8. The next section investigates effect of different parameter settings for ReversalRate2.

3.8.7 Parameter tuning for Reversal Rate2

In the final step, the effect of different parameter settings for the ReversalRate2 metric was investigated. The re-sampling frequency and order of the low-pass filter did not seem to have any great impact and, thus, the analysis focused on the effects of the other two free parameters: gap size and low pass filter cut-off frequency. Effect sizes were calculated for the combinations of the following values:

Gap-size (degrees): {0.1, 0.5, 1, 2, 3, 4, 5, 10}
LPF cut-off frequency (Hz): {0.6, 2, 5, 10}

The analysis was done by means of visual inspection of 3D-surface plots where effect size was plotted as a function of gap size and low-pass filter cut-off frequency (see Figure 26 to *Figure 28*). The following were the main findings from this analysis:

- For the visual task in straight driving conditions, the optimal gap size could be found in the range of 2-4 degrees in all straight-driving conditions and the cut-off frequency did not have a major influence. A representative example, from the field data, is given in *Figure 26*.
- For the visual task in curves, the low-pass frequency cut-off had a strong influence, where the highest sensitivity was achieved for the lowest cut-off value tested (0.6 Hz). Moreover, the optimum for the gap-size is increased to 5 degrees or even higher, depending on the LPF cut-off setting. An example is given in *Figure 27*, with data from the moving-base simulator.
- For the cognitive task, the optimal gap size is much smaller than for the visual task. In fact, the sensitivity is largest for the smallest the gap-sizes investigated (0.1 and 0.5 degrees). It should be noted that smaller gap sizes could not be detected since the spatial resolution of the steering wheel angle sensor was limited to 0.1 degrees. The cut-off frequency parameter had some influence, mainly in the moving base and field conditions, where the effect was reduced for the lowest value (0.6 Hz). This is expected since much of the variance related to the cognitive secondary task is in the higher frequencies (as shown in the frequency content analysis above). An example, from the field data, is given in *Figure 28*.

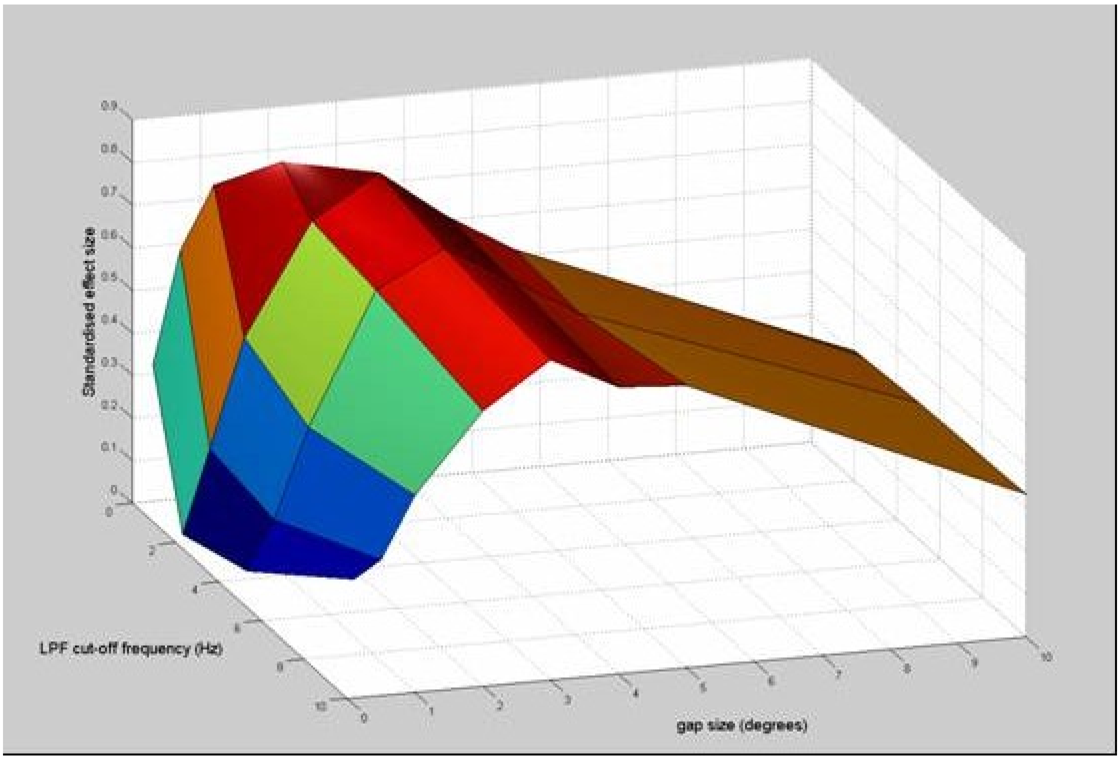


Figure 26 Effect size as a function of gap size and LPF cut-off frequency, visual task, field

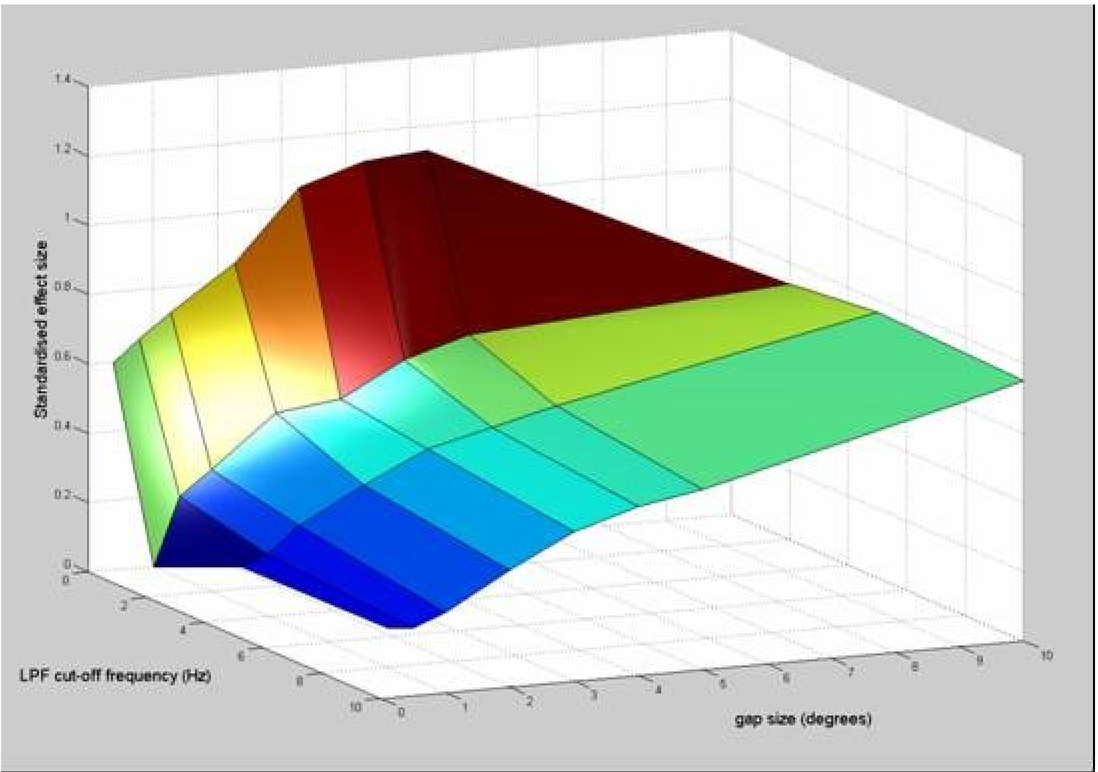


Figure 27 Effect size as a function of gap size and LPF cut-off frequency,visual task, moving-base simulator, curve

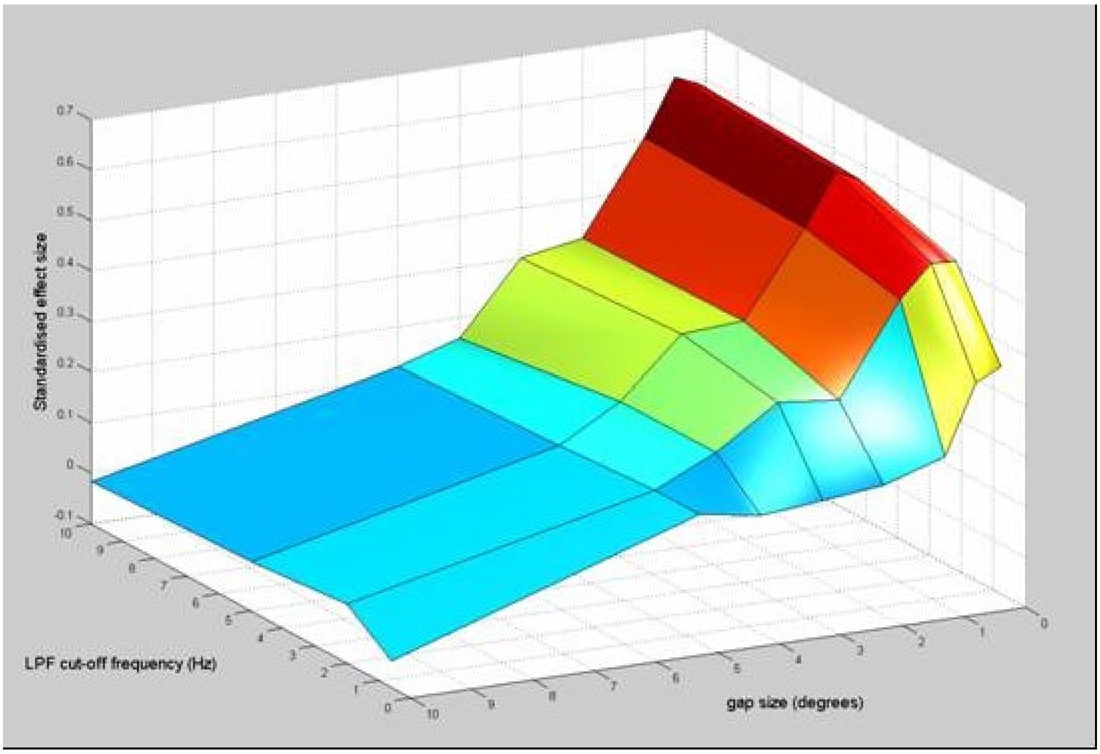


Figure 28 Effect size as a function of gap size and LPF cut-off frequency, cognitive task, field

Figure 29 plots effect size as a function of gap size for both the visual and the cognitive tasks in all conditions. The cut-off frequency was held constant at 2Hz. The plot clearly shows the differences between the two task types, where the largest sensitivity for cognitive load was obtained for the smallest gap sizes. By contrast, for the visual task, the optimum gap size varied between 2 and 10 degrees, with the largest optimal gap sizes in curves. The optima were also generally larger in the moving base simulator than in the fixed-bas simulator and the field. These consistent differences in optimal gap sizes clearly indicate the very different effects of visual and cognitive load on steering.

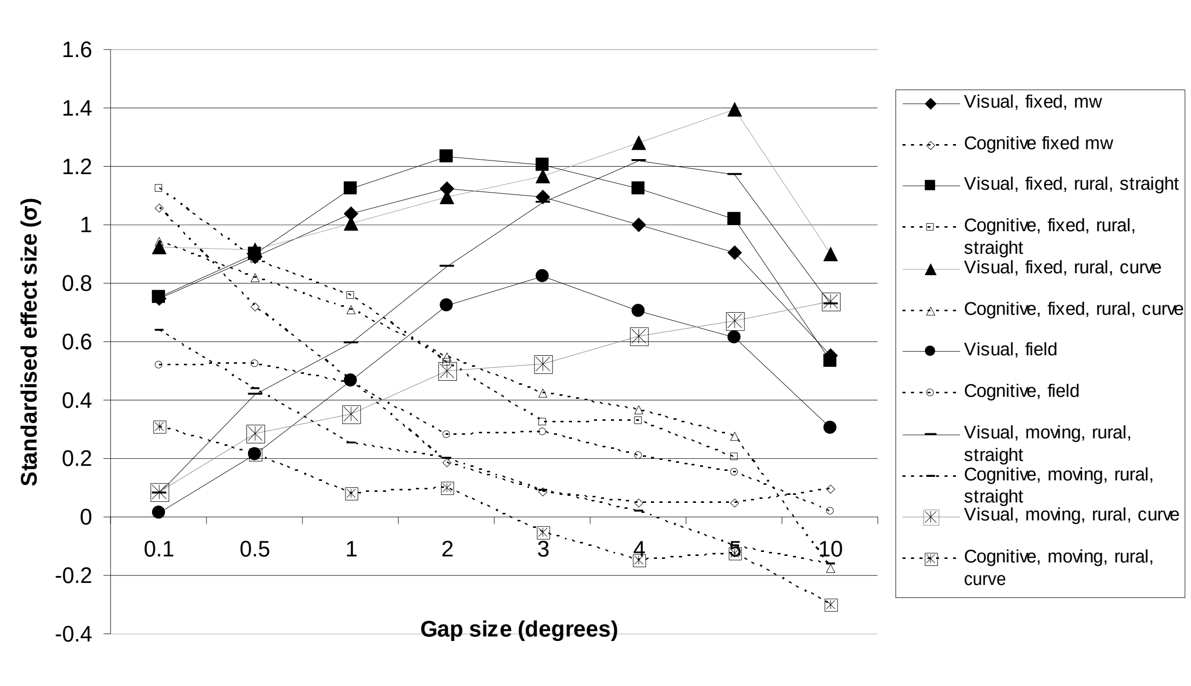


Figure 29 Effect size as a function of gap size for visual and cognitive secondary tasks. The cut-off frequency was held constant at 2Hz.

3.8.8 Discussion and conclusions: Steering wheel metrics

3.8.8.1 *Review of main findings*

The results from the HASTE WP2 studies clearly indicated qualitatively different effects of visual and cognitive secondary task load on driving performance (Östlund et al., 2004). A particularly strong result from the HASTE study was that visual load resulted in increased lane keeping variance while cognitive load had the opposite effect. However, both visual and cognitive load resulted in increased steering wheel activity, where the effects found in HASTE were weaker for cognitive load, and significant effects were only found in some of the sub-studies. In these cases, however, there were indications that the two task types resulted in different types of steering patterns where visual load tended to mainly induce large steering wheel corrections while cognitive load resulted in an increased amount of steering micro-corrections.

The results from the present study confirm these findings and provide further insight into the detailed effects of the HASTE S-IVIS on steering behaviour. It was found that visual load led to steering corrections, mainly in the range of about 2-6 degrees, where the corrections were somewhat larger in curves. By contrast, cognitive load mainly increased steering corrections smaller than 1 degree. Thus, the weak effects of cognitive load on steering wheel metrics in the HASTE study could be explained by non-optimal parameter settings. For example, for the reversal rate metric, a gap size of 1 degree was used for both tasks, which implies that a great portion of the variance related to the cognitive task was missed in the HASTE analysis. The present results showed that, for appropriate parameter settings, the effects of the cognitive task were almost as strong as for the visual task, at least in some test settings (effect sizes larger than 1 were found for several metrics). Thus, as further discussed below, different parameter settings should be considered for visual and cognitive tasks.

The present results also showed, in line with existing work, that steering wheel metrics are very sensitive to a variety of factors, e.g. individual differences, road curvature and test settings (e.g. type of simulator or simulator vs. field). The differences between the different test settings could not be entirely explained based on the present analysis. However, it could be hypothesised that these differences were mainly due to the presence of a moving base in the VTI simulator, to differences in steering dynamics and/or combinations of these factors. In any case, factors relating to the vehicle dynamics must be held under strict control when using steering wheel metrics for IVIS evaluation.

3.8.8.2 *Basic mechanisms*

An issue of critical importance is how the effects of secondary task load on steering wheel movements could be *interpreted*. Based on the present results and the proposed theoretical framework outlined in chapter 2, the effects of *visual* secondary tasks on steering behaviour could be given a quite straightforward explanation. Visual time sharing induced by tasks on the monitoring or targeting layers, or non-driving related tasks (as in the present study), interferes with the basic tracking control by intermittently reducing the visual tracking input. During glances away from the road, tracking errors build up, which are corrected by steering wheel movements (mainly in the range of 2-6 degrees). These steering corrections lead to increased steering wheel angle magnitude as well as frequency, where the frequency content increases across almost the entire spectrum. These effects could be quantified by metrics representing either steering magnitude, frequency content, change rate or a combination of these. Thus, the steering wheel metrics investigated in the present study represent different aspects of the increased steering effort needed to cope with visual time sharing. As such, they could be interpreted as direct measures of the consequences of visual distraction on driving performance. Due to their relative simplicity in terms of measurement, they are useful as a complement to other lateral control performance metrics, such as lane keeping variance.

Cognitive secondary tasks also yield increased steering activity, although the patterns are very different compared to visual task. In this case, the increase is mainly in smaller steering wheel movements, the majority of which are smaller than 1 degree. This often comes with increased gaze concentration towards the road centre and reduced lateral position variance (e.g. Engström et al., 2005; Victor et al., 2005), but no consistent effect on speed (existing studies report null effects as well as speed increase and decrease). While it is clear that these effects are quite different from the effects of visual demand, there is yet no agreed interpretation of the phenomenon. One interpretation is that the increased steering wheel activity (and resulting reduced lane keeping variance) is a side effect of the gaze concentration towards the road centre, which, in turn, is caused by a general cognitive interference that prevents the driver from scanning the environment in the normal way. The gaze concentration provides more visual input for lateral control, thus reinforcing the tracking loop, which results in enhanced steering control input (Engström et al., 2005). Another possible hypothesis is that the increased steering wheel activity reflects an adaptation of safety margins on the regulating layer, where control effort on the tracking layer is increased in order to compensate for the (subjectively perceived) reduction of safety margins (Boer et al., 2005). A problem with the latter explanation is that the easiest way for the driver to increase the safety margins would be to slow down rather than put more effort into steering, but, as mentioned in section 3.2, speed reduction has generally not been observed in existing studies with purely cognitive tasks. Further research is needed in order to establish a clearer interpretation of the effect of cognitive load on steering wheel movements, and driving performance in general. Thus, while being very relevant for basic research on driving performance, steering wheel metrics could not (yet) be recommended for IVIS evaluation with respect to cognitive load.

3.8.8.3 *Sensitivity analysis*

A wide range of candidate steering wheel metrics were investigated in the present study. It was found that the simplest metric, standard deviation of steering wheel angle, yielded the lowest sensitivity for both the types of tasks (for the cognitive task it was not sensitive at all). The main explanation for this is that the SD metric only represents the *magnitude* of steering wheel movements, and this is not where the main effect of secondary task is (especially not for cognitive tasks). The other, more sophisticated, metrics were roughly equally sensitive to visual load, but steering entropy and reversal rate metric where found to be more sensitive than the high-frequency steering metrics for cognitive load. As pointed out by Boer et al. (2005), the main effect of secondary task load is to be found in the *derivative* of the steering signal, rather than the in the magnitude and frequency. This is supported by the frequency content analysis performed in the present study, which showed that the frequency content tended to vary a lot between drivers as well as between test settings. Both steering entropy and reversal rate can be viewed as representing derivative information (i.e. change rate), albeit in different ways. Thus, the best versions of these metrics were selected for further consideration. The ReversalRate2 metric was finally selected due to its more straightforward implementation (no baseline data is needed for normalisation) and interpretation (the gap size parameter is somewhat more intuitive and interpretable than alpha).

3.9 General discussion and conclusions, vehicle control metrics

The objective of the present sub-part of Task 2.2.5 was to identify a limited set of driving control metrics suitable for the AIDE evaluation methodology, mainly based on further analysis of data from the HASTE project.

A key unresolved issue from the HASTE project is the strong dependency of variation metrics (e.g. variance or standard deviation) to data duration. This problem was addressed in the present work by means of high-pass filtering which resulted in modified variation metrics for speed and lateral position. For suitable settings of the filter parameters, it was found that this method reduces the dependency, at least for data sequences longer than 10 seconds.

Moreover, other existing performance metrics, such as TLC-related metrics, headway metrics and brake metrics were reviewed, and the ones most suitable for present purposes identified.

Steering wheel metrics are generally more complex than the “standard” performance metrics reviewed above. Also, since the results from HASTE were somewhat less conclusive for these types of metrics, they were given some more extensive treatment in the present study. In the present work, a number of existing and refined steering wheel metrics were compared with respect to their sensitivity to visual and cognitive load, but also taking into consideration interpretability and ease of implementation. The steering wheel metric finally selected was a modified version of steering wheel reversal rate.

In general the picture emerging from the HASTE WP2 studies (Östlund et al., 2004; Engström et al., 2005; Victor et al., 2005) was further confirmed by the present analysis. It is clear that visual and cognitive loading secondary tasks lead to qualitatively different effects on driving performance. As described in more detail above (section 3.7), visual time sharing leads to reduced lateral tracking control which is corrected for by relatively large compensatory steering manoeuvres. This effect can be quantified e.g. in terms of increased lane position variation, increased number of line crossings and medium- to large steering wheel reversals. However, for visual tasks, the reduced level of lateral control is normally compensated for by a speed reduction (e.g. Antin et al., 1990; Engström et al., 2005). This may also be accompanied by increased headway to a lead vehicle (if any). This increases the total available time for the tracking loop and can be interpreted as an adaptation on the regulation layer to maintain subjectively chosen safety margins at an acceptable level. Due to this compensatory behaviour, the relation between increased steering activity (related to visual time sharing) and actual road safety is not entirely straightforward. The actual safety consequences of the secondary task is related to the driver's ability to fully compensate for the reduced control and it is very difficult to tell, in a specific case, whether a given speed compensation is sufficient. This shows that, in order to obtain a full understanding of the behavioural effects of an in-vehicle system on driving performance, the different metrics must be interpreted together.

It is also possible that visual secondary tasks would have an effect on longitudinal tracking control, which could be quantified e.g. in terms of brake jerks. However, this effect was not investigated in HASTE or in the present study.

By contrast, purely cognitive tasks appear to improve lateral tracking control, an effect which can be measured e.g. in terms of reduced lateral position variation and increased number of small steering reversals. Moreover, it has been shown that these effects generally occur together with increased gaze concentration towards the road centre (e.g. Victor et al., 2004). These seemingly positive effects of cognitive load on tracking control are currently not well understood and further research is needed to identify the key mechanisms in play. There is evidence, however, that cognitive load impairs driving performance on the regulating and monitoring layers, e.g. the ability to adopt appropriate safety margins (Östlund et al., 2004), to detect critical events (Greenberg et al., 2003) and take in driving-relevant information such as speed limit signs (Patten et al., 2003). Some of these effects are reflected in control performance metrics. For example, in the HASTE WP2 studies, it was found that cognitive load leads to significantly shorter headways, indicating impaired safety-margin setting. However, the most straightforward way to capture these effects is probably by means of detection task performance (see e.g. AIDE D2.2.3, Merat et al., in preparation). Also, the Lane Change Task, further addressed in Chapters 4-6, has proved sensitive to purely cognitive load (Mattes, 2003).

Based on the present work, a sub-set of driving control performance metrics could be recommended. Correctly interpreted, these metrics should together yield a comprehensive picture of the effects of various types of IVIS (and to some extent ADAS) on driving performance. The selected metrics are summarised in *Table 5* below. Detailed specifications, together with interpretation guidelines and descriptions of use, are provided in Chapter 7 of this deliverable.

Table 5 The selected subset of driving control metrics, the main behavioural effects that they are intended to quantify, the related causes and the general interpretation in terms of the ECOM/COCOM framework.

