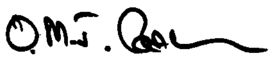
Human Machine Interface And the
Safety of Traffic in Europe
Project GRD1/2000/25361 S12.319626

Deliverable 1

- Development of Experimental Protocol -

INSTITUTE FOR TRANSPORT STUDIES
DOCUMENT CONTROL INFORMATION

Title	HASTE Deliverable 1 ***Development of Experimental Protocol***
Author(s)	A.J. Roskam, K.A. Brookhuis, D. de Waard, O.M.J. Carsten, L. Read, S. Jamson, J. Ostlund, A. Bolling, L. Nilsson, V. Anttila, M. Hoedemaeker, W.H. Janssen, J. Harbluk, E. Johansson, M. Tevell, , J. Santos, M. Fowkes, J. Engström, T. Victor
Editors	Samantha Jamson, Oliver Carsten
Reference Number	HASTE-D1
Version Number	1.1
Date	3 September 2002
Distribution	Unrestricted
Availability	Public
File	C:\Documents and Settings\ocarsten\My Documents\HASTE\WP1\HASTE Deliverable1 v1.1.doc
Authorised	O.M.J. Carsten
Signature	

Executive Summary

The *aim* of HASTE is to develop methodologies and guidelines for the assessment of in-vehicle information systems (IVIS), i.e. to formulate pass/fail criteria for IVIS. A major technical and scientific *objective* of HASTE is the identification and exploration of the relationship between traffic scenario, driver and IVIS. This relationship will be investigated by studying behavioural, vehicle, and psycho-physiological, and self-report measures.

In Deliverable 1 a preliminary experimental design is presented that aims to investigate and improve the understanding of this relationship. The approach has resulted in a rather detailed experimental design. In order to achieve an manageable operational design, a selection had to be made from independent variables constituting the triad scenario-driver-IVIS on the one hand, and the measures (dependent variables) within this triad on the other hand. For both independent and dependent variables the selection can be divided into a practical and a theoretical part. For both types of variables the practical part is feasibility, both in terms of resources (time, money) and expertise. The theoretical aspect of the selection of the dependent variables is methodology-related. Diagnosticity, validity and sensitivity for task load as well as the resulting driving capability determine the set of the measures.

The (theoretical) selection of the *in*dependent variables, i.e. the actual experimental design to be used in WP2 is mainly determined by the generic transferability of the experimental findings to the actual field environment. The combinatorial space of factors IVIS, Scenario, and Driver is far too great to test all possible combinations. Basically, the concepts of transferability (ecological validity) of the experimental results and feasibility of the experimental design may be inversely related.

The inclusion of the so-called surrogate IVIS is a special case of this transferability. The surrogate IVIS is an experimental analogue for a real IVIS. It is a computer task of which cognitive, perceptual and manual-motor characteristics can be configured independently, while keeping the basic task the same. Comparison of reaction patterns between surrogate and real IVIS would allow for inferences to be projected in terms of used mental resources and driving quality.

In sum, the main challenge of this deliverable was to find a balance between transferability of results and feasibility of design when formulating a preliminary experimental design. Data will have to be acquired in a diversity of environments, drivers, and IVIS that are together sufficiently broad to allow the formulation of generic guidelines. At the same time it should also be possible to conduct the experiment in such a way that it yields qualitatively adequate data, and carried out within the given amount of time and resources. Deliverable 1 is a first step in the sequence protocol (WP1), formalisation & experiments (WP2), and validation (WP3).

Table of contents

List of Figures

List of Tables

1 Introduction

1.1 Overview of the HASTE project

The aim of HASTE (**H**uman Machine Interface **A**nd the **S**afety of **T**raffic in **E**urope) is to develop methodologies and guidelines for the assessment of In-Vehicle Information Systems (IVIS). The following definition, which is stated in the EC Statement of Principles, will be adopted for IVIS in HASTE: *An IVIS (In-Vehicle Information System) is an In-vehicle Information & Communication System designed for use by the driver while driving.* To date, there have been attempts to provide manufacturers and testing authorities with a set of guidelines to assess the likely impacts of IVIS on the driving task, usually in the form of a checklist. Such checklists provide a tool that enables the identification of likely problems but they do not attempt to quantify safety problems. This project is fundamental to the development of a valid, reliable and efficient tool that will aid testing authorities in their safety evaluation of IVIS.

The objectives of the programme of research are:

- To identify and explore relationships between traffic scenarios in which safety problems with an IVIS are more likely to occur
- To explore the relationships between task load and risk in the context of those scenarios
- To understand the mechanisms through which elevated risk may occur in terms of distraction and reduced Situation Awareness
- To identify the best indicators of risk (accident surrogates)
- To apply the methods devised to evaluating real systems
- To recommend a pre-deployment test regime that is both cost effective and possesses the validity to predict performance
- To recommend an approach for the preliminary hazard analysis of an IVIS concept or design
- To review the possible causes of IVIS safety hazards, including those related to reliability, security and tampering

The current international state of the art in terms of methodologies for assessing the safety implications of IVIS is highly problematic, with the major drawback being that the tools and metrics that have been provided do not permit, in any straightforward way, judgements to be made about the safety of a particular IVIS during use while driving. There are, as a result, no criteria which can be used by a manufacturer, a system supplier or the public authorities to determine whether a particular design meets a minimum threshold of safety in actual use.