Metric	Behavioural effect	Likely secondary task-related cause	Interpretation
Mean speed	Speed reduction	Visual secondary task load	Safety margin compensation on regulation layer to increase available time in the tracking control loop.
	Large speed increase/reduction	Cognitive secondary task load	Reduced monitoring performance.
Maximum speed	Large speed increase	Cognitive secondary task load	Reduced monitoring performance.
Mean lateral position	Changed position in the lane during visual load.	Visual secondary task load	Safety margin compensation to increase time-to-contact to the side which is perceived the more risky (e.g. the one with oncoming traffic).
Modified lateral position variation	Increased variation	Visual secondary task load	Reduced lateral tracking control due to reduced visual input
	Reduced variation	Cognitive secondary task load	Enhanced lateral tracking performance, possible as a side effect of increased gaze concentration.
Line crossings	Increased frequency	Visual secondary task load	Reduced tracking control due to reduced visual input
Steering wheel reversal rate	Increased frequency of medium-large reversals	Visual secondary task load	Reduced lateral tracking control due to reduced visual input
	Increased frequency of small reversals	Cognitive secondary task load	Enhanced tracking control, possible as a side effect of increased gaze concentration.
Mean time headway	Increased headway	Visual secondary task load	Safety margin compensation on regulation layer to increase available time in the tracking control loop.
	Reduced headway	Cognitive secondary task load	Reduced regulating (safety margin setting) performance
Min time headway	Reduced min headway	Cognitive secondary task load	Reduced regulating (safety margin setting) performance
Brake reaction time	Increased BRT	Visual secondary task load	Reduced forward visual attention
	Increased BRT	Cognitive secondary task load	Reduced regulating/monitoring ability (generally reduced situation awareness)
Brake jerks	Increased frequency	Visual secondary task load	Reduced longitudinal tracking control
	Increased frequency	Cognitive secondary task load	Reduced regulating (safety margin setting) control

4 Lane Change Test – Origins of Workload

The Lane Change Test (LCT) (Mattes, 2003) was developed within the German project ADAM (Advanced Driver Attention Metrics; DaimlerChrysler, BMW, 2001-2004). It is a dynamic dual-task method that quantitatively measures human performance degradation on a primary driving-like task while a secondary task is being performed. The LCT is applicable to all types of interactions with in-vehicle information, communication, entertainment, and control systems. In September 2005 the LCT was approved as an active work item of the ISO (ISO/AWI 26022) with the title "Simulated lane change test to assess driver distraction". A brief description of the LCT will be given in the method section of the present deliverable. For a full documentation of the LCT see the ISO draft/document.

4.1 Objectives

Driver distraction caused by secondary tasks could be characterized according to its origin. One may well assume that some secondary tasks are distracting due to high visual effort (e.g. a target search on a map) whereas other tasks are distracting mainly due to demanding cognitive processing (e.g. a conversation on the telephone with a highly complex content). It would – at least for scientific reasons – be desirable to have simple means to trace the nature of driver distraction. Ideally one might want to measure to what extend cognitive and visual distraction contribute to the overall distraction caused by a given secondary task. The present analysis was carried out to test whether a refined analysis of LCT makes such detection of the origin of workload possible (cognitive distraction vs. visual distraction).

4.2 Method

Existing LCT data from experiments which employed cognitive and visual secondary tasks was used. However, new dependent variables from the driving data were calculated to check for selective sensitivity of these variables to cognitive distraction vs. visual distraction.

4.2.1 Standard LCT and Analysis

The primary task in the LCT is a simulated driving task. It can be carried out either in a desktop setup, in a seating buck or in a more advanced driving simulator (vehicle cabin, large screen). *Figure 30* shows the simplest setup with a regular PC monitor and a PC game steering wheel. The LCT requires the test participant to drive a straight 3 lane course at a constant speed of 60km/h. The speed is system controlled and can not be affected by the participant. Participants strive to maintain their position in the centre of the indicated lane. While driving forward, participants encounter traffic signs (about every 10 seconds), which prompt them to change lanes. The signs are blank as the participant approaches them. Only at a distance of 40m before the sign the content which indicates the lane change becomes visible (see *Figure 30*). In a typical experiment, the participants perform multiple secondary tasks either in a blocked mode or in a mixed mode. In the blocked mode, the same secondary task is repeated for about 3 minutes, whereas in the mixed mode, the secondary tasks under test are mixed within one test run. *Figure 31* illustrates the standard analysis of LCT.

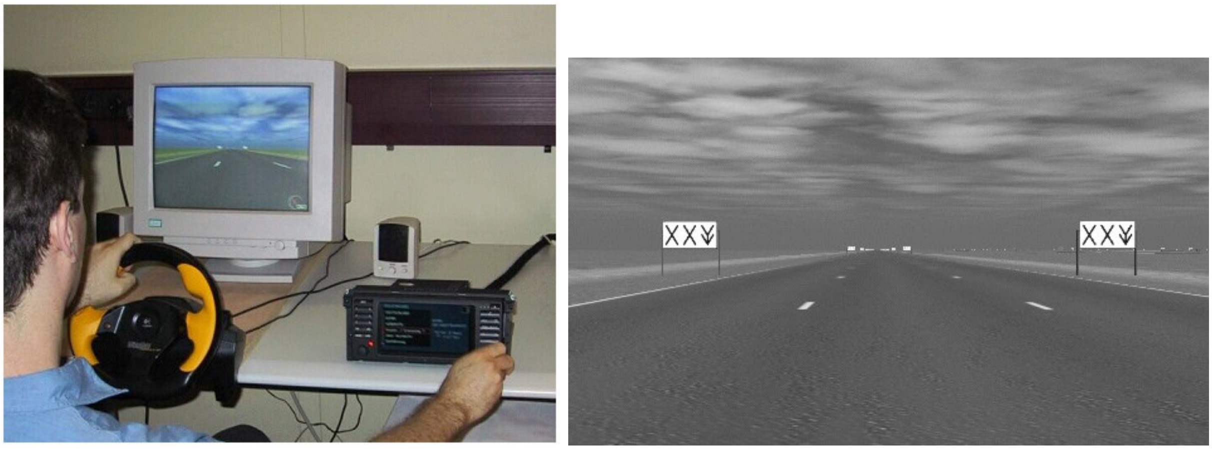


Figure 30 Physical layout of the PC based LCT (left), and a close up of the screen (right)

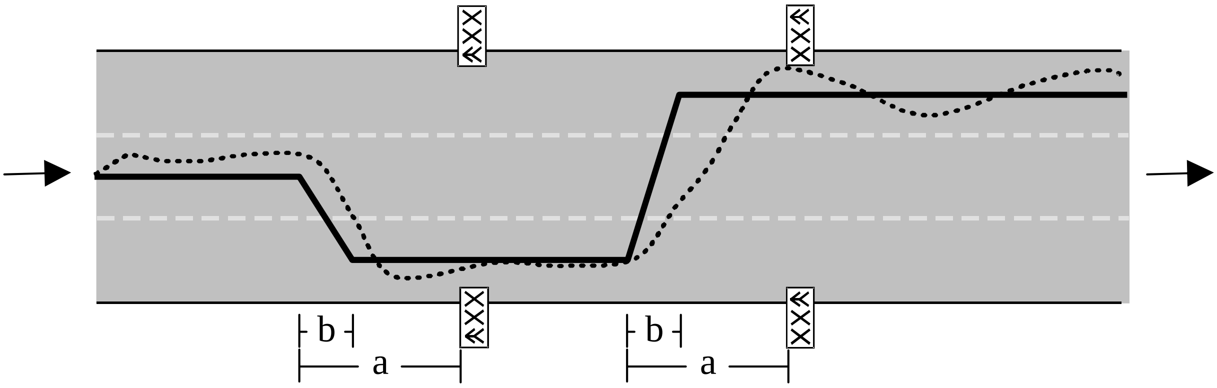


Figure 31 Standard analysis of LCT.

A simple normative model (solid line) is used to assess driving performance. The model assumes that the lane change starts 30 m before the sign (a) and is completed after 10 m (b). The dependent variable is the **mean deviation** of driving path (dotted line) from a normative model.

- Setting parameter a to 30 m proceeds from the assumption of a fixed minimum reaction time of 0.6s (Note that the information on the sign appears at a distance of 40 m).
- Identical values of parameter b for single as for double lane changes were employed due to empirical observations. Apparently, subjects are willing to perform more vehement steering manoeuvres when a double lane change is required.
- Note that all parameters of this model are fixed. That is, no fitting of the model to the behavioural data is necessary.

4.2.2 Modified Analysis

For the present analysis, additional parameters were calculated to test whether a modified analysis of LCT allows the detection of the origin of workload (cognitive distraction vs. visual distraction). The hypothesis, that LCT could be extended to discriminate between cognitive and visual distraction was based on results from the HASTE project (EU 5th framework, 2002-2005). Studies carried out within HASTE showed the following pattern:

Visual secondary tasks led to reduced event detection performance (e.g. increased PDT response time) and also to reduced lateral control (e.g. increased standard deviation of lane position). Cognitive secondary tasks, however, led only to reduced event detection performance while lateral control was unaffected or even improved. The LCT procedure contains both lane keeping components (straight part between the signs) and event detection components (reaction to sign). Based on the HASTE findings, one might expect that LCT should be differently affected by the two types of secondary tasks (visual vs. cognitive). In particular, the following pattern was predicted, see *Figure 32*:

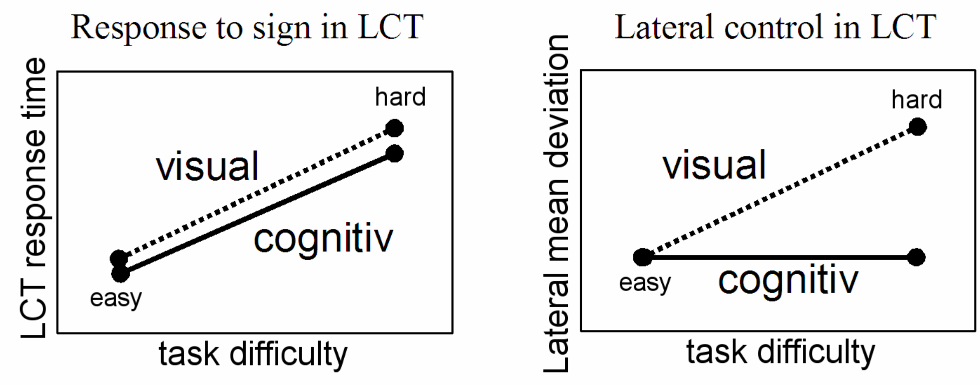


Figure 32 Expected differences between visual and cognitive tasks within the LCT

To test this, additional variables were calculated. These are illustrated in *Figure 33*.

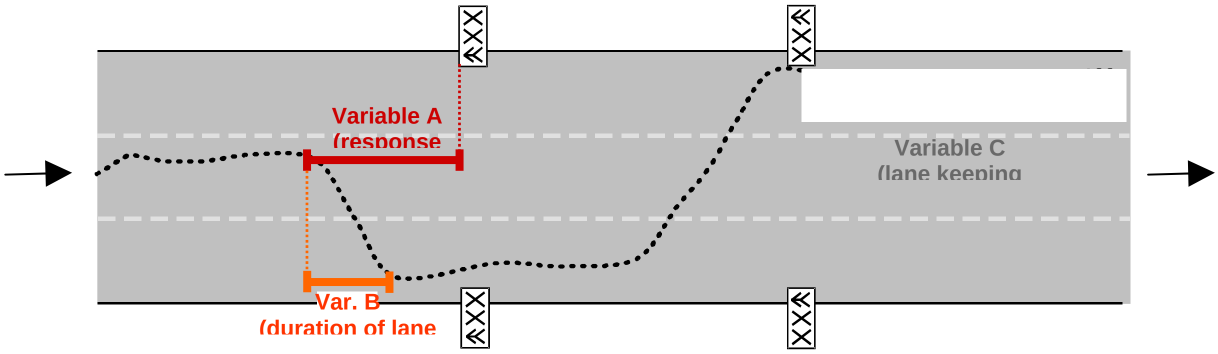


Figure 33 Definition of three additional variables used in the modified analysis

Variable A (response delay): an iterative algorithm estimates the onset of the lane change. The distance between this onset and the sign is used as an indicator of the participants response delay (smaller values indicating more delayed responses). This variable is used for testing the hypothesis that response time is delayed as an effect of visual and cognitive distraction.

Variable B (duration of lane change): in addition to the onset of the lane change, the end of the lane change was estimated. The distance between the two was used as a measure of the quality of the lane change manoeuvre. Note, that according to the instruction the participants are required to complete the lane change "quickly and efficiently". This variable is used for testing the hypothesis that visual, but not cognitive, load results in decreased lateral control performance.

Variable C (lane keeping): the mean absolute deviation from the centre of the lane between lane changes (i.e. on the straight part of the track) was calculated. As with variable B, this variable is used for testing the hypothesis that visual, but not cognitive, load results in decreased lateral control performance.

Data from two previous experiments were re-analyzed. In these experiments, an artificial visual secondary task was employed to manipulate visual distraction. *Figure 34* shows examples of the stimulus material. The participant's task was to search an array of circles for one target circle which was larger than the rest. Then they had to mark the target by a corresponding key press which turned the target side of the screen gray (see *Figure 34*). In the easy condition, the difference between target and distractor stimuli was large; in the difficult condition the difference was smaller, so that search time was increased. This manipulation revealed clearly significant main effects on Mean Deviation from the normative model in the LCT (Exp.1: $t(29)=2.57$, $p=0.016$; Exp.2: $t(19)=4.17$, $p<0.001$).

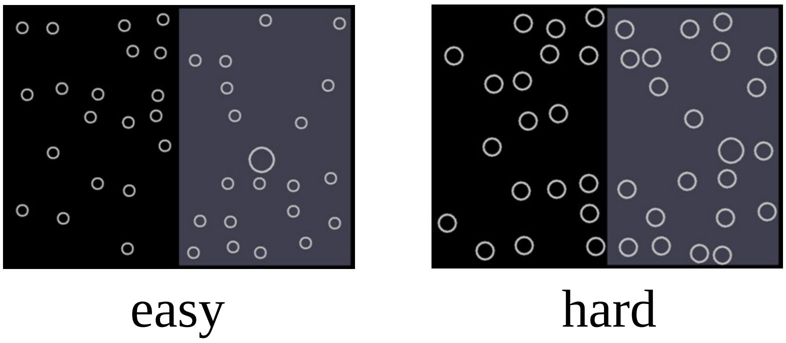


Figure 34 Two visual tasks layout easy and hard. The subject's task was to identify a circle with a larger radius.

As a cognitive secondary task, the participants had to count either up in steps of 2 (starting from 282) or count down in steps of seven (starting from 581). These two levels were also clearly discriminated by LCT Mean Deviation in either experiment (Exp.1: $t(29) = 4.58$, $p<0.001$; Exp.2: $t(19) = 3.62$, $p=0.002$).

4.3 Results

The results for data from Experiment 1 (n=30) are shown in *Figure 35*.

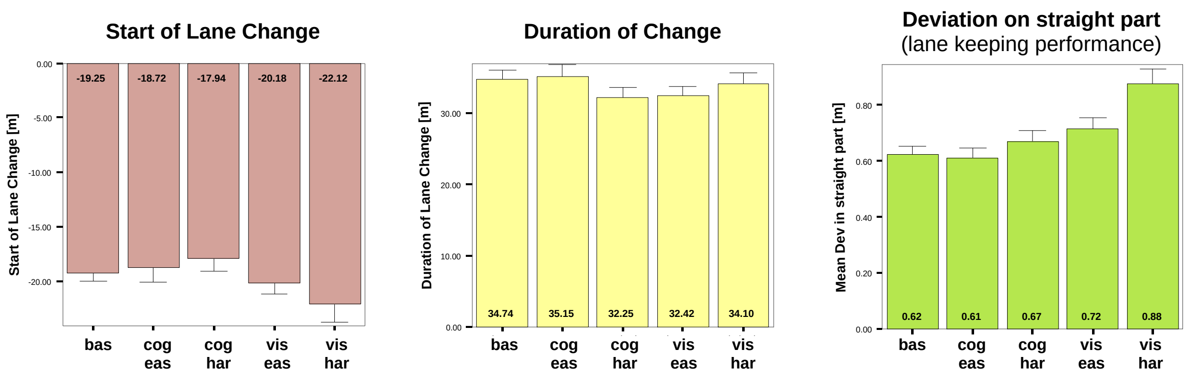


Figure 35 Results from Experiment 1

Variable A represents something like „reaction time“. A small value indicates an early reaction. Contrary to expectations, however, participants initiated the lane change in the most difficult condition (visual hard) earlier than in any other condition. However, the difference was not significant (5%).

Variable B represents one aspect of the quality of the lane changes, since participants were instructed to change the lane immediately. The only significant difference was between cognitive easy vs. cognitive hard. The difference, however, was contrary to the expectations! The lane changes were done more quickly in the hard cognitive condition than in the easy condition.

Variable C represents lane keeping quality between lane changes, i.e. on the straight part of the track. For both cognitive and visual secondary tasks the easy tasks lead to better lane keeping performance. However, only the visual tasks revealed a statistically significant effect.

To illustrate the results as compared to the standard analysis, Figure 36 shows p-values for single comparisons both for the classical analysis and for the new variables. As becomes obvious, only the standard analysis (top panel) brought satisfactory discrimination.

Analysis of the second data set brought roughly the same results. Again, the new variables showed hardly any significant difference even though the standard analysis (mean deviation from normative model over complete track) discriminated the secondary tasks very good.

Mean Dev (classic standard analysis of LCT)					
	Baseline	Cog_easy	Cog_hard	Vis_easy	Vis_hard
Baseline	---	0.007	0.000	0.000	0.000
Cog_easy		---	0.000	---	---
Cog_hard			---	---	---
Vis_easy				---	0.016
Vis_hard					---

Start of Lane Change					
	Baseline	Cog_easy	Cog_hard	Vis_easy	Vis_hard
Baseline	---	0.641	0.158	0.351	0.095
Cog_easy		---	0.422	---	---
Cog_hard			---	---	---
Vis_easy				---	0.337
Vis_hard					---

Duration of Lane Change					
	Baseline	Cog_easy	Cog_hard	Vis_easy	Vis_hard
Baseline	---	0.763	0.067	0.088	0.720
Cog_easy		---	0.006	---	---
Cog_hard			---	---	---
Vis_easy				---	0.386
Vis_hard					---

Mean Deviation in straight part between lane changes					
	Baseline	Cog_easy	Cog_hard	Vis_easy	Vis_hard
Baseline	---	0.605	0.201	0.054	0.000
Cog_easy		---	0.141	---	---
Cog_hard			---	---	---
Vis_easy				---	0.018
Vis_hard					---

Figure 36 Significance levels for t-Tests as an indicator of the sensitivity of each variable. Significant differences are indicated in green, tendencies in yellow and not significant differences in red. See text for more information.

4.4 Discussion and conclusions

The present re-analysis of LCT data was done in order to find out, if it is possible to track the origin of workload (namely, whether it is visual or cognitive in nature for a certain secondary task). This hypothesis was motivated by findings from the HASTE project, according to which visual load and cognitive load from secondary tasks exert different effects on driving behaviour. To test this hypothesis, new dependent variables were calculated for the data from two typical LCT experiments. However, it was not even possible to find a clear discrimination of the secondary tasks with any of the new variables. Of course, this overall negative finding rules out the possibility to search for more sophisticated interaction patterns as predicted in the hypothesis for this study (compare *Figure 32*). The finding was somewhat surprising since the new components calculated in the present analysis are implicitly contained in the standard analysis of LCT. The normative model which is used in the standard analysis of LCT is assumed to require lane keeping, early responses, and straight manoeuvres, i.e. those features of driving which were analyzed in the present study. Tracing the reasons for these discrepancies would most likely require specific data collection and much more detailed analyses. This, however, would go beyond the means of the present analysis. For the time being, it must be concluded that the hypothesized detection of the origin of workload within the standard paradigm of LCT is not easily possible.

5 Modified Lane Change Test for Assessing Adaptive IVIS Interfaces

5.1 Introduction

The standard version for the LCT was designed to measure the impact of typical secondary tasks such as the operation of car entertainment and information systems, but also other secondary tasks like eating or purely cognitive tasks. In the present study we sought to test whether a modified LCT would allow also testing adaptive features of an integrated system as it is planned for the AIDE system (AIDE metafunctions).

LCT in its standard version does not contain workload changes that could be employed to test adaptive IVIS. In order to manipulate workload, visibility conditions in the LCT were manipulated. The general scheme for the present study was as follows:

Primary Task: LCT with good/bad visibility conditions	+	Secondary Task: IVIS with/without workload adaptation

As no adaptive prototype was available, two simulated adaptive IVIS were employed in this study. In particular, the mode of information presentation with respect to the simulated workload was manipulated in two different ways.

5.2 Objectives

The aim of this study was to test whether a modified LCT can be used to evaluate adaptive features of future AIDE systems. The main question is whether a modified LCT can distinguish adaptive IVIS from non-adaptive IVIS, thus demonstrating one of the benefits of intelligent system adaptation.

5.3 Research Question

5.3.1 Hypothesis

The main hypothesis established to guide the study was: *A modified LCT with workload manipulation is able to discriminate adaptive and non-adaptive IVIS.* More specific questions were formulated later.

5.4 Method

5.4.1 Lane Change Test (LCT)

For a description of the lane change test, see 4.2.1. In the version of the LCT described in the ISO document, the symbols on the sign are presented at a distance of 40 meters. Contrary to this, the lane change signs and their content were always visible in the present study. The LCT was implemented in a driving simulator with a real car body. The road scene was projected on a large screen in front of the driver. The normative model which describes the idealized driving path was adjusted to the given visibility conditions. According to pre-tests, the ideal onset of a lane change was set to 100 m before the sign, both in the good and bad visibility condition. The bad visibility condition was realized by a diffusing hardware filter in front of the beamer which was used to project the LCT driving scene.

5.4.2 NASA TLX

The NASA Task Load Index (NASA-TLX) is a multi-dimensional subjective rating procedure on scales of 0 to 10 that provides an overall workload score based on a weighted average of ratings on six subscales: Mental Demands, Physical Demands, Temporal Demands, Own Performance, Effort and Frustration (Hart & Staveland, 1988). The NASA TLX which is used as the second measure of workload has been used in a variety of experimental studies and has shown its sensitivity in several studies of car driving (see D2.2.1). Its administration is sufficiently quick to be administered after every experimental condition in the LCT (2 tracks). The DALI developed in task 2.2.6 is quite similar to the NASA TLX (see D 2.2.6). It would have been interesting to use the DALI in our experiments but the validation of the scale has not been achieved at the time the experiments have been driven. The NASA-TLX consists of two parts: first, the estimation of workload on the six different subscales and second, the weighting process determined by pair wise comparisons among the six factors. However, Byers & Bittner (1989) found that the paired comparison is not necessary and Dickinson et al. (1993) stated that raw TLX scores based on equal weighting and scores who undergo the weighting process as proposed by the NASA TLX comparisons, correlate almost perfectly (r=.96). According to this, only the ratings on the six subscales were effectuated by the subjects in this study. Thus, the NASA Raw Task Load Index was used in this study. Furthermore, sum scores were calculated of the raw scores based on equal weights.

5.4.3 Demographic Questionnaire

A demographic questionnaire, established by DaimlerChrysler, was translated from German to French as the experiment was run in France. The questionnaire contains questions about age, gender, personal driving experience, activities while driving, experience with technical devices, use of technical devices while driving and preference of technical devices.

5.4.4 Sample

A total of 23 participants (11 male, 12 female) were invited to participate in the experiment. However, only 20 participants finally completed the simulator study, since three participants (2 female, 1 male) had to be excluded due to simulator sickness. The age ranged from 20 to 44 years (M = 32.9 years, SD = 8.6). Participants were recruited through announcements in the local newspaper or chosen from a registered database.

5.4.5 Experimental Design and Hypotheses

5.4.5.1 *Independent Variables*

Independent Variable 1: Workload

Two levels of workload were employed. Pilot studies with the LCT indicated that declined visibility (a "foggy" sight) resulted in higher subjectively estimated workload. In order to generate workload changes, visibility was thus changed several times in intervals of about 26 seconds, which corresponds to five changes of visibility within one 3-minutes track of LCT. *Figure 37* shows the two different levels of good and bad visibility.

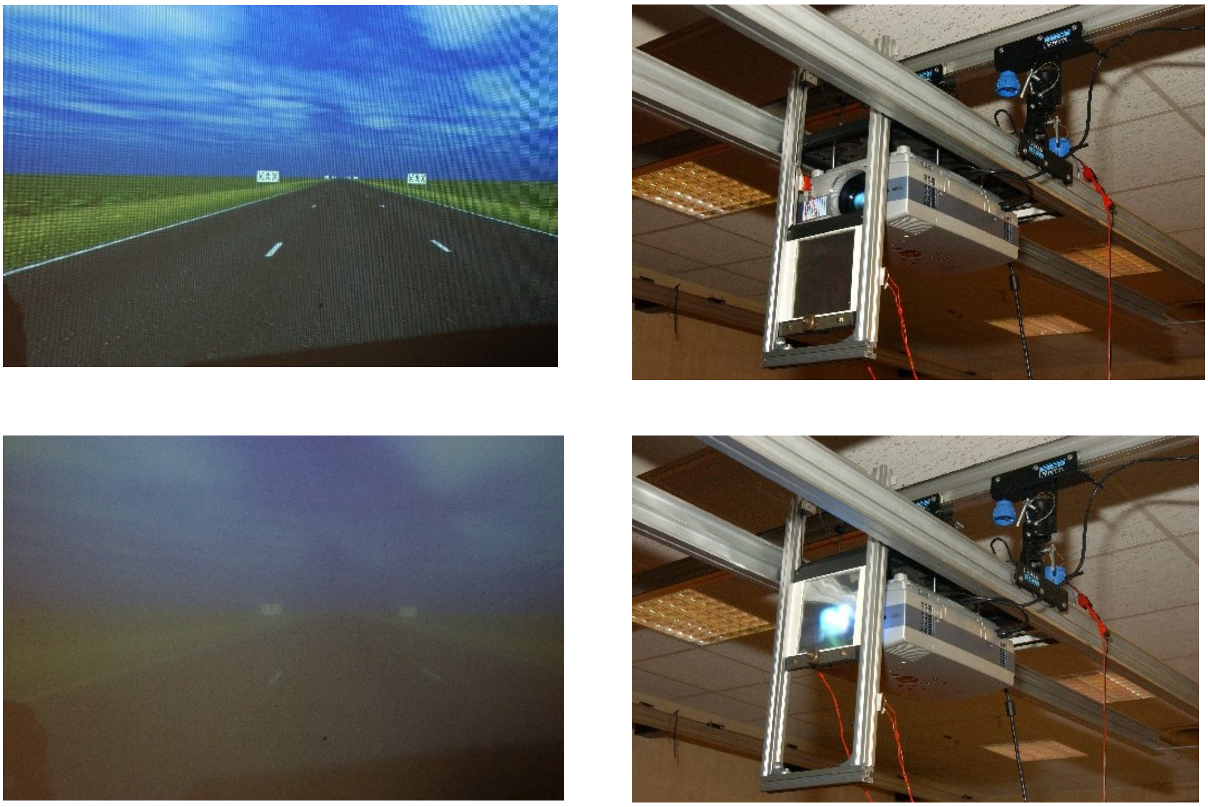


Figure 37 Good visibility due to open shutter (top pictures) and declined visibility due to closed shutter (bottom)

Independent Variable 2: Adaptive IVIS
Adaptation of the IVIS was simulated in two different ways:

- a change between sensory modalities for information presentation (visual vs. auditory).
- an increase of font size for visual displayed information (small font vs. large font).

Because no existing adaptive system could be employed for the present study, a simulation of a limited adaptive IVIS was used. The basic idea is to investigate the multiple resource model as described by Wickens (1984). According to this model the resources humans need to accomplish tasks are limited. The multiple resource model points out that there is interference and therefore decrease of performance when the modalities of information presentation are equal (e.g. the primary and the secondary task appeal mainly in a visual modality).

In our case, the hypothesis was that when visual workload is high, it is less distracting for the driver to get information presented in an auditory mode, because the information presentation modalities do not interfere. Therefore one way of adaptation of IVIS to higher workload was done by a switch to auditory information presentation. This version is called "*modality*". This adaptive behaviour was also described in the use cases in Deliverable 3.1.2 (Amditis et al, 2004). The use cases propose situations in which the modality of information presentation switches from a visual to an auditory mode when the driver is in a visually demanding or distracting situation (see use cases 2.2, and 2.3).

Another system feature which reflects IVIS adaptation to workload is the increase of font size for written information on displays. It is assumed, that increased letter size reduces the effort to read information and thus results in decreased workload. This version is called "*font size*".

To summarize, two different intelligent or "*good*" adaptive behaviours were chosen for the experiment:

1. Intelligent adaptation of the *modality* of the presentation of information (auditory-visually)
   - In the normal workload condition (good visibility) the information is presented visually
   - In the high workload condition (bad visibility), the information is presented in an auditory mode.
   - This System was called "auditory-visual good adaptive system"

2. Intelligent adaptation of the *font size* of the presented information (big-small)
   - In the normal workload condition (good visibility) the information is presented in a small font size (Arial 30)
   - In the high workload condition (bad visibility), the information is presented in a big font size (Arial 40)
   - This System was called "big-small font size good adaptive system"

One might argue, that effects due to the above described "good" system can not be unambiguously interpreted since it is not clear whether effects could be attributed to intelligent adaptation (e.g. large fonts when workload is high) or more generally to the fact that something in the IVIS changes. To facilitate the interpretation, the "good" adaptive system was compared to a system which also produces changes, but the opposite way. This system shall be called "*bad*" adaptive system. Two bad adaptive behaviours were simulated:

3. Bad adaptation of the modality of the presentation of information (auditory-visually)
   - In the normal workload condition (good visibility) the information is presented in an auditory mode.
   - In the high workload condition (bad visibility), the information is presented visually.
   - This System was called "auditory-visual bad adaptive system"

4. Bad adaptation of the font size of the presented information (big-small)
   - In the normal workload condition (good visibility) the information is presented in a big font size (Arial 40)
   - In the high workload condition (bad visibility), the information is presented in a small font size (Arial 30)
   - This System was called "big-small font size bad adaptive system"

A non-adaptive system with the following features served as a reference for comparison.

5. In both workload situations, the information was presented aurally only (auditory non adaptive). This served as a reference to the good and bad adaptive system behaviour: "*modality*".
6. In both workload situations, the information was presented visually (only) (visuall non adaptive) in small font size only. This served as a reference to the good and bad adaptive system behaviours: "modality" and "font size".
7. In both workload situations, the information is presented visually in big font size only. This served as a reference to the good and bad adaptive system behaviour: "font size".

See also *Figure 38* and *Table 6* for more information.

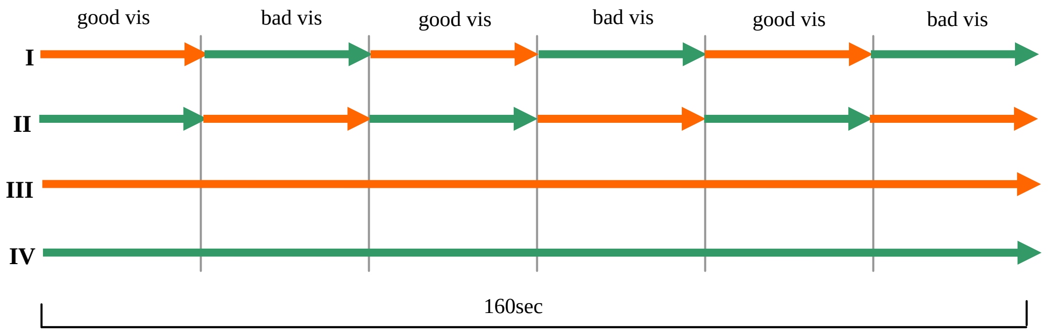


Figure 38 Information presentation according to workload, example: system behaviour modality (Vis = Visibility, I = auditory-visual good adaptive, II = auditory-visual bad adaptive, III auditory non-adaptive, IV = visually non-adaptive)

Table 6 Summary of Independent Variables

Workload	Normal	Good visibility condition in LCT
	High	Bad visibility conditions in LCT (fog)
Dimensions of information presentation	Modality	auditory vs. visual
	Font size	30 vs. 40
IVIS	1	Adaptive System 1 (“good” adaptive system), 2 adaptive system behaviours: 1. Modality: normal workload (good visibility) → visually presented information; high workload (bad visibility) → auditory presented information 2. Font size: normal workload (good visibility) → presentation visually in small font size (30); high workload (bad visibility) → presentation visually in big font size (40)
	2	Adaptive System 2 (“bad” adaptive system), 2 bad adaptive system behaviours: 3. Modality: normal workload (good visibility) → auditory presented information; high workload (bad visibility) → visually presented information 4. Font size: normal workload (good visibility) → presentation visually in big font size (40); high workload (bad visibility) → presentation visually in small font size (30)
	3	Non-adaptive Systems 5. auditory stimuli only 6. visual small font size information only (30) 7. visual big font size information only (40)

5.4.5.2 *Secondary tasks*

The secondary tasks were short messages, consisting of two positive statements, followed by a question. The statements were presented both aurally and visually (text message). The participants were asked to respond to the questions verbally. An example for these messages would be:

I	Mary eats a peach. George plays the guitar.	II	Who eats a peach?

The auditory messages were recorded by a female voice in wav-format while the written text messages were saved in text format. Both types of messages were presented on the integrated navigation system of the vehicle.

5.4.5.3 *Specific Hypotheses*

Our main hypothesis as described above was reformulated in terms of our independent variables:

Modified LCT with different visibility conditions is able to discriminate between adaptive and non-adaptive IVIS.

Furthermore, for different types of system adaptations (modality/font size change) the effect sizes might differ. E.g., the system feature "font size" might produce only a small effect for adaptive vs. non-adaptive IVIS, whereas the system feature "modality change" might produce a larger effect. Therefore a second hypothesis was:

The modified LCT allows discriminating different adaptive behaviours of adaptive systems ("modality change" vs. "font size").

This hypothesis is depicted in *Figure 39*. The figure illustrates the assumption that the modality change (right panel) produces a larger effect than the font size change (left panel).

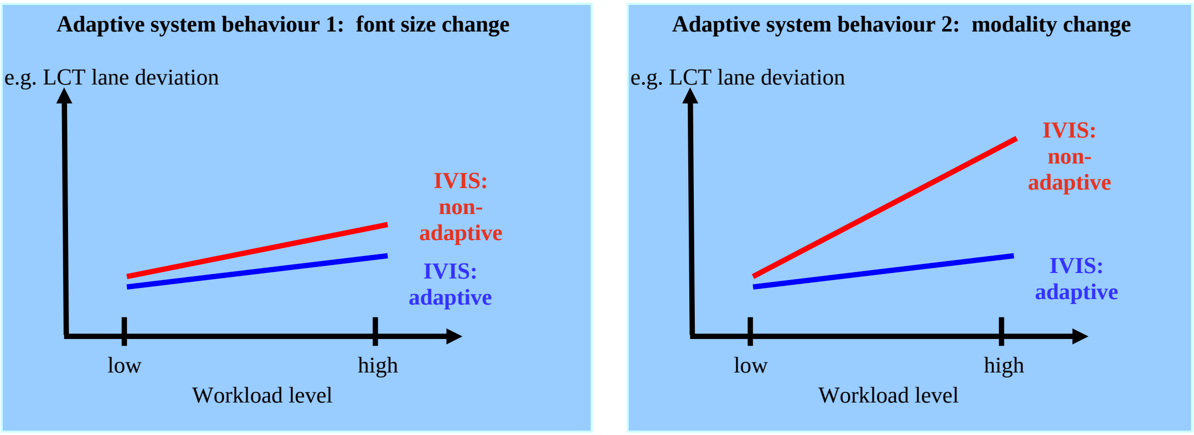


Figure 39 Illustration of the hypothesis, that the adaptive feature "modality change" (right panel) might produce a larger effect than "font size" (left panel).

In order to detect the benefits of a good adaptive system, it is compared with a bad adaptive system, which behaves the opposite way. Therefore, third hypothesis was formulated as:

The modified LCT allows distinguishing between good and bad adaptive systems.

5.4.6 Conditions and order

Altogether, seven different behaviours of systems were tested in a repeated measures design. These are the three non adaptive (auditory only, small font size only, big font size only), the two good adaptive systems (small font size in normal workload situation and auditory/ respective big font size presentation in high workload situation) and the two bad adaptive systems (big font size/ respectively auditory presentation in normal workload situation and small font size presentation in high workload situation).

Two tracks of LCT without secondary tasks (baseline) were added for each subject. To obtain at least two minutes of data per condition (as recommended in the ISO document), each system was used on two LCT tracks of approximately three minutes each. To avoid order effects, order of the conditions was randomized. The two workload conditions were randomized whereas the secondary tasks were balanced according to a Latin Square. However if a person started a track with a good visibility condition and the good adaptive system for modalities was applied, the second track would be the same system, as stated before, but starting with bad visibility. The two tracks of the same secondary task condition were always used consecutively. This randomisation of tracks was employed for all conditions except for the baseline, which were fixed for each participant such that one baseline was driven at the beginning and the other at the end of the experimental phase.

5.4.7 Procedure

At first the participant was asked to fill in the demographic questionnaire. Then the participant made himself comfortable in the car and a short explanation of LCT assisted by a screenshot projected on the screen was given. Then the phase of familiarization started. Each participant drove two tracks of LCT without any tasks or change of visibility to get used to the simulation. After that, the changing of visibility was introduced and each participant drove one track of LCT with visibility changes. Then the secondary tasks were explained and practised without driving. Three messages of each existing feature were presented. The last step of the familiarization phase was the driving of one track of LCT with visibility changes together with the secondary tasks. Finally, the participants were asked if they felt comfortable with the simulation and the tasks, and if they felt ready to start.

The experimental phase consisted of 16 tracks of LCT (each about 3 minutes), two of which were baseline. After each condition the participant filled in the NASA-TLX. A mandatory break of 10 minutes was made after seven LCT tracks (ca. 30 minutes). The participants were also offered other breaks if necessary, but none of the participants requested this. Overall duration of an experimental session was about 2 hours. For an overview see *Table 7* below.

Table 7 Procedure

	Time	Number of tracks	Activity	Order
Phase of familiarization	6 min	2	**familiarization with LCT** without tasks	
	3 min	1	**familiarization with LCT and visibility change** without tasks	
	3-5 min	0	**familiarization secondary tasks annexes with** LCT	
	3 min	1	**familiarization with secondary tasks, LCT visibility change**	
	3 min	1	Baseline	Random ordre
	3 min	0	NASA-TLX	
test 1	6 min	2	**Good adaptive system** visual information with good visibility and auditory information with bad visibility	
	3 min	0	NASA-TLX	
	6 min	2	**Bad adaptive system** auditory information with good visibility and visual information with bad visibility	
	3 min	0	NASA-TLX	
	6 min	2	**Good adaptive system** Small font with good visibility and big font with bad visibility	
	3 min	0	NASA-TLX	
	10 min	0	Break	
test 2	6 min	2	**Bad adaptive system** Big font with good visibility and small font with bad visibility	Random order
	3 min	0	NASA-TLX	
	6 min	2	Auditory information only (visibility changes every 30 sec.)	
	3 min	0	NASA-TLX	
	6 min	2	Visual small font information only (visibility changes every 30 sec.)	
	3 min	0	NASA-TLX	
	6 min	2	Visual big font only (visibility changes every 30 sec.)	
	3 min	0	NASA-TLX	
	3 min	1	Baseline	
	3 min	0	NASA-TLX	

5.4.8 Apparatus

The static simulation was realized with a Peugeot 407, which was equipped with a serial navigation system of 7.87 inch screen diagonal. The Peugeot was equipped with a Logitech Formula Force GP steering wheel and pedals in order to record the movements of the steering rod. The LCT was run on a Pentium PC, 4.2 GHz, with a GeForce 5600 FX video card. The road scene was projected on a big screen with a diagonal of 147 inches and the distance of between the driver's eyes and the screen was 124.4 inches (ca. 3.16 m). The eye-to-navigation system distance was 31.5 inches (ca. 0.80 m). Speakers were placed under the bonnet to simulate a realistic engine sound. Speakers for the messages, which were presented auditory, were placed in the driver's cab and the volume was adjusted for each of the participants individually. The messages were generated on a PC and presented on the display of the integrated car navigation system. To manipulate visibility, a hardware shutter was build. A filter made of a pile of ten standard transparencies (as used for overhead projection) was placed in front of the beamer to simulate bad visibility condition. This filter was moved up and down with an electronic motor. It should be noted that the experiment was run highly automated. A central master PC controlled both the visibility changes and the presentation of the messages for the secondary tasks. *Figure 40* and *Figure 41* illustrate the general experimental setup.

Figure 40 Test vehicle, screen with simulation

Figure 41 Test vehicle and technical equipment

5.4.9 Dependent variables

The main dependent variable was the mean deviation from the normative model as typically calculated for the LCT (ISO, 23005), see also Figure 36. This variable consists of a comparison of the trajectory of the participant with a normative trajectory, in which the lane changes begin at 100m before the sign and ends 10 meters later for each single and double lane changes[4]. The calculation of the mean deviation between the normative model and the real driving task is assumed to cover a set of characteristics of the driving behaviour, such as

[4] When using a setting in a simulator with just the computer screen it is recommended by ISO to set the start of the lane change of the normative model to 60m. Since in this study a projection on a screen was used, the content of the panels could be detected at a distance of 100 m, thus the normal trajectory model was adapted.

lane keeping quality, delayed responses to the sign and the quality of the lane change manoeuvre (Mattes, 2003).

Errors in the answers to the questions were recorded as the main indicator of secondary task performance. Task completion time could not be employed as a meaningful index of secondary task performance since the procedure was highly automated and there was hardly any variation in reaction time. For the NASA-TLX the sum score of the scale was used. Additionally, the single subscales were used for some in-depth analyses.

5.4.10 Statistical Analysis

The LCT raw data was pre-processed with the standard LCT analysis software (v1.99) to calculate the mean deviation value for each subject and each condition. Then, two factorial repeated measures ANOVAS were calculated with factors workload (2 levels) and system (4 levels). Three such ANOVAs were run, namely one for each adaptive system (modality and font size) and one ANOVA in order to test the differences between the two adaptive systems. Furthermore, repeated measures ANOVAS were calculated for the sum score of the secondary task performance and NASA-TLX. Six additional ANOVAS were calculated for the different subscales of the NASA-TLX. The effect size η^2 as calculated by the repeated measures ANOVA with SPSS were reported if a main effect or interaction was found to be significant. For validation on external criteria Pearson's Correlation Coefficients were calculated for the NASA-TLX scores and LCT. A 5% significance level was used in the analysis.

5.5 Results

5.5.1 Discrimination of the different systems with LCT mean deviation

5.5.1.1 *System: Modality*

A 2 (visibility) x 4 (features of system) repeated measures ANOVA was calculated to investigate possible significant differences with respect to LCT mean deviation. A significant main effect of workload was found, $F(1, 19) = 150.12$, $p < .05$, $\eta^2 = .89$. There were no main effects of system feature or interactions between system feature and workload. See Table 8 and Figure 42 below.

Table 8 Means and critical values for system: modality

Feature of system	**Auditory**	**Visual small font**	**Auditory visual bad adaptive**	**Auditory visual good adaptive**
Normal workload *M (SD)*	1.36 (0.42)	1.42 (0.37)	1.37 (0.39)	1.30 (0.33)
High workload *M (SD)*	2.34 (0.50)	2.35 (0.57)	2.38 (0.45)	2.41 (0.56)
	F(19)	SE		
Workload	150.12*	0.89		
Feature of System	0.15			
Workload x Feature	0.40			
*$p < .05$.				

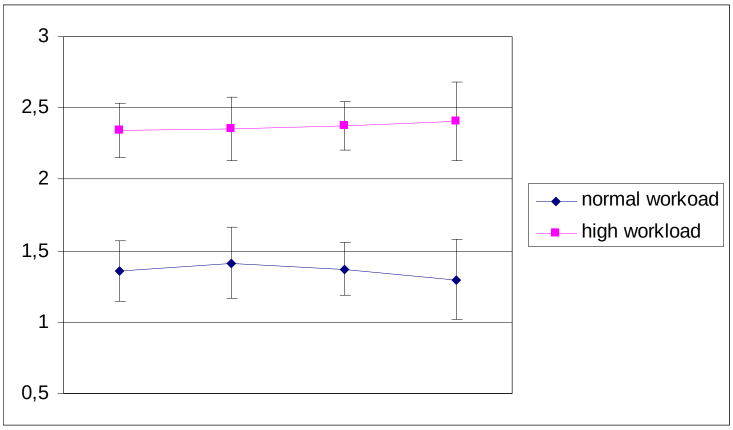


Figure 42 Group means of LCT deviations for the two workload conditions and four system modalities. (AUD = auditory only, SFO = small font size only, AVB = auditory-visual bad adaptive, AVG = auditory-visual good adaptive, LCT lane dev = LCT mean deviation)

5.5.1.2 *System: Font size*

A 2 (visibility) x 4 (features of system) repeated measures ANOVA was calculated with respect to LCT lane deviation. A significant main effect of workload was found, $F(1, 19) = 93.91$, $p < .05$, $\eta^2 = .83$. There were no main effects of system feature (font) or interactions of system feature and workload. See Table 9 and Figure 43.

Table 9 Means and critical values for system: font size

Feature of system	**Visual big font size**	**Visual small font size**	**Big-small font size bad adaptive**	**Big-small font size good adaptive**
Normal workload *M (SD)*	1.42 (0.52)	1.42 (0.37)	1.30 (0.40)	1.48 (0.47)
High workload *M (SD)*	2.30 (0.49)	2.35 (0.57)	2.32 (0.53)	2.41 (0.54)
	F(19)	SE		
Workload	93.91*	0.83		
Feature of System	1.47			
Workload x Feature	0.30			
*p < .05.				

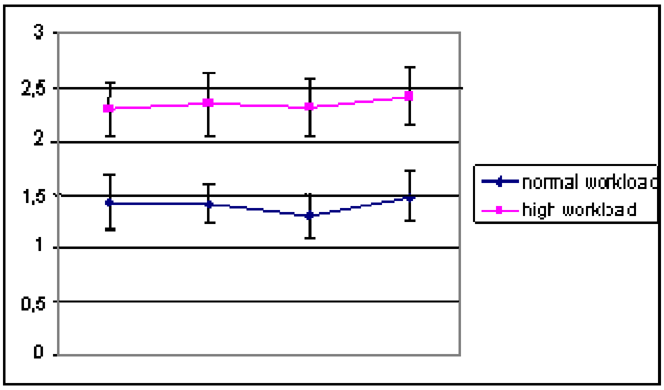


Figure 43 Group means of LCT deviations for the two workload conditions and four font conditions. (SFO = small font size only, BFO= big font size only, BSB = big-small font size bad adaptive, BSG = Big-small font size good adaptive, LCT lane dev = LCT mean deviation)

5.5.1.3 *System Comparison*

A 2 (visibility) x 4 (system) repeated measures ANOVA was calculated was calculated with respect to LCT lane deviation. A significant main effect of workload was found, $F(1, 19) = 145.03$, $p < .05$, $\eta^2 = .88$. There was no main effect of system or interaction of system and workload found, see Table 10 and

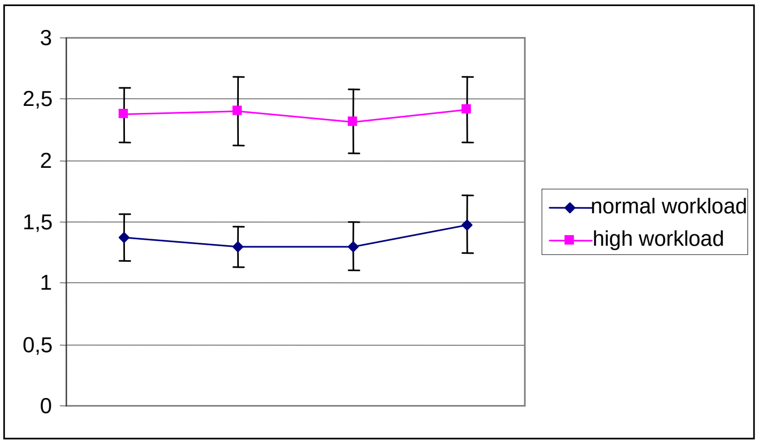


Figure *44* below.

Table 10 Means and critical values for system comparison

Feature of system	Auditory- visual bad adaptive	Auditory-visual good adaptive	Big-small font size bad adaptive	Big-small font size good adaptive
Normal workload *M (SD)*	1.37 (0.39)	1.30 (0.33)	1.30 (0.40)	1.48 (0.47)
High workload *M (SD)*	2.38 (0.45)	2.41 (0.56)	2.32 (0.53)	2.41 (0.54)
	F(19)	SE		
Workload	145.03*	0.88		
Feature of System	1.47			
Workload x Feature	0.30			
*$p < .05$.				

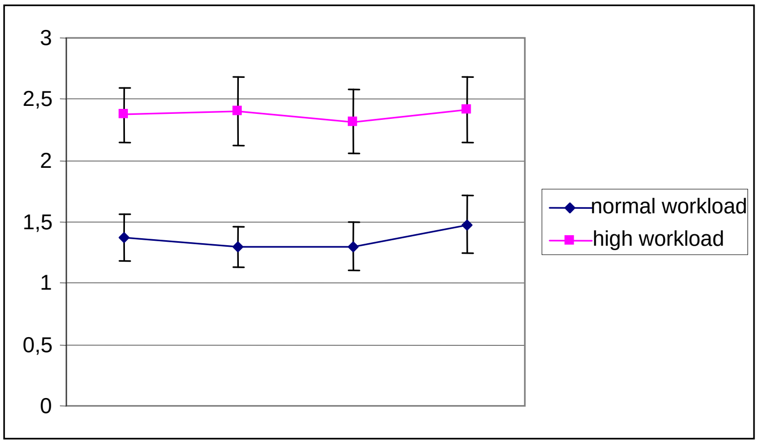


Figure 44 Group means of LCT lane deviations for the two workload conditions and four system conditions (AVB = auditory-visual bad adaptive, AVG = auditory-visual good adaptive, BSB = big-small font size bad adaptive, BSG = Big-small font size good adaptive, LCT lane dev = LCT mean deviation)

- The different visibility conditions were distinguished in terms of LCT mean deviation for all features of the systems, whereas there was no difference found for the different systems or the features of the systems.

For the comparison of the different levels of workload the same normative model was used, indicating the lane change at 100 meter before the actual panel. However, to make sure that the significant effects concerning workload are not only due to the possible fact that the panels in the high workload condition (bad visibility) could not be clearly detected at 100m, the same statistical analyses was calculated with an adapted normative model for the high workload situation. A normative model suggesting the lane change at 80m before the panel was used. This model was established by several experts, who drove the simulation and indicated the point at which they could definitely see the signs on the panels. Most of the experts even detected the content of the panel at 100m; however at 80m they were surely detected. However, the statistical analyses using the mean deviation calculated with this new model led to the same results as the formerly reported analyses.

5.5.2 Discrimination of the different systems with NASA TLX

5.5.2.1 *Sum Score*

As with the NASA TLX only the whole tracks of LCT and not the different workload conditions were judged, one overall repeated measure ANOVA was calculated. No significant effects could be found, $F(6, 114) = 1.89$, $p = .09$. See Table 11 and *Figure 45* below.

Table 11 NASA TLX un-weighted means and critical value of sumscore

Feature of system	Auditory	Visual small font	Auditory visual bad adaptive	Auditory visual good adaptive	Visual big font size	Big-small font size bad adaptive	Big-small font size good adaptive
M (SD)	4.48 (1.40)	4.77 (1.36)	4.71 (1.34)	7.75 (1.64)	5.18 (1.40)	4.89 (1.32)	5.03 (1.41)
F (19)	1.89						
$^{*}p < .05$.							

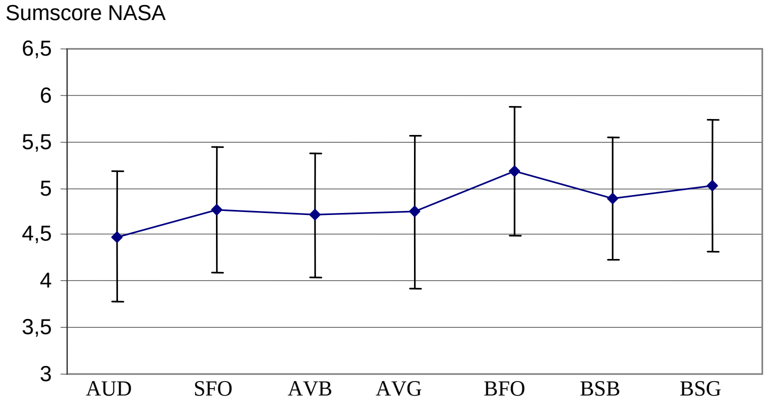


Figure 45 Group means for NASA-TLX sumscores. (AUD = auditory only, SFO = small font size only, AVB = auditory-visual bad adaptive, AVG = auditory-visual good adaptive, BFO= big font size only, BSB = big-small font size bad adaptive, BSG = Big-small font size good adaptive)

- The sum score of the NASA TLX, did not detect any differences between the different IVIS.

5.5.2.2 *Sub scales of the NASA TLX*

To see if there were some significant differences between the systems for the different subscales of the NASA TLX one repeated measures ANOVA was calculated for each of the subscales. Significant results were found only for the subscale “effort”, $F(6, 114) = 2.34$, $p < .05$, $\eta^2 = .11$. See Table 12, Table 13 and *Figure 46* below.

Table 12 NASA TLX un-weighted means of subscales

Feature of system	Auditory	Visual small font	Auditory visual bad adaptive	Auditory visual good adaptive	Visual big font size	Big-small font size bad adaptive	Big-small font size good adaptive
Mental Demand *M (SD)*	5.01 (1.80)	5.68 (2.21)	5.86 (2.07)	5.80 (2.31)	6.52 (1.82)	6.14 (2.11)	5.86 (2.18)
Physical demand *M (SD)*	3.91 (2.21)	4.67 (2.30)	4.25 (2.34)	4.39 (2.58)	4.54 (2.35)	4.45 (2.53)	4.58 (2.22)
Temporal Demand *M (SD)*	4.62 (2.20)	4.92 (2.28)	5.18 (2.25)	5.00 (2.45)	5.58 (1.67)	5.35 (2.47)	5.44 (2.36)
Own Performance *M (SD)*	6.57 (2.27)	5.40 (2.74)	5.44 (2.12)	5.85 (2.24)	6.03 (2.29)	5.45 (2.49)	5.73 (2.20)
Frustration *M (SD)*	2.14 (1.86)	2.58 (1.91)	2.57 (2.28)	2.49 (1.81)	2.91 (2.06)	2.36 (1.96)	2.94 (1.95)
Effort *M (SD)*	4.51 (2.40)	5.35 (2.25)	4.88 (2.24)	4.87 (2.64)	5.49 (2.23)	5.56 (2.32)	5.51 (2.43)

Table 13 NASA TLX critical values of subscales

Feature of system	Mental Demand	Physical Demand	Temporal Demand	Own Performance	Frustration	Effort
F (19)	1.83	0.82	1.27	1.77	1.20	2.34*
SE						0.11
*$p < .05$.						

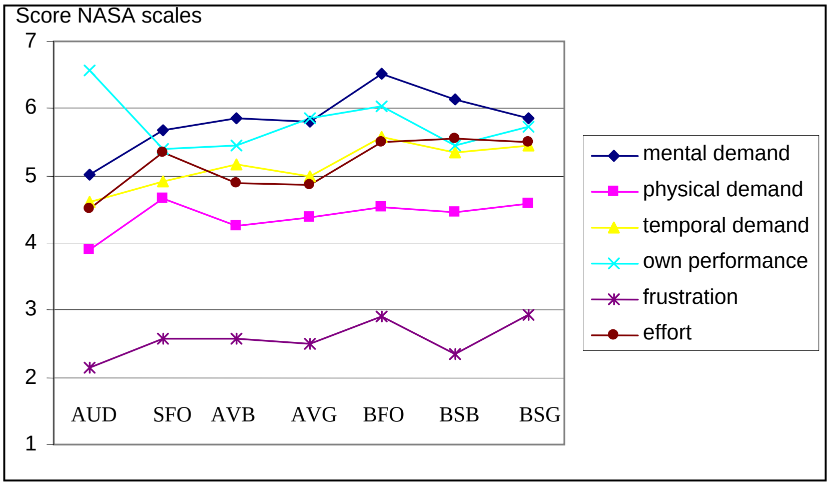


Figure 46 Group means for NASA TLX six subscales and the six system conditions. (AUD = auditory only, SFO = small font size only, AVB = auditory-visual bad adaptive, AVG = auditory-visual good adaptive, BFO= big font size only, BSB = big-small font size bad adaptive, BSG = Big-small font size good adaptive)

- The subjective measure of workload, the sum score of the NASA TLX, did not detect any differences between the different IVIS.
- Only the Subscale “Effort” discriminated the different IVIS.

5.5.3 Correlation of LCT Mean Deviation and NASA TLX sum score

Only the correlation between LCT Mean Deviation and NASA TLX sumscore for the auditory-visual good adaptive system was found to be significant, $r(\text{NAVB}, \text{LCTAVB}) = .57$, $p < .05$. All other relevant correlations were not significant. See Table 14 below.

Table 14 Pearson's Correlation NASA TLX sum score with LCT mean deviation for whole tracks

	Auditory only	**Visual small font only**	**Adaptive bad auditory-visual**	**Adaptive good auditory-visual total**	**Visual big font only**	**Adaptive bad small-big fontsize total**	**Adaptive good small-big fontsize total**
r(NASA sumscore, LCT mean deviation)	-.426	-.113	-.223	-.569(*)	-.150	-.312	-.129
$^{*}p < .05$.							

5.5.4 Secondary Task Performance

5.5.4.1 *Secondary Task Performance – system: modality*

A 2 (visibility) x 4 (system) repeated measures ANOVA was calculated for task performance (errors) across system modalities. There were a main effects of workload, $F(1, 19) = 7.51$, $p < .05$, $\eta^2 = .36$, system modality, $F(3, 57) = 10.61$, $p < .05$, $\eta^2 = .28$, and an interaction, $F(2.25, 42.66) = 19.49$, $p < .05$, $\eta^2 = .51$. See Table 15 and Figure 47.

Table 15 Means and critical values for secondary task performance, system: modality

Feature of system	**Auditory**	**Visual small font**	**Auditory visual bad adaptive**	**Auditory visual good adaptive**
Normal workload *M (SD)*	0.00 (0.00)	0.70 (0.92)	0.15 (0.37)	1.05 (0.83)
High workload *M (SD)*	0.10 (0.31)	0.85 (0.88)	1.45 (0.89)	0.35 (0.59)
	F(19)	SE		
Workload	7.51*	0.36		
Feature of System	10.61*	0.28		
Workload x Feature	19.49*	0.51		
*$p < .05$.				

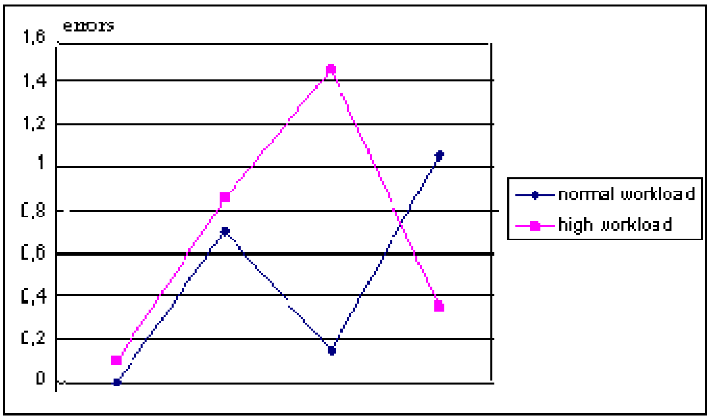


Figure 47 Group mean errors for the four system modalities and two workload conditions (AUD = auditory only, SFO = small font size only, AVB = auditory-visual bad adaptive, AVG = auditory-visual good adaptive)

5.5.4.2 *Secondary Task Performance – system: font size*

A 2 (visibility) x 4 (system) repeated measures ANOVA was calculated for task performance (errors) across system font size. There were no significant main effect or interactions found. See Table 16 and Figure 48 below.

Table 16 Means and critical values for secondary task performance, system: font size

Feature of system	**Visual big font**	**Visual small font**	**Big-small font size bad adaptive**	**Big-small font size good adaptive**
Normal workload *M (SD)*	0.80 (1.01)	0.70 (0.92)	1.00 (1.26)	0.85 (1.14)
High workload *M (SD)*	0.65 (0.88)	0.85 (0.88)	0.65 (0.75)	0.85 (0.59)
	F(19)	SE		
Workload	0.51			
Feature of System	0.15			
Workload x Feature	0.58			
$^{*}p < .05$.				

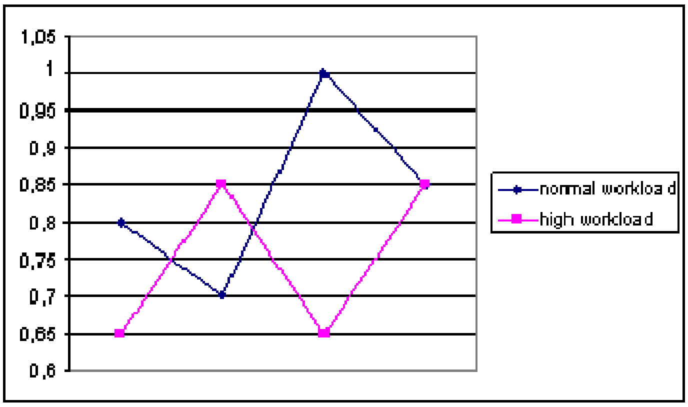


Figure 48 Group mean errors for system font and two workload conditions (SFO = small font size only, BFO= big font size only, BSB = big-small font size bad adaptive, BSG = Big-small font size good adaptive)

5.5.4.3 *Secondary Task Performance – comparison of systems*

A 2 (visibility) x 4 (system) repeated measures ANOVA was calculated for task performance (errors) across system (adaptive feature). There were no significant main effects. However, the interaction between workload and adaptive feature was significant, $F(8.55, 0.65) = 13.07$, $p < .05$, $\eta^2 = .41$. See Table 17 and Figure 49.

Table 17 Means and critical values for secondary task performance, comparison systems

Feature of system	**Auditory visual bad adaptive**	**Auditory visual good adaptive**	**Big-small font size bad adaptive**	**Big-small font size good adaptive**
Normal workload *M (SD)*	0.15 (0.37)	1.05 (0.83)	1.00 (1.26)	0.85 (1.14)
High workload *M (SD)*	1.45 (0.89)	0.35 (0.59)	0.65 (0.75)	0.85 (0.59)
	F(19)	SE		
Workload	0.24			
Feature of System	0.29			
Workload x Feature	13.07*	0.41		
*$p < .05$.				

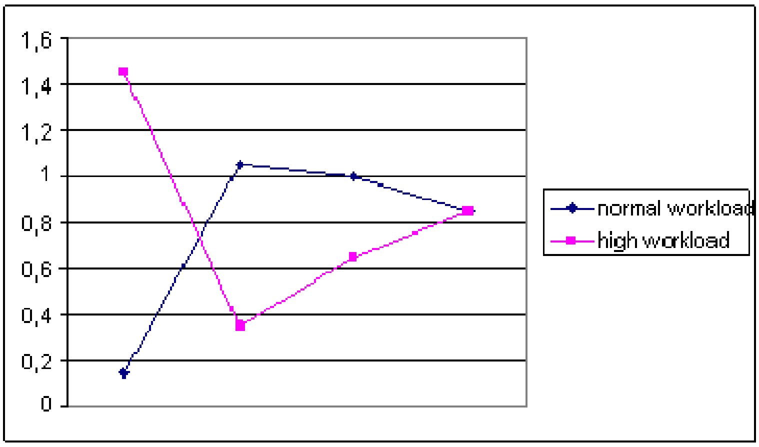


Figure 49 Group mean errors for the four system adaptive features and two workload conditions: AVB = auditory-visual bad adaptive (auditory for normal workload, visually for high workload), AVG = auditory-visual good adaptive (auditory for high workload, visually for normal workload), BSB = big-small font size bad adaptive (big font for normal workload, small font for high workload), BSG = Big-small font size good adaptive (big font for high workload, small font for normal workload)

In order to locate the differences, single t-tests have been calculated. Workload conditions were compared for each feature of the system and also the systems as a whole were compared to the other systems. Altogether 14 t-tests have been calculated, finding that the secondary task performance is significantly better for the auditory-visual bad adaptive system in the normal workload condition, $t(19) = 5.94$, p<.05 and for the auditory-visual good adaptive system there are significantly less errors in the high workload situation, $t(19) = 4.27$, $p<.05$. Concerning the task performance for the system as a whole only the differences for the auditory system with the visual small font size, $t(19) = 4.31$, p<.05, and the auditory system with the visual big font size system, , $t(19) = 4.24$, p<.05, are significant, stating that there are significantly less errors for the auditory condition.

- In terms of the secondary task performance (number of errors) significant differences could be found for workload only for the auditory-visually good and bad adaptive systems. Whenever the presentation within these two systems was auditory, fewer errors were made, meaning that the secondary task performance was better.

- As for the main effect of the IVIS itself, significantly fewer errors were made in the auditory condition as compared to messages presented in big and small font size.

5.6 Discussion and conclusions

The main purpose of this study was to test, whether advantages of intelligent car system management could be demonstrated with a modified version of the LCT. The LCT was introduced as a simple and cost effective method to estimate driver distraction caused by secondary tasks on IVIS. The present study was sought to be a first test whether higher-level features of HMI systems such as intelligent workload dependent information management could also be tested in such a simple and efficient way. To this end, a modified version of the LCT was developed and tested. In this modified version the visibility of the driving scene was manipulated in order to simulate different levels of environmentally induced external workload. Furthermore, simple visual and auditory tasks were generated as fictitious information from a simulated adaptive system. As a subjective measure of workload, NASA TLX was applied.

It turned out that neither the objective measure of LCT mean lane deviation nor the subjective workload measure could discriminate between the simulated system levels. However, such discrimination would have been a prerequisite to test whether adaptive features could be discriminated. The only significant differentiation was between the simulated levels of workload (good vs. bad visibility) which is, however, neither surprising nor overly helpful. Therefore, it must be concluded, that it was not possible to demonstrate in the present study that the modified LCT scenario is a method to investigate higher level adaptive function of a vehicle HMI.

Concerning the correlations between the NASA TLX and the LCT mean lane deviation, only one correlation between the LCT mean deviation and the NASA TLX sum score was found to be significant, namely a medium, negative correlation for the auditory-visual good adaptive system. This correlation is even somewhat paradoxical, since it means that low LCT mean deviation is associated with high subjective estimated workload. However, it is not surprising that in general the correlations did not turn out to be significant, since neither of the instruments could discriminate the different features of the system and without variation co-variation cannot be expected. Therefore one has to be careful when interpreting the one significant correlation, since neither NASA TLX nor LCT showed a difference between the auditory-visual good adaptive system and the other systems.

A closer look at the secondary task performance shows that the participants took great care to fulfil the secondary tasks. Very few errors were made throughout all of the different tasks. Only a few significant differences could be found concerning the systems for the plain auditory condition when compared to the visual only conditions. Concerning workload, there were only significant differences for the auditory-visual good and bad adaptive system. Whenever the task presentation was auditory, there were significantly less errors than when it was visually presented.

One reason for the failure to find significant differences for the IVIS in this study might be that the simulated level of the IVIS was simply not different enough. This view is supported by the fact that both objective (LCT) and subjective variables (NASA-TLX) revealed no difference. Probably, overall difficulty of the secondary tasks was to low to produce a meaningful difference. This assumption is supported by the observation, that the NASA-TLX values for the baseline condition (LCT tracks without secondary tasks; M = 3.69 SD = 1.23) and for LCT tracks with secondary tasks (M = 4.83, SD = 1.41) were in the same order of magnitude.

Generally, the difficulties to demonstrate effects of system adaptation in a simulated laboratory environment agree with previous observations, for example in the COMUNICAR project (Weiland, et al., 2003). These authors concluded that testing of adaptive functions turned out to be an extremely intricate subject. Nevertheless, it can be expected, that comparison of sub-features of an adaptive HMI (e.g. visual vs. auditory presentation) is clearly possible in a driving simulator environment and also with the standard version of the LCT. However, to design a highly standardized environment in which tests of super ordinate features of adaptive systems (information management) were possible would require more extensive research. Probably there are even logical limits to this when it turns out that the environmental and psychological setting specific for the application of an intelligent vehicle information system cannot be reproduced in a laboratory environment.

6 Lane Change Test – Effects of Scenarios and Simulators

6.1 Experimental objectives:

The LCT has mainly been used with a standard PC set-up (see ISO, 2005). However, the LCT as a method should not be tied to a specific set-up. Many different set-ups are possible, ranging from a simple PC set-up to a mock-up with beamer and screen (Hallen, 2004) to a set-up using static or dynamic driving simulators. Furthermore, it could be possible to also use the LCT with real cars, using some special equipment to measure the steering signal from the wheels. The objective of the following experiment was to test if different LCT set-ups could lead to different results or if they provide comparable results. To compare the different set-ups we used tasks that are of interest for the AIDE project. These include IVIS as well as nomadic devices. Furthermore, we used a so-called content-free calibration task (secondary/reference task) that should help to ensure stable results between trials and test sites. The results of this study should also contribute to the ISO standardization process of the LCT and related calibration tasks.

6.2 Research question

The basic research question for this experimental study was to clarify if the LCT as described in the ISO draft is scalable over different experimental set-ups. Therefore, two typical static driving simulator set-ups (large projection vs. smaller plasma screens) were compared. For the evaluation typical IVIS tasks like destination entry have been chosen along with so called nomadic devices tasks (e.g. smartphone tasks). A subset of the simulator set-up tasks has also been compared to a simple standard PC set-up e.g. described in Mattes (2003).

6.3 Method

6.3.1 Subjects

A total of 30 subjects (Age M= 38.9, SD = 9.35, 2 female) participated in the study.

6.3.2 Experimental design

A two factorial mixed design was used for this study. The factor LCT set-up had two realizations, driving simulator plasma and driving simulator projection. This factor was realized between-subjects. The second factor was constituted by the tasks and was realized within-subjects. Results for some tasks were compared with the results of the same tasks for a standard PC set-up of the LCT.

There were three plasma screens used for the plasma condition. Only a front view was used and the field of view was horizontal 105° and a vertical 20°. For the projection condition also a three channel front view was used with a horizontal field of view 140°, and vertical 37°. In both cases we used a sampling rate of 25Hz. In both conditions objects had the same size on the retina (same viewing angle).

As dependent variables we measured the LCT deviation as described in chapter 4 (4.2.1). (see also ISO, 2005). Furthermore we measured performance and reaction times for the calibration task. Participants were instructed according to ISO draft (ISO, 2005).

6.3.3 Calibration task

The calibration task we used in this experiment was developed by DaimlerChrysler in the ADAM project and presented in the ISO working group on standardization of the LCT. The task draws mainly on visual and motor resources while cognitive effort can be neglected. The task may help to compare different set-ups of the LCT over different laboratories and can be described as a visual search task. Furthermore the task is easy to be installed and can be used in a standardized way. Currently there is a Task Force within ISO/WG8 for the standardization of such a secondary calibration task.

The task for the participants was to discriminate one larger circle (target) from various smaller ones (distractors). To indicate that the right circle was found, the participant has to move a grey vertical rectangle over the screen (see *Figure 50*).

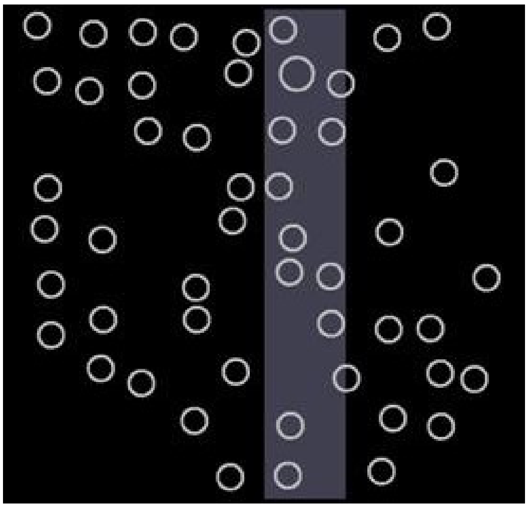


Figure 50 The screen of the Calibration task. The grey cursor indicates the position of the larger circle.

The movement of the cursor had to be done with the arrow keys of a keypad that was located in the middle of the console (see *Figure 51*).

Figure 51 The keypad for the Calibration task

The target circle had always the size of 44 arc minutes to the participant (distance 78 cm). By increasing the size of the distractors (3 levels) and decreasing the size of the cursor (3 levels) three levels of difficulty were realized. By decreasing the size of the cursor it was assumed that the amount of visual-manual load is increased, not only because step size is increased but also because it requires more visual-manual fine-tuning. *Figure 52* gives examples of the three levels of difficulty for the calibration task. *Table 18* gives an overview of the specification for the three levels of difficulty.

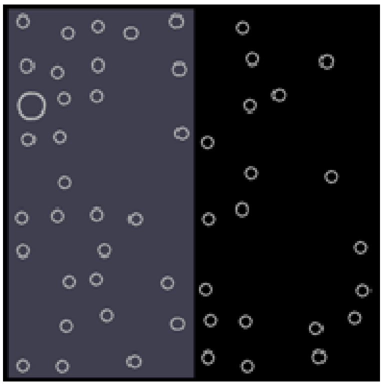
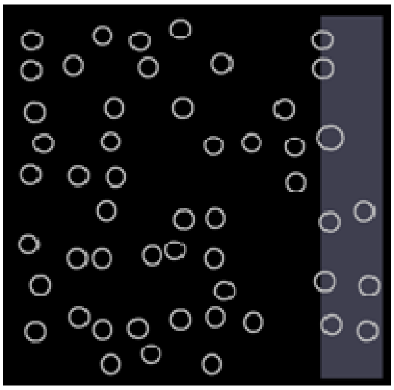
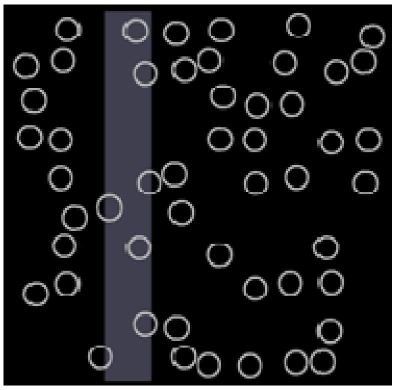
		
Easy	**Average**	**Difficult**

Figure 52 Three levels of difficulty for the calibration task.

Table 18 Specification for three levels of Difficulty of the calibration task.

Difficulty	Amount Distractors	Size Distractors	Cursor Stepsize
Easy	50	22 arc minutes	2
Average	50	35 arc minutes	6
Difficult	50	40 arc minutes	8

6.3.4 IVIS applications to be tested

The following nine tasks were used in this experiment. The tasks were chosen to satisfy the need to evaluate traditional, integrated IVIS (e.g. destination entry CARIN BMW system) and upcoming nomadic devices (e.g. different PDA functionalities). *Table 19* gives a short overview of the tasks and devices used for the experiment.

Table 19 Overview of Tasks and devices.

	Task	**Device**
Task 1	Destination Entry	CARIN
Task 2	Radio Tuning	CARIN
Task 3	Destination Entry	TomTom Navigator
Task 4	Writing SMS	Motorola MPx220
Task 5	Writing SMS	PalmOne (Treo)
Task 6	Select date in calendar	PalmOne (Tungsten)
Task 7	Calibration Task Easy	Calibration Task Set-up
Task 8	Calibration Task Average	Calibration Task Set-up
Task 9	Calibration Task Difficult	Calibration Task Set-up

The destination entry task, the radio tuning task (both with CARIN system) and writing SMS tasks have already been used within the ADAM context. However, in this study some other devices have been used to keep up with the technological development. A more detailed description of the tasks is given below.

Task 1: Destination entry using CARIN system
The task was to enter a street name as a destination and then start the navigation. As input device the participants used the rotary knob that can be pushed and rotated to select e.g. letters from a speller (see *Figure 53*). The CARIN system was mounted in the centre stack of the mock-up. Step lengths for street names were constant over tasks.

Figure 53 The speller functionality of the CARIN system.

Task 2: Radio tuning using the CARIN system
With this task participants had to do a manual radio-tuning starting from a predefined radio station. Step length for the manual radio-tuning task was the same for all task repetitions. *Figure 54* shows the radio of the CARIN system. Again, the input device was the rotary knob of the CARIN system.

Figure 54 Radio menu of the CARIN system.

Task3: Destination entry using TomTomGo
The TomTomGo is a mobile navigation device that was mounted on the windshield and has a touch screen as input device. The task was to enter a street name and start navigation. Letters are selected by selecting them directly from the screen (see *Figure 55*). Again, step lengths for street names were constant over tasks.

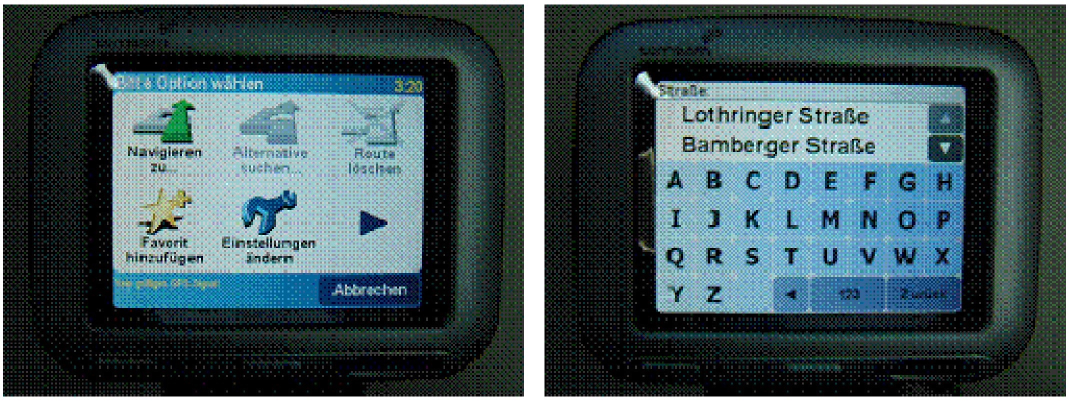


Figure 55 Destination entry with mobile navigation device TomTomGo.

Task 4: Writing SMS using a mobile phone
A SMS containing of three words had to be typed in (without using the T9 mode) on a mobile phone. The mobile phone is shown in *Figure 56*.

Figure 56 Mobile phone used for manual SMS entry.

Task 5: Writing SMS using PalmOne Treo.
This task was merely the same as task 4(writing three words), but a different input device had to be used. While the mobile phone had a conventional phone keyboard the PalmOne had a QWERTY-keyboard (see *Figure 57*) which had to be used by the drivers.

Figure 57 PalmOne by Treo.

Task 6: Select a date in a calendar using PalmOne Tungsten
The task was to select a specific date, by opening the calendar function of the organizer (see *Figure 58*), changing to the monthly view and then opening the specific date, reading aloud what it was. The Palm was mounted on the windshield. Step length for each task was constant over tasks.

Figure 58 PalmOne Tungsten for Calendar task.

Figure 59 shows the driving simulator set-up with the devices for the secondary tasks.

Figure 59 Secondary tasks in a mock-up for LCT in static driving simulator.

6.4 Results

A first analysis was performed to check if baseline measures (without any secondary task) differed between the two LCT driving simulator set-ups. *Figure 60* shows that there were no significant differences between the set-ups in terms of baseline lane deviation. However, the mean deviations for baseline were quite high. This may be due to the instructions to the participants. The latest ISO draft specifies a training criterion for baseline drives that is a mean (M) lane deviation of M <=1. While this has been included in the latest ISO draft, the experiment reported here has been carried out according to a previous ISO draft without recommendations for specific instruction and training.

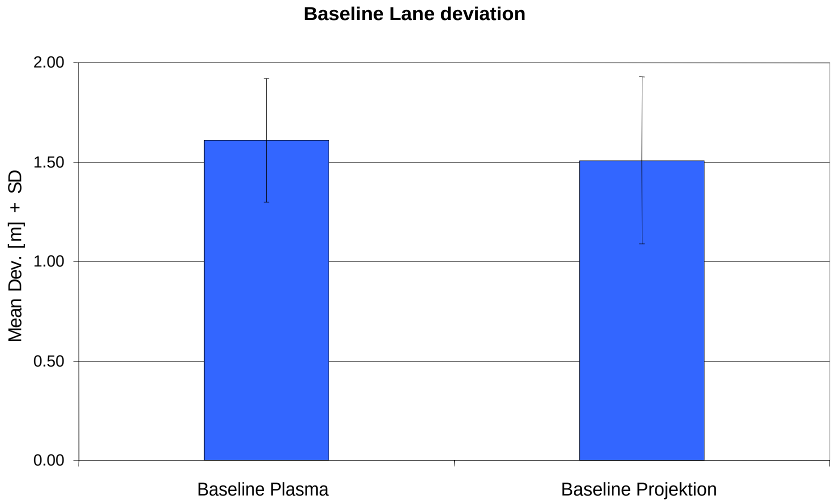


Figure 60 Baseline Lane Deviations betweens driving simulator set-ups.

In a second step we analysed the performance while working on the secondary tasks described above. *Figure 61* shows the results for both driving simulator conditions and all secondary tasks.

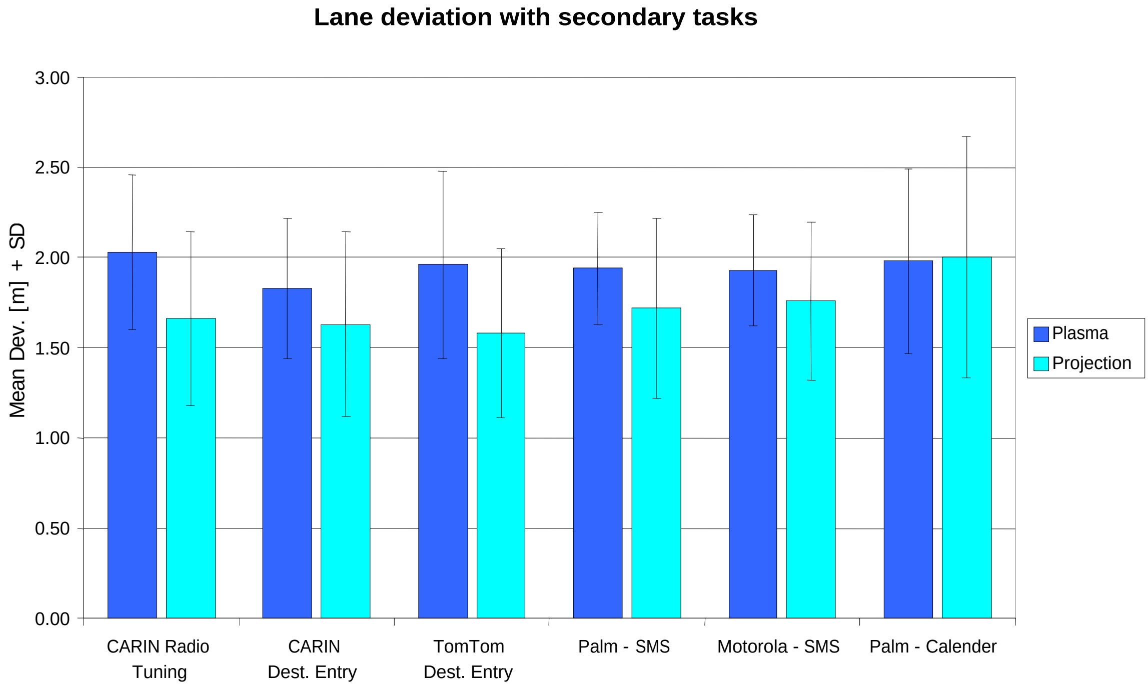


Figure 61 Lane Deviations for type of set-up and secondary tasks.

We did not find a significant effect for the factor secondary tasks but there was a significant factor for simulator set-up ($F = 9.1$, $p < 0.01$). However, the small $\eta^2 = 0.06$ reveals only a small effect. Comparable with this result was the finding for the calibration task, see *Figure 62*. Again we only found a significant result for the factor simulator set-up ($F= 8.01$, $p < 0.01$) but not for the factor secondary task or the interaction. Again, only a small effect was revealed ($\eta^2 = 0.09$).

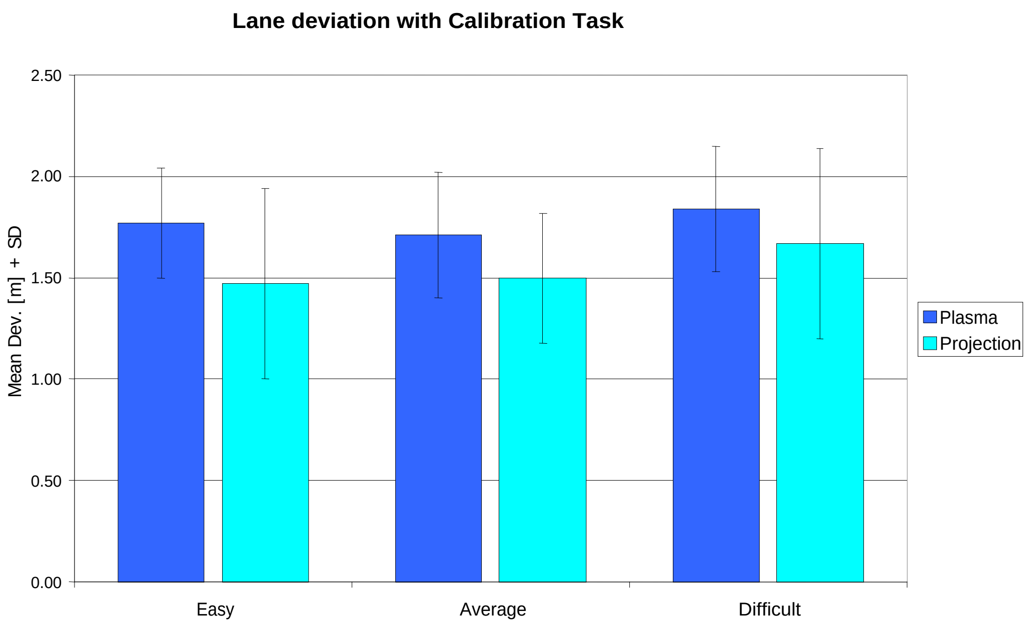


Figure 62 Lane Deviations for type of set-up and Calibration Task.

A comparison with results from the standard PC LCT set-up revealed that the standard PC set-up had the highest lane deviations. An analysis of variance of variance revealed a significant effect for set-up ($F = 7.2$, $p < 0.01$) but no effect for the factor task ($F = 1.2$) or the interaction ($F = 0.3$). However, because there were no significant differences between the tasks, no analysis of the task ranking could be performed. An important result is that both simulator conditions lead to a reduced variance in secondary task performance (see *Figure 63*) compared to a PC set-up. A post hoc analysis revealed medium effects for the differences between PC set-up and Projection for the destination entry task ($d = 0.5$), a small effect for the radio tuning task ($d = 0,29$), a small effect for the palm task ($d = 0.29$) and a medium effect for the SMS task ($d = 0.66$). The results implicate that at least between set-ups of different fidelity (PC vs. projection) differences can be found. Because we did not find any differences between tasks no conclusions can be drawn if these differences between simulators would also affect a task order.

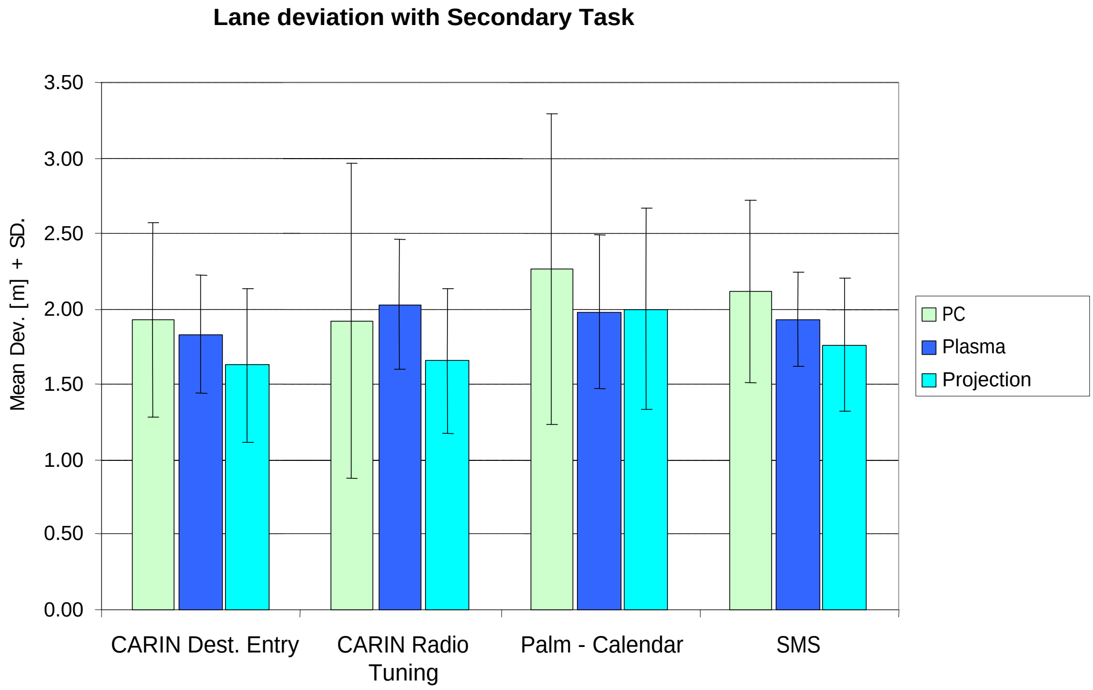


Figure 63 Lane Deviations for type of set-up and Secondary Task.

Although we did not find differences in LCT performance between tasks we found a highly significant result for Total Task Times (TTT, $F = 216.16$, $p < 0.01$). Again, there was no effect of TTT over both simulator set-ups, see *Figure 64*.

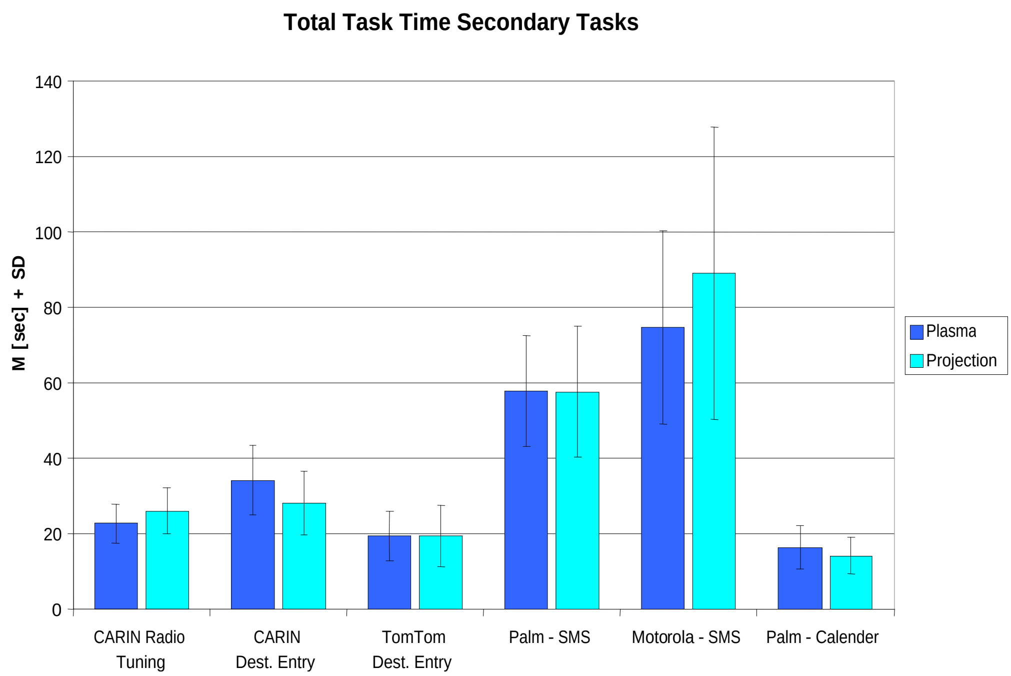


Figure 64 Total Task Times of secondary tasks for each LCT set-up.

A comparable result was found for the calibration task. While we did not find a difference in the performance measure, we found differences in the TTT for the different levels of the calibration task (see *Figure 65*). A significant effect for level of difficulty (F = 70.4 $p < 0.01$) was found, while there was no difference between the set-ups.

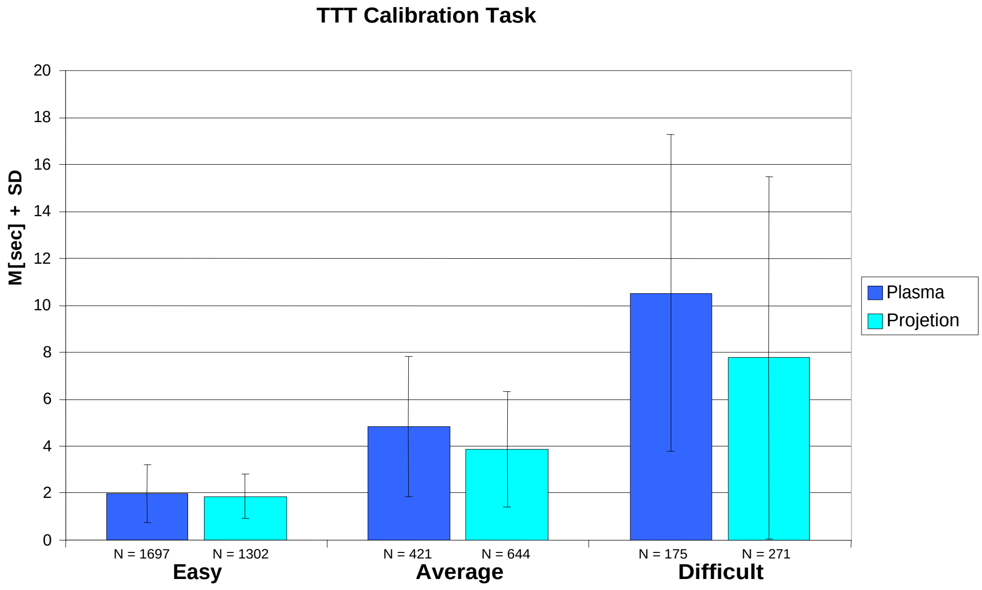


Figure 65 TTT for each type of Calibration Task for both simulator set-ups.

6.5 Discussion and conclusions

The results show a mixed picture. While we have found differences in Total Task Times (TTT) for the secondary tasks we used, we could not identify any performance differences with the LCT. There are two possible reasons for this. First, the LCT was not sensitive to the visual/cognitive/manual load of the tasks. Second, there was no difference in visual/cognitive/manual load of the tasks itself. Previous studies of the LCT suggest (e.g. Mattes, 2003) that the second interpretation is true, that is, there was no difference in terms of performance degradation between the tasks while there was a difference in TTT. The sensitivity of LCT to different task characteristics has been shown by the ADAM project, for example (e.g. Mattes, 2003).

Furthermore, we could not find any differences in lane deviations between the three types of the calibration task. However, there have been significant differences in TTT for the calibration tasks, as well as for the frequency of solved tasks. Only these dependent variables represented the three steps of difficulty correctly. The results show again that participants were able to adjust their performance in the secondary task (increased TTT, frequency of solved tasks) while not being affected in the LCT performance.

More important is the finding of differences in lane deviations for the same tasks between different set-ups. The mean values for the Plasma condition have always been slightly higher than those in the projection condition. However, the effect was small. A possible reason for this small effect may be the larger image size of the projection condition that enabled an earlier decision on which lane to change.
A comparison of the results between the simulator set-ups and a PC-set-up showed small to medium effects. This result indicates that changing from one set-up to another may alter the

values for LCT performance. However, it is unclear if this is only a linear transformation or if this may affect task ordering. Therefore in an ideal comparison of set-ups we would have found differences in task performance showing consequently that task ordering is not affected by changing from one set-up to another. This could not be realized in our case, as the distraction potential of all tasks has been found the same over all set-ups. Despite the differences in the mean values between the set-ups, the results show that in the PC set-up an increased variance and the highest mean values for all three set-ups were found. This may be due to the simple steering wheel used in the PC condition (game steering wheel) while in the simulator condition a real steering wheel with a more naturalistic behaviour was used. It may be speculated that the high variance found in the PC condition will make it harder to discover small effects between systems. Further, the experiments reported here have been carried out according to an early ISO draft version. As pointed out before, the new ISO draft instructions seem to be designed to reduce variance by paying more attention to training of the subject. As a consequence one should specify what steering wheel as well as which set-up was used when reporting LCT results. The results of this study did not address different vehicle dynamics or manipulated other simulation characteristics concerning the LCT. In this experiment the underlying simulation characteristics between the plasma and projection set-up were held constant. Yet, the vehicle dynamics of the PC set-up and the driving simulations have been different. However, the main differentiation should lie in different presentation conditions. Nevertheless it is clear that simulated vehicle dynamics characteristics may affect LCT results and should therefore be an important study goal for further research.

To summarise, a fruitful effort has been done, to implement the LCT in a driving simulator environment. This gives the opportunity to test the distraction potential of IVIS (integrated systems as well as nomadic devices) in a real mock-up. The experiment showed that the LCT is capable of measuring performance degradation while working on a secondary task, despite the effect of task length. However, we could not prove an effect of the calibration task on performance data. Participants in this experiment seemed to adopt a strategy where the driving task had higher priority and therefore an effect could only be shown in the performance of the secondary task. The use of this calibration task, its scalability and effects on performance should be an issue of further research.

To summarize, a fruitful effort has been done, to implement the LCT in a driving simulator environment. This gives the opportunity to test the distraction potential of IVIS (integrated systems as well as nomadic devices) in a real mock-up. The experiment showed that the LCT is capable of measuring performance degradation while working on a secondary task, despite the effect of task length. However, we could not prove an effect of the calibration task on performance data. Participants in this experiment seemed to adopt a strategy where the driving task had higher priority and therefore an effect could only be shown in the performance of the secondary task. The use of this calibration task, its scalability and effects on performance should be an issue of further research.The LCT is a simple laboratory method that can be easily implemented in a driving simulator as well and is therefore useful in different stages of the design cycle as an evaluation tool. However, it has to be assured that LCT experiments are referenced according to the ISO draft they have been set up to guarantee consistency between test sites. Furthermore, the experiment reported here does not tangle the question if the LCT is capable of evaluating IVIS incorporating meta-functions that was done in chapter 5.

7 Specification of Selected Driving Control Metrics

In chapter 3 relevant potential driving control metrics were analysed and discussed, and a set of metrics to be recommended for the inclusion in the AIDE pilots was selected. Some new metrics were developed based on the analyses and discussions.

7.1 Introduction

This chapter provides definitions, descriptions of use, interpretation guidelines and limitations of driving performance metrics. Note that requirements on sensors for the measurements are not included in this report. Instead, these are given in *D3.3.1 Requirements for AIDE HMI and safety functions and also in the internal deliverable "Vehicle Sensor Requirements for the AIDE prototype evaluation in WP 2.4*." (Peters & Östlund, 2004) However, in some cases guidelines concerning data processing are given.

7.2 Mean Speed

7.2.1 Definition

Mean speed is defined as the average travel speed (km/h).

7.2.2 Description of use

This is one of the easiest metrics to include in field experiments. However, if the pulse generator/tachometer used for speed calculation is not calibrated, speed data will not be accurate. Note that also the speedometer most likely is not accurate. Data collected from different vehicles should therefore not be directly compared – unless the accuracy of the speed recordings and speedometers has been verified.

Mean speed is a reliable and valid metric of driving performance in scenarios where the driver chooses and controls the speed, such as in rural road and motorway driving during low traffic intensity hours. In urban roads and when there are many surrounding vehicles, speed is rather chosen collectively (by the traffic flow) than individually. Thus, consider excluding speed data during car following situations and other situations in which the driver does not choose speed herself/himself.

7.2.3 Interpretation guidelines

Reduced speed is commonly interpreted as the driver compensating for increased distraction or workload. Reduced speed may however be unsafe, e.g. when driving in a fast lane of a motorway. Occasionally, cognitive workload has been found to increased speed. This may possibly be explained by failing to monitor the own and other vehicles' travel speed due to e.g. decreased situation awareness or distraction.

Mean speed is of course influenced by speed limitations – unless the driver is too distracted to see the road signs. If speed supporting ADAS controls speed, the speed should be influenced accordingly.

7.2.4 Limitations

Mean speed is not useful in scenarios where speed is controlled by traffic flow or other environmental factors. Urban road is not a feasible scenario for speed measurement.

7.3 Maximum Speed

7.3.1 Definition

Maximum value of vehicle speed

7.3.2 Description of use

See mean speed (7.2)

7.3.3 Interpretation guidelines

See mean speed (7.2). However, maximum speed can reveal occasional severe loss of speed monitoring. Occasional speeding behaviour can be caused by visual and cognitive distraction.

7.3.4 Limitations

Since this metric is based on one sample – the maximum value – this metric is less reliable than e.g. mean speed, which is based on all samples.

7.4 Mean Lateral Position

7.4.1 Definition

Mean lateral position is defined as the average distance between the right side of the *front or rear* right wheel and the inner (closest) edge of the right hand lane marking. Please, pay attention to this definition to guarantee the comparability of data collected in different experiments. Lateral position (LP) should be measured perpendicular to the lane marking. If this is not possible, make an estimation of the effects on the collected data. Note! It should be noted which of the two wheels that is used for LP measurement (front or rear) when the metric is used. Left-hand wheel and left-hand lane marking are used in the UK. In case of several lanes (e.g. motorways), the lane markings of the current lane should be used. Mean lateral position is not defined during lane changes.

7.4.2 Description of use

Mean lateral position can only be used when there are well defined lane boundaries and lane markings. Urban driving including junctions and parking lots is thus not suitable for this metric. Rural road and motorway are however suitable. Please note that lane shifting on motorways and take-overs will cause variations in lateral position not caused by distraction! These data should be excluded. This is done by manually identifying these situations, and then remove data from 200 m before any part being over the centre line to 200 m after the vehicle leaving the centre line. Field collected data in particular has to be inspected before used for any LP calculations. Any errant data must be excluded or repaired.

7.4.3 Interpretation guidelines

A change in mean lateral position indicates a shift of goal in the regulating level; during distraction the driver targets at driving more towards a safer lateral position. This may either be the left or the right side, depending on traffic intensity, emergency shoulder width etc. This metric has to be interpreted with the current driving scenario in mind.

Control for any shifts in lane in the analysis.

7.4.4 Limitations

Not suitable for urban driving. Shifts in lane width will affect this metric, this is why road sections of constant width should be used.

7.5 Modified Lateral Position Variation (MSDLP)

7.5.1 Definition

Modified lateral position variation is defined as standard deviation of 0.1 Hz high pass filtered lateral position data, where lateral position is defined as described under Mean Lateral Position. This metric will in short be called MSDLP (a Modified way to calculate SDLP). Modified lateral position variation is only defined for data sets longer than 10 seconds!

As for mean lateral position, modified lateral position variation is not defined during lane changes.

7.5.2 Description of use

See Mean Lateral Position. Road curvature should be controlled for since it influences lateral position variation.

The reason for the high pass filtering before lateral position variation calculation is that this metric otherwise becomes highly correlated with data length (see 3.3, *page 37*). The high pass filtering of data removes any variations of longer periods than 10 seconds. A second order Butterworth-filter (or equivalent) with cut off frequency 0.1 Hz should be used. Make sure that the filtering does not cause any spikes in the beginning or end of data. Data recording interruptions may cause spikes in the filtered data. Make sure this does not happen.

Any errant data should be excluded, of course. Field collected data in particular has to be inspected before used for any LP calculations. If the data is noisy, apply a suitable filter to the LP-data before calculating lateral position variation. A low pass filter with 3-5 Hz cut off frequency can be a good starting point (does not remove effects of lateral control).

7.5.3 Interpretation guidelines

Increased lateral position variation is often an effect of visual distraction, but may also occur when heavily loaded by secondary cognitive tasks. This is often found in conjunction with increased steering effort. However, decreased lateral position variation is more commonly found during cognitive workload, which also is found in conjunction with concentrated forward gaze. This is however not necessarily a good thing since the attention to peripheral information is reduced and there will be less mental resources left for the driving task and other secondary tasks.

An ADAS lane keeping support system can be expected to reduce lateral position variation.

7.5.4 Limitations

Not suitable for urban driving. Shifts in lane width will affect this metric. Not applicable on data of less duration than 10 seconds.

7.6 Mean TLC – Mean Time-To-Line crossing

7.6.1 Definition

Mean TLC is defined as the mean of the TLC local minima, where 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. TLC metrics are only defined if the vehicle is within a lane. Unit: Seconds.

7.6.2 Description of use

Before identifying any min TLC values, TLC itself has to be calculated. We however use an approximation. Calculate TLC with the procedure described in Appendix 2. Please note that poor lateral position data will create even worse TLC data.

7.6.3 Interpretation guidelines

Decreased mean TLC indicates a decreased lateral control performance on either a regulatory or tracking level. Visual distraction and heavy cognitive load lead to decreased mean TLC.

If frequent line crossings are found, the lane markings may not represent the lane boundaries to the driver. If so, mean TLC is not a valid metric of driving performance.

7.6.4 Limitations

TLC metrics can only be used when there are well defined lane boundaries and lane markings. Urban driving including junctions and parking lots is thus not suitable for this metric. Rural road and motorway are however suitable.

Not valid if the driver often chooses to cut in curves or by other reasons chooses to cross the lane markings.

If mean TLC is measured during short periods of time (less than 10 seconds), there is a great risk that there are no valid TLC waveforms, and you will get no mean TLC. TLC metrics require rather long data to produce reliable data.

7.7 Line Crossings

7.7.1 Definition

Number of line crossings per kilometre. A line crossing is defined as any part of any wheel touching any lane marking. Voluntary line crossings (lane changes and take-overs) are excluded from data.

7.7.2 Description of use

If the lateral position becomes negative (out on the right side, or left in the UK) or larger than *lane width – vehicle width*, this is considered a line crossing. In order to exclude the effects of take-overs and lane changes, you should make sure that you can identify these occasions, either in data, in video recordings or in protocols filled in during driving by an accompanying observer. It should be possible to use lateral position data for this metric.

7.7.3 Interpretation guidelines

Line crossings normally do not occur often, but if they occur, they may be an indicator of very high visual and/or cognitive distraction. Other effects preceding line crossings are increased lateral position variation, decreased mean TLC and possibly increased steering reversal rate (and similar metrics). It is very likely that this metric will not produce any significant effects in statistical analyses (as was found for the similar LANEX metric in the HASTE project), but studying the situations in detail when lines were crossed may give valuable qualitative data on distraction effects. It is highly risky to unintentionally travel outside the lane boundaries.

7.7.4 Limitations

Line crossings occur very seldom if the lane markings do not represent the lane boundaries to the driver. If they occur often, the metric may not reflect high distraction, but rather voluntary shifts in lateral position.

7.8 Minimum Time Headway

7.8.1 Definition

Minimum Time Headway is defined as the minimum time gap (seconds) to a lead vehicle, travelling in the experimental vehicle’s path of travel.

7.8.2 Description of use

Headway is calculated as the distance to the lead vehicle (bumper to bumper) divided by the experimental vehicle travel speed.

In field experiments the headway sensor defines if a vehicle or object is detected. Road scenarios should be chosen so that the risk of the sensor locking on vehicles and objects others than in the travel path is avoided. Knowing what the sensor locks on is the greatest problem concerning field collected headway data. This problem can be reduced by video recording the drive. More reliable data are achieved if the sensor range is reduced – the closer objects, the more reliable data.

In simulator experiments, the definition of what vehicles and objects to detect is entirely up to the experiment designer. If driving performance is evaluated, independently of any prototype sensor, headway should be measured only to vehicles travelling in the same lane. Overtakings should be excluded from headway data. If a prototype is evaluated, the sensor should be simulated as correctly as possible.

7.8.3 Interpretation guidelines

Small headways are related to high risk of collision. Headways larger than 3 seconds can be considered safe and are of little interest. Headways less than 1 second can be considered unsafe, but the subjective estimation of safe headway varies a lot between drivers. Decreased and increased headway may reflect loss of situation awareness if the driver is engaged with a distracting (visual or cognitive) task. Increased headway may also indicate that the driver chooses to increase the distance to the lead vehicle in order to compensate for increased distraction/cognitive load.

7.8.4 Limitations

Minimum time headway requires a lead vehicle. If different lead vehicles drive at different, speeds there is a risk that time headway is influenced. There may be difficulties evaluating the quality and accuracy of field collected data.

7.9 Mean Time Headway

7.9.1 Definition

Mean Time Headway is defined as the mean value of the time gap (seconds) to a lead vehicle, travelling the own vehicle's path of travel.

7.9.2 Description of use

See Minimum Time Headway. There is however an additional issue that has to be addressed. The car-following situations that are used for headway measurement have to be steady state, i.e. the initiation and ending of the situations must be identified and excluded from data. The definition of steady state car following situation is not made here.

7.9.3 Interpretation guidelines

See Minimum Time Headway

7.9.4 Limitations

See Minimum Time Headway

7.10 Brake reaction time

7.10.1 Definition

Brake reaction time is defined as the time from safety critical event onset to brake onset, indicated by onset of brake lights.

7.10.2 Description of use

Brake reaction time is intended only for safety critical situations requiring very quick brake reaction, typically within two seconds, to avoid incident or crash. In simulator experiments, the onset of events should be defined as onset of lead vehicle brake lights, onset of turning indicator of vehicle cutting into the own travel path etc. Keep in mind that very small variations in scenario will cause brake reaction time variations. Consider using identical events for different subjects (between subject design) rather than similar events for the same subject (within subject design). Brake reaction time is very difficult to use in field experiments. If included, however, you will need camera based detection of events in the traffic environment.

7.10.3 Interpretation guidelines

Brake reaction time to such obstacles and sudden firm braking of a lead vehicle is a straightforward metric of driving performance on a regulating or monitoring level. Visual distraction leads to the driver not being able to regulate speed properly – does not see and react fast. Cognitive workload leads to decreased situation awareness and less performance in evaluating the state and behaviour of other vehicles – does not interpret a vehicle or object as causing a risk of collision.

7.10.4 Limitations

Brake reaction time is only practical in simulator experiments, not in field experiments. It is very sensitive to scenario variations. Further, if an event causes a crash (in simulator), this may affect the realism of the simulation and even make the participant upset and unsuitable to continue driving.

7.11 Brake jerks

7.11.1 Definition

The Brake jerks metric is a binary metric, which is *yes* or =1 if there was one or several abrupt onsets of the brakes during driving, and *no* or=0 otherwise. Abrupt onset of the brakes is defined as the occurrence of a deceleration change higher than 10 m/s^3, induced by braking.

7.11.2 Description of use

This metric can be used in field and simulator experiments. In field experiments, it may capture the effects of distraction in real unpredicted hazardous situations, which is of high value for the assessment of ADAS/IVIS effects on driving performance and risk.

Deceleration change rate is achieved by measuring and calculating the change in deceleration, which can be measured by an accelerometer. Using speed for calculating this metric is the second choice. Make sure that speed data has such precision that it can be derived twice without adding noise more than +/- 2m/s^3. Low pass filtering the speed data is recommended.

7.11.3 Interpretation guidelines

As with line crossings, analysis of situations where brake jerks are found may give valuable understanding of the effects of distraction and cognitive load on the driver. Brake jerks of the specified amplitude only occur in very hazardous situations and thus indicate high risk of accident. In the analysis, it should be assessed what caused the brake jerk. Only brake jerks that would have been avoided if the driver was not distracted or under cognitive workload indicate that driving performance was deteriorated.

7.11.4 Limitations

Since it is a binary metric, and since there most likely will be very few abrupt brake onsets, this metric will seldom produce significant results.

7.12 Steering Wheel Reversal Rate

7.12.1 Definition

The *steering wheel reversal rate* is defined as the number, per minute, of *steering wheel reversals* larger than a certain minimum angular value, referred to here as the *gap size*. The idea of steering wheel reversals is intuitively simple to grasp, but a rigorous explicit definition is not entirely straightforward. Here we content ourselves with first giving an approximate definition, and then in practice letting the metric calculation algorithm operationally define what is considered as a reversal.

Roughly, given a steering wheel angle signal $\theta(t)$, a steering wheel reversal is taken to be a portion $[t_1; t_2]$ of the signal such that θ is stationary at both t_1 and t_2 (i.e. $d\theta(t_1)/dt = 0$ and $d\theta(t_2)/dt = 0$), and such that $|\theta(t_1) - \theta(t_2)| \geq \theta_{min}$, where θ_{min} is the gap size. Reversals can not be overlapping and are chosen so as to fit as many as possible of them into the interval on which θ is defined. Note that a stationary point can be a local minimum, a local maximum, a point of inflection, or a point on a constant segment of the steering wheel angle signal.

Figure 66 shows an example of a steering wheel angle signal, with calculated reversals for a gap size (θ_{min}) of 1°.

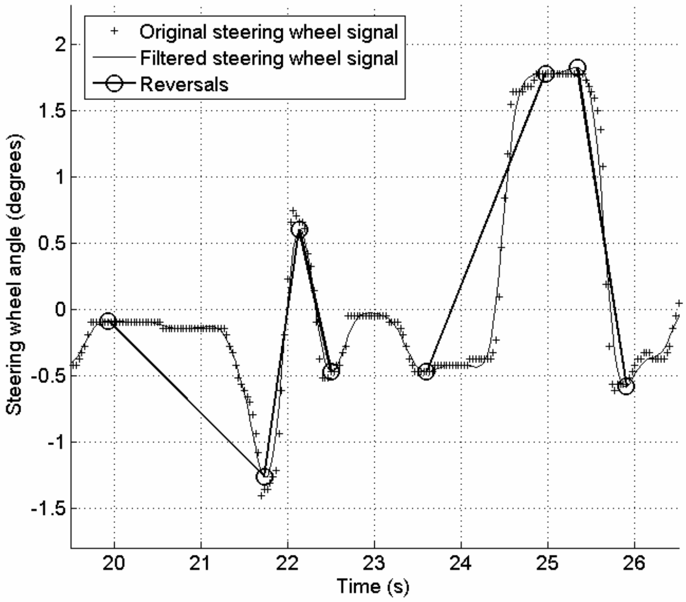


Figure 66 Example of steering wheel reversals. Low pass filtering is performed to facilitate localization of stationary points of the steering wheel angle signal. When the difference in

angle between two stationary points exceeds the gap size (one degree, in this figure) a reversal is counted. Reversals are marked as lines extending between their respective starting and ending stationary points, marked with circles.

Below, the different steps of the prescribed method of reversal rate calculation are described in detail.

1. *Low pass filtering.* A low pass second order Butterworth filter with cut-off frequency *0.6 Hz* is applied. The filter reduces high-frequency noise in the steering wheel angle signal, and makes it possible to find stationary points using the method described below.

2. *Finding stationary points.* Let θ_i be the value of the low pass filtered steering wheel angle signal at time step i, with $i \in \{1, 2, 3, \ldots, T\}$, where T is the total number of samples in a measurement. We calculate the following quantity:

$$\theta'_i = \begin{cases} 0 & i = 1 \\ \theta_i - \theta_{i-1} & i > 1 \end{cases}$$

θ'_i is a scaled version of $\theta'_i / \Delta t$, an approximation to the first order derivative of the steering wheel signal at time step i. Δt is the difference in time between two time steps, but we don't need to include it in order to find the stationary points. We instead use θ'_i directly, and find all i such that either:

$$\theta'_i = 0 \qquad 2 \le i \le T \qquad (1)$$

or:

$$\left| sign(\theta'_i) - sign(\theta'_{i+1}) \right| = 2 \qquad 1 \le i \le T-1 \qquad (2)$$

where we have defined:

$$sign(x) = \begin{cases} -1 & x < 0 \\ 0 & x = 0 \\ 1 & x > 0 \end{cases}$$

Any i satisfying equation 1 or 2 is thus a position in the steering wheel angle signal where the approximate first-order derivative of the steering wheel angle is either zero (equation 1), or just about to pass zero (equation 2). We thus take any such point to be a stationary point. This procedure is illustrated in Figure 67.

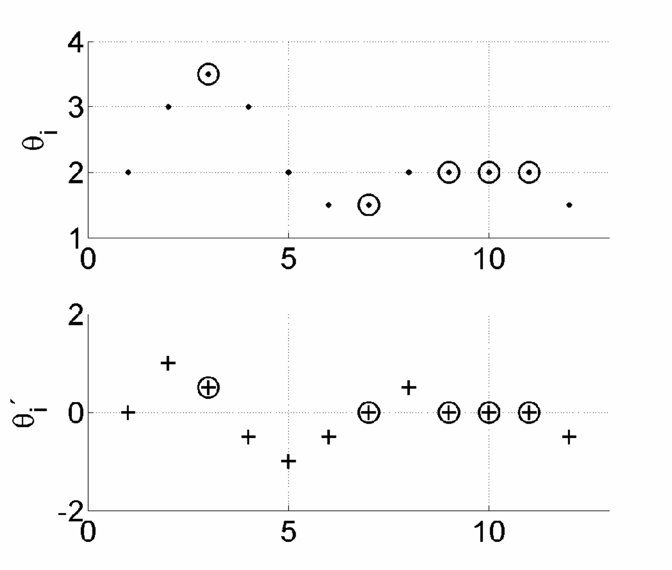


Figure 67 An illustration of the method for finding stationary points of the steering wheel angle signal. An example signal θ_i is plotted in the top graph, and corresponding values of θ'_i are plotted in the bottom graph. i = 3 satisfies equation 2, and i $\in$ {7, 9, 10, 11} satisfies equation 1, so all i $\in$ {3, 7, 9, 10, 11} are stationary points of the steering wheel angle signal.

3. *Finding reversals.* Let $e(k)$ be the kth value of i such that i is a stationary point, sorted in time order so that $e(k) > e(l)$ if $k > l$. For the example of Figure 67, we thus have $e(1) = 3$, $e(2) = 7$, $e(3) = 9$, $e(4) = 10$, and $e(5) = 11$. Let N be the total number of stationary points. Then the following algorithm counts all "upwards" reversals (from a stationary point of lower angle value to one of higher angle value, e.g. from a local minimum to a local maximum) in the steering wheel angle signal that are bigger than the gap size threshold θ_{min}, which is set to 3 degrees.

1. $k \leftarrow 1$
2. $N_r \leftarrow 0$
3. For each l in [2, 3, 4, …, N]
   a. If $\theta_{e(l)} - \theta_{e(k)} \geq \theta_{min}$:
      i. $N_r \leftarrow N_r + 1$
      ii. $R(N_r) \leftarrow [e(k), e(l)]$
      iii. $k \leftarrow l$
   b. Else if $\theta_{e(l)} < \theta_{e(k)}$:
      i. $k \leftarrow l$

This algorithm positions k at the first stationary point (k=1), and then iterates through the subsequent stationary points until either a stationary point l is found that is more than θ_{min} bigger in angle value than the stationary point at k, or a stationary point l is found that is smaller in angle value than the stationary point at k. In the first case an "upwards" reversal has been found. In either case, k is set to l and the iteration is continued. Setting k to l in the latter case, when l is a stationary point with smaller angle value than k, ensures that an "upwards" reversal will be found as soon as possible, since this will require a lower

When the algorithm above has terminated, N_r is the number of “upwards” reversals, and $R(m)$ is a vector with two elements where the first is the time step where the mth reversal begins, and the second element is the time step where it ends. $R(m)$ is useful for visualizing the results of the algorithm, as in Figure 67, but if this is not needed step 3.a.ii of the algorithm can be omitted.

To count also the “downwards” reversals, the same algorithm is then applied on the negative of the steering wheel angle, $-\theta_i$, instead of on θ_i, and the total number of reversals in the steering wheel angle signal is obtained as the sum of “upwards” and “downwards” reversals.

4. Calculating the reversal rate. The steering wheel reversal rate is finally calculated as the total number of reversals detected in the steering wheel angle signal, divided by this signal’s total length in minutes.

7.12.2 Interpretation guidelines

For visual tasks, SRR represents the control effort needed to cope with visual time sharing induced by a secondary task, and thus provides a direct measure of the consequences of visual demand on lateral control. Since, in this case, SRR is directly related to visual distraction, increased SRR could be interpreted in terms of increased risk. However, the picture is complicated by the fact that visual time sharing is normally accompanied by a reduction of speed (which increases the total time available for the tracking loop and thus reduces the risk), and the actual safety consequences depends on the degree to which the driver succeeds in compensating for the increased risk. The metric should thus be interpreted together with other performance metrics (especially lane keeping variance and speed).

As discussed in section 3.8.8, the relation between cognitive load and steering performance is still not entirely understood. Thus, further studies are needed to assess the different possible explanations (including those outlined in 3.8.8) before clearcut interpretation guidelines can be provided.

Like other steering wheel metrics, SRR is strongly influenced by both primary- and secondary task load. Examples of primary task-related factors influencing steering wheel movements are speed, lane width, road surface friction and curvature. Hence, if steering wheel metrics are to be used for IVIS evaluation, it is important to distinguish the variance of interest (i.e. that induced by the secondary task) from the variance associated with changes in primary task demand (i.e. the driving scenario).

7.12.3 Description of use

With respect to the AIDE evaluation methodology, steering wheel reversal rate can mainly be recommended for evaluation of the effects of visually demanding IVIS on driving performance. The metric is sensitive to cognitively loading tasks as well, but due to the lack of a clear scientific interpretation, it cannot (yet) be recommended for IVIS evaluation applications. However, the metric is still very useful for further basic research on the effects of cognitive load.

The steering wheel reversal rate metric can be used in both simulator and field experiments, and on all types of roads, except built-up areas. The effects are generally larger in simulators than in the field, where fixed-base simulators seem to yield stronger effects than moving-base simulators. The implementation is straightforward, albeit relatively complex, and requires Matlab or some equivalent computation tool.

Based on the present results, two versions of the Reversal Rate metric could be proposed, defined in terms of the parameter setting:

1. LargeReversals: {Gap size= 3, LPF cut-off=0.6 Hz)
2. SmallReversals: {Gap size=0.1, LPF cut-off=2 Hz)

If the purpose is to measure the effects of an IVIS that impose any type of visual time sharing, the LargeReversals version should be used. However, if the task is purely cognitive (i.e. it requires no visual time-sharing), the SmallReversals versions should be used.

However, due to differences between test settings, the optimal gap size for visual load will vary somewhat (roughly between 2-6 degrees for passenger cars). If possible, it is recommended to identify a suitable value for the gap size the first time the metric is used in a specific simulator or test vehicle. This is absolutely necessary if the metric should be used for heavy trucks (the present derivation was used based on passenger vehicle data only).

Moreover, it should be pointed out that the spatial and temporal resolutions used in the present study (0.1 degrees, and 0.033 seconds respectively) were higher than what is specified in the AIDE Requirements deliverable (D3.2.1, Kussmann et al., 2005). In AIDE deliverable the required resolutions for demonstrator evaluation are 0.2 degrees and 0.1 seconds respectively. These requirements may be sufficient, although this cannot be guaranteed from the present results. Especially, the measurement of cognitive load could be expected to require data with higher resolution.

7.12.4 Limitations

Like all steering wheel activity metrics, steering wheel reversal rate is not suited for use in built-up areas, due to the large variation induced by the road geometry. Another main limitation is the yet unclear interpretation of the effects of cognitive load.

8 Recommendations for the Lane Change Test Usage

Three different investigations of the LCT were carried out within T2.2.5. One experiment (chapter 4, Origins of Workload) was aimed at a way to develop new metrics for the LCT analysis in order to identify the origin of workload. It was concluded that this is not easily achievable within the standard paradigm of LCT. Furthermore, the investigation of the modified LCT (chapter 5, Modified Lane Change Test for Assessing Adaptive IVIS Interfaces) showed that it seems inappropriate at the moment for evaluation of AIDE metafunctions. In the third experiment (chapter 6, Lane Change Test – Effects of Scenarios and Simulators), it was concluded that the LCT seems to be capable to measure the distraction potential of integrated systems as well as of nomadic devices. However, it has to be assured that LCT experiments are set up to guarantee consistency between test sites and also consider the new ISO draft instructions to reduce variance by paying more attention to training of the subject. Reports of results should include a specification setup used including type of steering wheel.

It was concluded that the recommendation for LCT is to use the standard or “classic” version of the LCT for both formative and summative evaluation in the AIDE project, specifically for IVIS applications. Furthermore, it seems feasible to use the LCT in different setups ranging from PC versions, to advance driving simulators or even real field conditions.

9 Discussion and Conclusions

9.1 Mechanisms

In this document, and particular in chapters 3 and 7, driving control metrics have been investigated and specified. During this work, it has become clear that there are mainly two mechanisms that we identify as effects of distraction; Visual distraction leads to deteriorated lateral control, and when severe, also to deteriorated speed adaptation. However, speed decrease is often found as a behavioural adaptation to the increased visual load. These effects could mainly be attributed to changes in compensatory control in the tracking layer of the ECOM model (see *Figure 3*). Cognitive load leads to deteriorated event detection and interaction with other road users, mainly in the monitoring and regulating layers. Thus, it has more to do with impaired anticipatory control. It has been found that there are more metrics sensitive for the effects of visual distraction than cognitive distraction.

9.2 Scenarios and metrics

Several metrics can be used to identify the effects of visual and cognitive distraction on driving behaviour. These metrics are generally based on continuous driving on rural road, motorway or similar, where there is a minimum of lane changes, take-overs intersections and interfering vehicles. Metrics based on instantaneous reactions are less feasible for the identification of distraction effects on driving performance. Also, lane changes, intersections, take-overs and urban driving generally cause variations in data that are difficult to control for – at least in field experiments. The driver's intentions can be obscured in more complex situations and thus it will be more difficult to interpret behaviour deviations in terms of workload and distraction.

In *D2.1.3 Considerations on Test Scenarios* the following can be read:

> As decreased driving skills are likely to become salient in complex situations, scenarios including demanding interactions with other road users should have highest priority in the case of older drivers.

Among such situations, there are merging, yielding right of way, over-taking, emergency braking, turning on a narrow lane etc. Driving performance in such situations may surely be affected by IVIS and ADAS interaction and are of high relevance. However, driving performance in these situations is very difficult to evaluate since there are few metrics that can be used. Also, due to the nature of these situations, metrics are based on instantaneous reactions rather than prolonged driving, resulting in low data reliability – caused by the fact that short measurements cannot reflect prolonged behaviour. Rather, these situations should be studied qualitatively in detail. Applicable metrics are however brake reaction time, brake jerks and minimum time headway.

In field experiments the control over the effects of the environment and other road users is poor. To be able to answer such questions as "why is the steering activity so high in this situation?" and "why is the speed so low in this condition?" the route must be chosen to enhance the chances to get reliable answers. The aim should be to eliminate any confounding effects. The best speed and lateral position data, including a minimum of discontinuities, is collected on one lane roads with a minimum of intersections. In urban environments, most measures are not useful. There are some suggested metrics that are designed to indicate possibly hazardous situations. Those measures are minimum time headway, brake jerks, line crossings. If safety critical values are found in these metrics, the identified situations should be examined qualitatively in order to search for any undesired effects from IVIS/ADAS under these conditions.

9.3 Task length and metrics

Modified standard deviation of lateral position (MSDLP) requires a minimum data length in order to be defined, and some metrics become unreliable for short data durations (e.g. line crossings and steering reversal rate). Thus, there is a risk that MSDLP becomes undefined for a subset of the experimental population in an IVIS/ADAS evaluation, resulting in a biased data set where e.g. only the worst performing IVIS/ADAS users are represented. The source of this problem is not the metrics as such. It is rather the fact that certain aspects of driving behaviour cannot be reflected within too short time frames. This applies to all variation based measures; variation is strongly linked to time.

There are two ways to handle the described problem of biased population. Either an analysis of the effect of missing data is made, or MSDLP is excluded from the analysis. This problem does however need to be addressed when it occurs.

9.4 The Lane Change Test

The investigation of the modified LCT showed that it seems inappropriate at the moment for evaluation of AIDE metafunctions. Thus, it was concluded that the recommendation for LCT is to use the standard or “classic” version of the LCT for both formative and summative evaluation in the AIDE project, specifically for IVIS applications. Furthermore, it seems feasible to use the LCT in different setups ranging from PC versions, to advance driving simulators or even real field conditions. Thus, it will not be further discussed in the examples below as there are no specific considerations to account for.

9.5 Examples of metric applications

In this section, two examples of metrics selection is given, based on the evaluation use cases in *2.5.2 Formative and summative evaluation*.

9.5.1 Metrics in formative evaluation, example 1

The purpose of formative evaluation is to compare different design solutions in order to improve a system design. This type of evaluation will most likely be carried out in a driving simulator or lab setting. The systems to be compared and evaluated are assumed to be two different traffic information systems. We assume that the chosen driving scenario is driving on rural road with constant speed limit, specifically defined curvature and random oncoming traffic. No traffic in the same lane. At two occasions an object appears in the road requiring a brake reaction within two seconds to avoid a crash. The experimental factors are IVIS and straight/curved road. The hypothesis is that one system will be better than the other one, and that the difference is greater on a curvy road than on straight road sections. It is further expected that one system generates worse reaction times for the critical events. Note that if you do not cater for road geometry you will find confounding effects in steering reversal rate and lateral position variation of road curvature.

For this experiment, mean speed, maximum speed, mean lateral position, modified lateral position variation, line crossings and reversal rate should be recorded and analysed for the “normal” driving conditions. For the critical events, brake reaction time and minimum headway should be used.

9.5.2 Metrics in formative evaluation, example 2

Suppose the traffic information system in the previous example only is designed for urban driving. If so, the scenario may have to include turning in intersections and possibly several situations where the driver have to adapt the speed to lead vehicles and avoid other road users cutting in. Applicable measures are brake reaction time, brake jerks, minimum headway and, if there are prolonged car following situations, mean headway. Speed and lateral position measures are most likely not applicable. Steering reversal rate may be applicable if situations of turning in intersections and evasive manoeuvres are excluded from data.

9.5.3 Metrics in summative evaluation, example 1

The purpose of the AIDE summative evaluations is to evaluate the AIDE prototypes in real traffic. If the scenario is the same as in example 1, formative evaluation, the same metrics can be used in this example. Care should however be taken considering the reliability and control of the headway and lateral position measures. Sensor problems may result in the absence of valuable data. Road curvature and road width will influence steering reversal rate and the lateral position metrics. The route should therefore be chosen so that any effects of road curvature and width are eliminated or balanced. It is recommended that the modified lateral position variation and steering reversal rate is used also in this case.

9.5.4 Metrics in summative evaluation, example 2

If the scenario is urban driving including several complex traffic situations, speed and most lateral position metrics will not provide any relevant data. Reversal rate may be used in road sections where there are no turnings or lane changes. The most useful measures will however be minimum headway to lead vehicles and brake-jerks, which reflect the driving behaviour in conflict situations. Brake reaction time is very difficult to measure automatically in field experiments because there is no control of what triggers the brake reaction. There are most likely no lines to which lateral position can be measured – and there will most likely not be any longer continuous uniform driving anyway required by lateral position metrics. Speed can be measured, but is rather influenced by restrictions imposed by the road environment and other road users rather than by behavioural adaptation or loss of speed monitoring. If, however, there are longer sections of uniform driving, without turnings, stops and interfering vehicles, speed and possibly lateral position metrics can be analysed.

9.6 Conclusions

There are a number of driving control metrics that are highly feasible for the evaluation of IVIS and ADAS. However, care should be taken in experimental design, recording and analysis of these metrics. Also IVI/ADAS task length will influence the usefulness of driving control metrics. These metrics are rather easy to measure, most even in real traffic. The applicability of the metrics are however highly dependent on driving scenario. In complex scenarios, including e.g. conflict situations, there are few and not very reliable metrics of driving performance that will provide interpretable data; too many factors influence driving behaviour, few data points are collected and drivers may react very differently to unexpected events if they are not sufficiently well designed. In uniform, conflict free driving on rural road or motorway, there are much more to be measured and analysed; influencing factors are much less and there is also more data to analyse (e.g. several minutes of data compared to one brake reaction).

10 Innovation

While the present work has not resulted in any brand new driving performance metrics, it has still provided many innovative results. First, a generally agreed conceptual framework for describing driving performance does not exists and Chapter 2 of the present report represent the first comprehensive attempt to apply such a framework (based on the COCOM/ECOM model) to the development of driving performance metrics.

Second, a number of previously unresolved issues were addressed and new solutions proposed. In particular, the issue of task duration for variation metrics, which has not been comprehensively treated in the literature, was addressed and a working solution was proposed.

Moreover, the analysis of steering wheel metrics revealed the strong differences between visual and cognitive load on steering performance. This is a truly new result that has not been demonstrated previously. Although there were some indications in this direction in the HASTE results (Östlund et al., 2004), it remained speculative since the metrics HASTE were not sufficiently tuned to detect this difference. A further key innovation of the present work was the development of a new reversal rate metric able to capture this effect. The metric was demonstrated to exhibit superior sensitivity compared to the metrics used in HASTE.

The LCT work involves innovative aspects, both in terms of further development of the method (e.g. the investigation of new metrics for capturing differences between visual/cognitive load, and the effect of simulator grade) and the application of the LCT to new areas (in particular its application to the evaluation of adaptive interfaces).

1 Contribution to overall AIDE objectives

The main concrete expected result of the AIDE project is the demonstration of the Adaptive Integrated Driver-vehicle InterfaceE concept in a set of three validated prototype vehicles. The human factors validation will take place by the end of the project, employing the AIDE Evaluation Methodology, which is the main expected result of SP2. The results from the present task (T2.2.5) provide an essential component of this methodology, i.e. metrics of driving performance. The detailed specifications of metrics provided in this deliverable are expected to be used in the subsequent experimental work towards the final definition of the AIDE Evaluation Methodology.

With respect to the work plan, the results from this task, and the other tasks from T2.2.2-T2.2.7, feeds directly into T2.2.7 where the different metrics and tools will be empirically compared (D2.2.7, due month 26) before being further passed on to the general synthesis of the different components into the general test regime in WP2.1 (T2.1.4, due month 36).

11 References

Allen, T. M., Lunefeld, H., & Alexander, G. J. (1971). Driver Information needs. Highway Research Record, 102-115.

Amditis, A., Bolovinou, A., Engstrom, J., Kussmann, H., Placke, L., Bekiaris, E., Panou, M., Gaitanidou, E, Andreone, L., Deregibus, E., Kompfner, P. and Robertson, P. (2004) AIDE scenarios and use cases definition. EU project AIDE, project IST-1-507674-IP, Deliverable 3.1.2

Antin, J.F., Dingus, T. A., Hulse, M.C. and Wierwille, W.W. (1990) An evaluation of the effectiveness of an automobile moving-map navigation display. *International Journal of Man-Machine Studies*. 33, 581-594.

Bertalanffy, L. V. (1968). General Systems Theory. New York: George Braziller.

Blaauw, G.J., Godthelp, H., and Moraal, J. 1977. Driver's lateral control strategy as affected by task demands and driving experience. Society of Automotive Engineers, SAE Paper 770876.

Bloch, H. et al. (1999). *Grand Dictionnaire de la Psychologie*. Paris: Larousse.

Brookhuis, K., de Ward, D. and Mulder, B. (1994). Measuring performance by car-following in traffic. Ergonomics, 1994, Vo. 37, No 3, 427-434.

Brookhuis, K. A., de Vries, G., & de Ward, D. (1991). The effects of mobile telephoning on driving performance. *Accident Analysis and Prevention*, 23(4), 309–316.

Brooks, R.A. (1986) A robust layered control system for a mobile robot. *IEEE Journal of Robotics and Automation*, 2: 14-23.

Boer, E. (2000) Behavioural Entropy as an Index of Workload. Proceedings of the IEA 2000/HFES 2000 Congress.

Boer, E., Rakauskas, M., Ward, N.J., and Goodrich, M.A. (2005) Steering Entropy Revisited. *Proceedings of the third International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design*. Rockport, Maine.

Burns, P.C. (2001) Behavioural Adaptation to an Advanced Driver Support System. Volvo Technology Corporation, Internal Report.

Byers, J. C., Bittner, A. C., & Hill, S. G. (1989). Traditional and Raw Task Load Index (TLX) Correlations: Are paired comparisons necessary? Paper presented at the Advances in Industrial Ergonomics and Safety I.

Carver, C. S., & Schreier, M. F. (1982). Control Theory: A Useful Conceptual Framework for Personality-Social, Clinical and Health Psychology. Psychological Bulletin, 92(4), 111 - 135.

Damasio, A. (1994) Descarte's Error: Emotion, Reason, and the Human Brain. New York: G.P. Putnam's & Sons.

de Waard, D. (1996) *The Measurement of Drivers' Mental Workload*. ISBN 90-6807-308-7. Traffic Research Centre. University of Groningen.

Donges, E. (1978) A Two-Level Model of Driver Steering Behaviour. *Human Factors*, 20(6), 691-707

Engström, J. (2005) Interaction Plan. AIDE Deliverable 4.0.1. EU project AIDE, project IST-1-507674-IP, Deliverable 4.0.1

Engström, J., & Hollnagel, E. (2005) Towards a conceptual framework for modelling drivers' interaction with in-vehicle systems. Paper presented at the EUR 21677 - The International Workshop on Modelling Driver Behaviour in Automotive Environments, Ispra, Italy.

Engström, J., Johansson, E and Östlund, J. (2005) Effects of visual and cognitive load in real and simulated motorway driving. *Transportation Research Part F*, (8), pp. 97-120.

Evans, L. (1983). *Traffic safety and the driver*. New York: van Nostrand Reinold.

Floudas, N., Amditis, A., Keinath, A., Bengler, K., Engeln, A. (2004) Review and Taxonomy of IVIS/ADAS applications, EU project AIDE, project IST-1-507674-IP, Deliverable 2.1.2

Fosser, S., Saetermo, I. F. and Sagberg, F. (1997). An Investigaton of behavioural adaptation to airbags and antilock brakes among taxi drivers. *Accident Analyis and Prevention*. 29 (3).

Fuller, R. G. C. (1984). A conceptualisation of driving behaviour as threat avoidance. Ergonomics, 27, 1139 - 1155.

Fuller, R. (2005) Towards a general theory of driver behaviour. *Accident Analysis and Prevention*. 37, pp 461-472.

Gibson, J.J. (1979). The Ecological Approach to Visual Perception. Boston: Houghton Mifflin.

Gibson, J. J., & Crooks, L. E. (1938). A theoretical field-analysis of automobile-driving. The American Journal of Psychology, 51(3), 453 – 471

Godthelp, H., & Konings, H. (1981) *Levels of Steering Control; Some notes on the Time-To-Line Crossing concept.* Paper presented at the Proceedings of the first European Annual Conference on Human Decision and Manual Control, Technical University, Delft, Netherlands.

Godthelp, H., Milgram, P. and Blaauw, J. (1984) The Development of a Time-related Measure to Describe Driving Strategy. Human Factors, 26(3), 257-268.

Green, P. (1993, revised March 1994, August 1994, final version in 1995) Measures and Methods Used to Assess the Safety and Usability of Driver Information Systems. Technical Report UMTRI-93-12. FHWA-RD-94-088

Greenberg, J., Tijerina, L., Curry, R., Artz, B., Cathey, L., Grant, P., Kochlar, D., Kozak, K., and Blommer, M. (2003) Evaluation of driver distraction using an event detection paradigm. Journal of the Transportation Research Board, No 1843. Washington D.C.: TRB.

Greenshields, B.D. (1963). Driver behaviour and related problems. Highway research record. 25, 14-32.

Hallen (2004). Presentation given on ISO WG8 LCT task force meeting Munich, 21. january 2005.

Harbluk, J.L., Noy, Y.I. (2002) *The Impact of Cognitive Distraction on Driver Behaviour and Vehicle Control*. Report no. TP13889E, Transport Canada, Ontario, Canada

Hart, S. G., & Staveland, L. E. (1988). Development of NASA - TLX (Task Load Index): Results of Empirical and Theoretical Research. In P. A. Hancock & N. Meshkati (Eds.), Human Mental Workload (pp. 139 - 183). Amsterdam: Elsevier Science Publisher B.V.

Hildreth, E.C., Beusmans, J.M.H., Boer, E.R. and Royden, C.S. (2000) From vision to action: Experiments and models of steering control during driving. *Journal of Experimental Psychology: Human Perception and Performance*, 26, 1106-1132.

Hollnagel, E., Nåbo, A and Lau, I.V (2003) A systemic model for driver-in-control. In *Proceedings of the Second International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design*. Park City Utah, July 21-24.

Hollnagel, E. & Woods, D. D. (2005) Joint cognitive systems: Foundations of cognitive systems engineering. Boca Raton, FL: Taylor & Francis/CRC Press.

Horrey, W. J. and Wickens, C. D. (2004) *The impact of cell phone conversations on driving: A meta-analytic approach* (Technical Report AHFD-04-2/GM-04-1) Warren, MI: General Motors Cooperation.

Hjälmdahl, M., and Varhelyi, A. (2004) Speed Regulation by in-car active accelerator pedal – effects on driver behaviour. *Transportation Research Part F*, Vol. 7(2), pp. 77-94.

ISO (2005). Road vehicles - Ergonomic aspects of transport information and control systems - Simulated lane change test to assess driver distraction. ISO/AWI 26022 (Draft). ISO/TC22/SC13/WG8

Jagacinski R.J. and Flach, J.M (2003) *Control theory for humans: Quantitative approaches to modelling performance*. Mahwah, NJ: Lawrence Erlbaum

Johansson, E., Engström, J., Cherri, C., Nodari, E., Toffetti, A., Schindhelm, R., Gelau, C. (2004) Review of existing techniques and metrics for IVIS and ADAS assessment. EU project AIDE, project IST-1-507674-IP, Deliverable 2.2.1

Knipling, R.R. and Wierwille, W.W. (1994) Vehicle-Based Drowsy Driver Detection: Current Status and Future Prospects. IVHS America Fourth Annual Meeting, Atlanta, GA, Washington, DC, April, 1994, pp 245-256.

Lee, D.N. (1976) A theory of visual control of braking based on information about time-to-collision. *Perception*, 5, 437-459

Liu, Schreiner and Dingus. (1999) Development of human factors guidelines for Advanced Traveler Information Systems (ATIS) and Commercial Vehicle Operation (CVO): Human Factors Evaluation of the Effectiveness of Multimodality Displays in ATIS. NHTSA FHWA-RD-96-150

Mattes, S. (2003). The lane change task as a tool for driver distraction evaluation. In H. Strasser, K. Kluth, H. Rausch, and H. Bubb (Eds.). Quality of Work and Products in Enterprises of the Future. (pp. 57-60). Stuttgart: Ergonomia Verlag.

Marken, R.S. (1986) Perceptual organisation of behaviour: A hierarchical control model of coordinated action. *Journal of Experimental Psychology: Human Perception and Performance*, Vol. 12, No. 3, 267-276.

McDonald, W. A., & Hoffman, E. R. (1980). Review of relationships between steering wheel reversal rate and driving task demand. *Human Factors*, 22(6), 733–739.

McRuer, D. T., Allen, R. W., Weir, D. H., & Klein, R. H. (1977). New Results in Driver Steering Control Models. Human Factors, 19(4), 381 - 397.

McLean, J.R. & Hoffman, E.R. (1971) Analysis of Drivers' Control Movements. Human Factors, 1971, 13(5), 407-418.

Michon, J.A. (1985) A critical review of driver behaviour models: What do we know? what should we do? In L.A Evans and R.C. Schwing (Eds.) *Human Behaviour and Traffic Safety*. (pp. 487-525) New York: Plenum Press.

Muzet, A., Pébayle, T., Otmani, S., Peters, B., Nikolaou, S., & Bekiaris, E. (2004, 24-26 May). Preventing driver drowsiness at the wheel: Can steering grip senor measurements contribute to its prediction? Paper presented at the ITS Europe Congress, Budapest, Hungary.

Nakayama, O., Futami, T., Nakamura, T. and Boer, E.R. (1999) Development of a Steering Entropy Method for Evaluating Driver Workload. SAE Technical Paper Series: #1999-01-0892

Neisser, U. (1976) Cognition and reality: principles and implications of cognitive psychology. New York: W.H. Freeman and Company.

Nilsson, L. (1995) Safety Effects of Adaptive Cruise Controls in Critical Traffic Situations. Proceedings of Steps Forward, Volume III, the Second World Congress on Intelligtent Transportation Systems, Yokohama, Japan.

Nilsson, R. (2001) *Safety Margins in the Driver.* PhD thesis Uppsala University, Sweden.

Nygård, M. (1999) *A method for analysing traffic safety with help of speed profiles* (Master's thesis) Tampere University of Technology: Department of civil engineering Transportation engineering.

OECD. 1990. Behavioural Adaptations to Changes in the Road Transport System. Report Prepared by an OECD Expert Group, Road Transport Research Programme.

Patten, C. J. D., Kircher, A., Östlund, J. & Nilsson, L. (2004) Using mobile telephones: Cognitive Workload and Attention Resource Allocation. *Accident Analysis and Prevention* (36), pp. 341-350

Peters, B. (2004) Evaluation of Adapted Passenger Cars for Drivers with Physical Disabilities. Doctoral Thesis, Institute of Technology, Linköping University, Linköping.

Peters, B., & Nilsson, L. (2005) Driver behaviour models and driver support. Paper presented at the HUMANIST (EUR 21677) - The International Workshop on Modelling Driver Behaviour in Automotive Environments, Ispra, Italy.

Peters, B. Thorslund, B. & Östlund, J. (2005) *AIDE_VTI_T2.2.5_Steering Grip Analysis of Haste data, AIDE Internal deliverable.*

Peters, B., & Östlund, J. (2004, 5 - 9 September 2004) Alternative joystick designs for drivers of adapted cars. Paper presented at the International Conference on Traffic and Transport Psychology 2004, Nottingham.

Peters, B. & Östlund, J. (2005) *AIDE_VTI_WP2.2 Vehicle Sensor Requirements, AIDE Internal deliverable.*

Peters, B., & Östlund, J. (2005) Joystick Controlled Driving for Drivers with Disabilities - a driving simulator experiment (VTI Rapport No. 506A) Linköping: Statens Väg och transportforskningsinstitut.

Powers, W. T. (1998) Making Sense of Behavior - The Meaning of Control. New Canaan, CT: Benchmark.

Recarte, M. A., Nunes, L. M. (2000) Effects of verbal and spatial-imagery tasks on eye fixations while driving. Journal of Experimental Psychology 2000, Vol 6 No.1.

Recarte, M. A., Nunes, L. M. (2003) Mental Workload While Driving: Effects on Visual Search, Discrimination, and Decision Making. Journal of Experimental Psychology 2003, Vol. 9, No. 2.

Rumar, K. (1988). Collective risk but individual safety. Ergonomics, 31(4), 507 - 518.

Saad, F., Hjälmdahl, M., Canas, J., Alonso, M., Garayo, P., Macchi, F., Nathan, F., Ojeda, L., Papakostopoulos, M., Panou, M. and Bekiaris, E. (2004). Literature Review of Behavioural Effects. Deliverable 1.2.1, AIDE Integrated Project, Sub-project 1. IST-1-507674-IP

Schindhelm, R., Gelau, C., Montanari, R., Morreale, D., Deregibus, E., Hoedemaeker, M., de Ridder, S., & Piamonte, P. (2003) Human factor tests on car demonstrator *Deliverable no D6.4. Final draft. COMUNICAR Programme IST KA1. Testing demonstration and pilot evaluation Work package WP6.* Report nr: 2003 P 134.

Smiley, A. (2000) Behavioural Adaptation. *Safety and Transportation Systems Research Record*, Vol. 1724, pp. 47-51.

Summala, H. (1985). Modeling driver behavior: a pessimistic prediction? In L. E. a. R. C. Schwing (Ed.), Human Behavior and Traffic Safety (pp. 43-65). New York: Plenum.

Summala, H. (1988) Risk control is not risk adjustment: the zero-risk theory of driver behaviour and its implications. Ergonomics, 31(4), 491-506.

Vaa, T. (2005) Modelling Driver Behaviour on Basis of Emotions and Feelings: Intelligent Transport Systems and Behaviour Adaptation. Paper presented at the HUMANIST (EUR 21677) - The International Workshop on Modelling Driver Behaviour in Automotive Environments, Ispra, Italy.

van der Horst, R. (2005) Time-Related Measures for Modelling Risk in Driver Behaviour, Paper presented at the HUMANIST (EUR 21677) - The International Workshop on Modelling Driver Behaviour in Automotive Environments, Ispra, Italy.

van der Horst, R. and Godthelp, H. (1989) Measuring road user behaviour with an instrumented car and an outside-the-vehicle video observation technique. Transportation Research Record # 1213, Washington DC: Transportation Research Board, 72-81.

van der Hulst, M. (1999). Adaptive control of safety margins in driving. Doctoral Thesis, University of Groningen, The Netherlands, Groningen.

van Winsum, W. (1996). From Adaptive Control to Adaptive Driver Behaviour. Dotoral Thesis, University of Groningen, The Netherlands, Groningen.

van Winsum, W., & Godthelp, H. (1996) Speed Choice and Steering Behaviour in Curve Driving. *Human Factors, 38*(3), 434-441.

Verwey, W. B. (2000). On-line driver workload estimation. Effects of road situation and age on secondary task measures. Ergonomics, 43(2), 187-209.

Victor, T.W, Harbluk, J.L and Engström, J.A. (2005) Sensitivity of eye-movement measures to in-vehicle task difficulty. *Transportation Research Part F* (8), pp. 167-190.

Vogel, K. 2002. *Modelling driver behaviour: A Control theory based approach*. PhD Thesis, Linköping University, Sweden, Diss. No. 751.

Wickens, C. D. (1992) *Engineering Psychology and Human Performance* (2nd ed.) New York: Harper Collins.

Weiland, J., Mattes, S., Kuhn, F., Gelau, Ch., Schindhelm, R., Hoedemaker, M. and Belotti, F. (2003). Deliverable 6.3 – Driving Simulator Test results: Communication Multimedia Unit Inside Car (COMUNICAR; IST 11595) Project.

Weir, D. H. and McRuer, T.M. (1968) A theory of driver steering control of motor vehicles. *Highway Research Record*, 247, 7-39.

Weir, D and Chao, K.C. (2005) Review of Control Theory Models for Directional and Speed Control. In *Proceedings of Modelling Driver Behaviour in Automotive Environments*, Ispra, Italy.

Wilde, G.J.S. (1982) The theory of risk homeostasis: Implications for traffic safety and health. *Risk Analysis*, 2, pp. 209-225.

Wiener, N. (1948). Cybernetics: Control and communication in the animal and the machine. Cambridge, Massachusetts: MIT Press.

Wierwille, W., Tijerina, L., Kiger, S., Rockwell, T., Lauber, E. And Bittner, A Jr. (1996). Heavy Vehicle Driver Workload Assessment. Task 4: Review of Workload and Related Research. US Department of Transportation, NHTSA. DOT HS 808 467 (4).

Östlund, J., Nilsson, L., Carsten, O., Merat, N., Jamson, H., Jamson, S., et al. (2004) *Deliverable 2 - HMI and Safety-Related Driver Performance* (No. GRD1/2000/25361 S12.319626): Human Machine Interface And the Safety of Traffic in Europe (HASTE) Project.

Östlund, J. (2004) *Design of Critical Events for Assessing IVIS in Simulators - a VTI Driving Simulator Study*. Paper presented at the DSC 2004 Europe, Paris, France.

12 Appendix 1 - Aide Glossary

Term	Definition	References	Alternative definition	References alt. definition
Action	An event initiated by the driver or an **application**	Original AIDE definition		
ADAS	Systems that interact with the driver with the main purpose of supporting the **driving task** on the tracking and regulating levels.	Original AIDE definition		
AIDE design scenario	A driving situation, specified by at least one **action** and one or more **DVE state** parameters, acted upon by **the AIDE system**.	Original AIDE definition		
AIDE meta-function	The response of the AIDE system to an **AIDE design scenario**.	Original AIDE definition		
AIDE system	The Adaptive Integrated Driver-vehicle Interface targeted by the AIDE IP, implementing the AIDE meta-functions	Original AIDE definition		
Application	A program (such as a word processor or a spreadsheet) that performs one of the important tasks for which a computer is used	EAST-EAA (Webster)		
Application Programming Interface (API)	A software interface that enables **applications** to communicate with each other. An API is the set of programming language constructs or statements that can be coded in an application program to obtain the specific functions and services provided by an underlying operating system or service program.	EAST-EAA (http://www-3.ibm.com/ibm/terminology/.)		
Behavioural adaptation	The whole set of behaviour changes that are designed to ensure a balance in relations between the (human) organism and his surroundings, and at the same time the mechanisms and processes that underlie this phenomenon	Grand Dictionnaire de la Psychologie	Those behaviours which may occur following the introduction of changes to the road-vehicle-user system and which were not intended by the initiators of the change.” (p. 23).	OECD. 1990. Behavioural Adaptations to Changes in the Road Transport System. Report Prepared by an OECD Expert Group, Road Transport Research Programme.
Data	Information output by a sensing device or organ that includes both useful and irrelevant or redundant information and must be processed to be meaningful.	EAST-EAA (Webster)	Data is the software implmentation of an information. It can be exchanged between software components. A data	EAST-EAA (Webster)

			is persistent. It is persistent in memory.	
Device	Functional unit of hardware or software, or both, capable of accomplishing a specified purpose.	EAST-EAA (Functional safety: safety instrumented systems for the process industry section; Part 1: Framework, definitions, system, hardware and software requirements; IEC2002.)		
Distraction	see **driver distraction**			
Domain	A problem space	EAST-EAA (IEEE Standard for Information Technology – Software Life Cycle Processes – Reuse Processes; IEEE Standard 1517-1999;1999.)	Used as a synonym for the groups of products defined in the EAST project: chassis, powertrain, telematic, body and HMI.	EAST-EAA (IEEE Standard for Information Technology – Software Life Cycle Processes – Reuse Processes; IEEE Standard 1517-1999;1999.)
Driver distraction	Attention given to a non-driving related activity	ISO TC22/SC13 WG8 CD 16673 (Occlusion Committee draft)		
Driving demand	The demands of the driving task	de Waard, D. (1996). The Measurement of Drivers' Mental Workload. ISBN 90-6807-308-7. Traffic Research Centre. University of Groningen.		
Driving performance	Performance of the **driving task**			
Driving task	All aspects involved in mastering a vehicle to achieve a certain goal (e.g. reach a destination) , including tracking, regulating, monitoring and targeting.	Original AIDE defnition		
DVE state	A set of **parameters** representing certain aspects of the driver, the vehicle and the environment	Original AIDE definition		
Element	A component of a system; may include equipment, a computer program, or a human.	EAST-EAA (IEEE Guide for Developing System Requirements Specifications; IEEE Standard P1233a, 1998.)		
Feature	User-visible aspects or characteristics of a system.	EAST-EAA (P. Clements; L. Northrop; Software Product Lines – Practices and Patterns; SEI series in software engineering; Addison-Wesley, 2002.)	(B) A feature is a functionality that is specifically perceptible by the customer/stakeholder. (C) A feature is a prominent or distinctive user-visible aspect, quality, or	EAST-EAA (P. Clements; L. Northrop; Software Product Lines – Practices and Patterns; SEI series in software engineering; Addison-Wesley, 2002.)

			characteristic of the system.	
Function	A task, action, or activity that must be accomplished to achieve a desired outcome (EAST-EAA).	EAST-EAA (IEEE Guide for Developing System Requirements Specifications; IEEE Standard P1233a, 1998.); D3.2.1 draft	A specific service that can be offered by a **system**. The same Function can be offered from different systems	Federal Information Processing Standards Publications (EAST-EAA)
Functional Specification	Specification of the normal function of the system.	EAST-EAA (Safety terms for automation systems reliability and safety of complex systems; VDI/VDE 2000.)		
Functionality	A synthesis of functions to provide a major functional entity of a unit.	EAST-EAA		
Human Machine Interface (HMI)	A set of compoents that govern the interaction between the user and one or more vehicle systems	Original AIDE definition		
Integrated system	Two or more in-vehicle devices, which provide information to, or receive output from, the driver of a motor veehicle, whose output have been combined or harmonised	ISO TC22/SC13 WG8 PWI LCT 018		
Interface	Abstraction of a service that only defines the operations supported by that service (publicly accessible variables, procedures, or methods), but not their implementation.	EAST-EAA (Szyperski, Clements.; Component Software – Beyond Object-Oriented Programming; Addison-Wesley, 1997;	An interface consists of the set of assumptions that users of the component may safely make about it – nothing more, but nothing less.	Parnas, D.; Information Distribution Aspects of Design Methodology”; In Proceedings of the 1971 IFIP Congress, Ljubljana, Slovenia, August 1971. pp339-344.)
Interface-Specification	A set of language and message formats used for the communication between software components.	EAST-EAA		
IVIS	Systems that interact with the driver with the main purpose to support tasks on the targeting and monitoring levels, or do not support driving at all.	Original AIDE definition		
Mental workload	The specification of the amount of information processing capacity that is used for task performance	de Waard, D. (1996). The Measurement of Drivers' Mental Workload. ISBN 90-6807-308-7. Traffic Research Centre. University of Groningen.		
Message (in architecture)	A message is a group of data values that must be exchanged together. A typical reason for grouping data is the temporal consistency of different data	EAST-EAA		

	values: a control algorithm may require for example that the temperature and the pressure are measured at the same time. Depending on the size of the message and the maximal size of frames, several messages may be transported by one frame or it may be necessary to split a message into several segments for being able to send it over a network.			
Parameter	An independent variable used to express the coordinates of a variable point and function of them.	EAST-EAA (Webster)		
Parameterisation	To express in terms of **parameters**.	EAST-EAA (Webster)		
Performance	see **driving performance**			
Primary task	The task with the highest priority in a multi-tasking situation.	Original AIDE definition		
Realtime	System which has to finish the processing within a specific time interval (deadline) dedicated by its environment.	EAST-EAA		
Reference task	**Secondary task** activity which consitutes a primary task performance reference level when performed concurrently with the **primary task**	ISO TC22/SC13 WG8 PWI LCT 018		
Requirement	A condition or capability needed by a user to solve a problem or achieve an objective.	EAST-EAA (IEEE Guide for Developing System Requirements Specifications, IEEE Standard P1233a, 1998.)	(B) A condition or capability that must be met or possessed by a system or system component to satisfy a contract, standard, specification, or other formally imposed document. (C) A documented representation of a condition or capability as in definition (A) or (B). (D) The necessity that a system has a particular feature. (E) A requirement expresses condition or capability that must be met or possessed by a system or system component to satisfy a contract,	EAST-EAA

			standard, specification or other formally imposed properties.	
Secondary task	A task with lower priority than the **primary task** in a multi-tasking situation.			
Specification	Precise (formal if possible) description of an object within the scope of the task.	EAST-EAA (Safety terms for automation systems reliability and safety of complex systems; VDI/VDE 2000.)		
System	A collection of components organized to accomplish a specific **function** or set of functions.	EAST-EAA (IEEE Recommended Practice for Architectural Description of Software-Intensive Systems; IEEE Standard P1471, IEEE Architecture Working Group (AWG), 2000.)	Set of elements, which interact according to a design; an element of a system can be another system, called a subsystem, which may be controlling system or a controlled system and may include hardware, software and human interaction	Functional safety: safety instrumented systems for the process industry section; Part 1: Framework, definitions, system, hardware and software requirements; IEC2002.
System response delay (SRD)	Interval during which the driver has to wait for an interface to respond or update in order to continue the task	ISO TC22/SC13 WG8 CD 16673 (Occlusion Committee draft)		
Task	Process of achieving a specific and measurable goal using a prescribed method	ISO TC22/SC13 WG8 CD 16673 (Occlusion Committee draft)		
Use case	An intended or desired flow of events or tasks that occur within the vehicle and are directed to or coming from the driver in order to accomplish a certain system-driver interaction.	Original AIDE definition	A model of the usage by the user of a system in order to realise a single functional feature of the system.	EAST-EAA
Visual demand	Degree of visual activity required to extract information from an object to perform a specific **task**	ISO TC22/SC13 WG8 CD 16673 (Occlusion Committee draft)		

13 Appendix 2 – Calculation of TLC and identification of TLC minima

Here, a procedure for calculating the TLC data is described. The method was designed by van Winsum and Godthelp (van Winsum, W., & Godthelp, H. (1996). Speed Choice and Steering Behaviour in Curve Driving. *Human Factors, 38*(3), 434-441).

1. LP_right = conventional lateral position, as described in this document
2. LP_left = distance from left wheel to left lane marking
3. LV = lateral velocity relative road
4. LA = lateral acceleration relative road (change in lateral velocity)
5. TLC = LP_right/(LV+LA) if LA<0 (accelerating to the right)
6. TLC = LP_left/(LV+LA) if LA>0 (accelerating to the left)
7. TLC is undefined for the following conditions:
   - LA=0
   - LP_right < 0 or LP_left < 0 (outside the lane)
   - TLC > 20 seconds (proposed and used in the HASTE project (Östlund et al, HASTE Deliverable 2, 2004)
8. Now you should have something like in the figure below. Note that TLC calculated for the right side is negative, and for the left side positive.
9. In order to achieve local minima, identify TLC wave forms that are at least *1 second long* (proposed and used in HASTE). For these, identify the local minima (maxima for negative waveforms), as in *Figure 68*.
10. Calculate the relevant TLC-metrics.

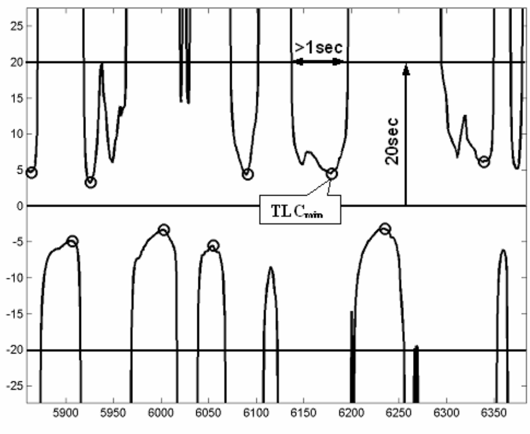


Figure 68 – TLC wave forms and identification of TLC min values. Figure taken from HASTE Deliverable 2 (Östlund et al, 2004).

If TLC is used in field experiments, some filtering will have to by applied to the LP, LV and LA data in order to avoid amplification of noise. Please, look into this carefully and visually verify the quality of your data. You should get something like in the figure above. A low pass filter with C/O frequency no less than 3 Hz should be a good starting point for LP, LV and LA. Poor precision in LP data can however not fully be compensated by filtering. Make a competent assessment of the usability of your TLC data. Do not use if poor.