Three important steps in meeting the above mentioned objectives are (1) *refinement of knowledge* about the impacts of IVIS on driving performance. An IVIS may have negative or positive consequences on driving. However, since IVIS-related performance decrements are crucial for the final safety judgement, in HASTE the emphasis will be on the negative effects of IVIS on driving performance. Deliverable 1 (D1, this document) comprises the first phase of this knowledge refinement step. It encompasses the establishment of methods and metrics and the definition of scenarios in which IVIS-related safety problems are likely to occur. These methods, metrics, and scenarios will be used in various experimental settings (D2). The experiments are the second phase of this process of knowledge refinement. They aim to identify the best risk indicators for IVIS use during driving. These risk indicators will be used for the (2) *development of a testing regime* for IVIS that is both simple and valid (D3). Both surrogate and real IVIS will be used in the development of this testing regime. The conclusions of the tests comprise the final report (D6) of the project: (3) *formulation of guidelines* for the future development of IVIS. These guidelines aim to provide authorities with a practical pass-fail procedure for IVIS.

In sum, the HASTE project aims to develop a generic safety judgement procedure for IVIS. The ultimate goal of the project is primarily a practical one. However, prior to this, research is needed in order to ensure the pass-fail procedure for IVIS is both valid and predictive. Several ‘candidate measures’ will have to be investigated for these characteristics. Therefore, a theoretical, explorative phase precedes the practical goal of formulating pass/fail criteria and safety guidelines.

1.2 Scope

When multiple IVIS are simultaneously used, it seems likely these IVIS will mutually interact or interfere. Overall task complexity is likely to increase when more IVIS (sources of information) are added. The specific configuration of number, mode ((dis)similarity) and relative timing (contiguity) of device inputs and/or outputs together determine the overall task complexity. This leads to the conclusion that the effects of the use of multiple IVIS on driving quality needs to be studied. This conclusion is underlined by the fact that the simultaneous use of multiple IVIS is something that happens in reality.

However, this issue is beyond the scope of HASTE. The interaction of a driver with only one IVIS is complex as it is. The lack of safety criteria necessitates the interaction of a single IVIS with the driver to be scrutinised first. By comparison, the Advisors project (GRD1-1999-10047), which deals with advanced driving assistance systems (ADAS), also studies one device at a time. The effects of multiple ADAS on (amongst others) driving quality are to be studied in the (proposed) project Advisors II.

1.3 In-Vehicle Information Systems

Central to the HASTE project is the consideration of “In-Vehicle Information Systems (IVIS)”, i.e. on-board systems that provide information to the driver. IVIS is a collective noun for a very diverse set of devices, with functions varying from navigation and traffic information to feedback on driving ability. In addition to this heterogeneity of IVIS, the diversity of drivers and driving environments complicates a straightforward assessment of whether ‘doing two things at the same time’ (i.e. driving a car and operating an IVIS) compromises traffic safety. The fact that the cost of such technology is decreasing, could mean an increase in use of IVIS in the future.

The diversity in IVIS, drivers, and driving environments (see Figure 1), complicates the formulation of safety criteria for IVIS. Probably for that reason, such a ‘pass-fail’ criterion does not exist (see also Brookhuis et al., 1999). Therefore, the *primary objective* of the HASTE project is the creation of a benchmark for the safety of an IVIS. In the following text, this justification for the HASTE project will be described more elaborately, addressing the interacting factors human, machine, and their surrounding.

Errors in the processing of information are probably the major contributory factor in traffic accidents. Relevant causes of traffic accidents include problems in perception, attention, distraction, etc., in other words, basically human errors. The cause of human error lies in the limited capacity of the human information processing system, in particular when information comes from different sources, forcing divided attention. For instance, with increasing information rate from external sources while driving, mental workload increases rapidly. There are limits to the mental demands that can be imposed on the driver participating in traffic. Information processing capacity of human operators is specifically limited in that respect. However, this can depend on the driver, the relevant factors being age, experience, and idiosyncratic characteristics. Information load and the extent to which it “consumes” capacity are related to accident involvement, dependent of these factors as well.

In addition to the ability to effectively cope with a relatively high information load, there is another factor determining the efficacy of one’s behaviour. This is the ability to handle sudden shifts of workload. The introduction of a novel event requires an attentional switch from one task to another. This task switch may result in a quality trade-off from one task to another. Good examples of sudden events requiring attentional shifts are a beeping IVIS, but also external events like falling cargo from a leading car. It has to be investigated whether, where, and how a ‘performance breakdown’ occurs during situations that require a sudden attentional shift.

A host of driver information and support systems have been developed to counteract the likelihood and impact of human errors in traffic, although most of these systems were not developed, or at least not introduced for that specific reason in the first place. The systems available on the market now and/or in the near future might be categorised as to the impact on the driver, from noncommittal information supply via active support at driver’s will to leaving control to technology completely. Examples of IVIS are route navigation systems, traffic information systems and driver monitor systems. The effect of any system, even if explicitly developed for a positive contribution to safety, is (potentially) negative as well. Basically, information provision by an IVIS distracts, draws attention and takes away perception from the road.

A further consideration is the diversity of the road environment. Variations in road type, traffic density, illumination level, to mention a few, all affect the mental demands of the driving task (see Brookhuis et. al., 1999). Given the limitations of the information processing capacity of the human brain, it can be inferred that performance of a dual task will affect performance of one or both tasks. There is a host of evidence that shows that a trade-off between primary and secondary task commonly occurs and that this trade-off is caused by limited information processing capacity, although this be partially counteracted by using scheduling strategies (Wickens, 1992) and the use of different resources for both tasks (Tsang & Wickens, 1988). Most importantly, there are bottlenecks in perceptual, attentional, and short-term memory processes in the brain.

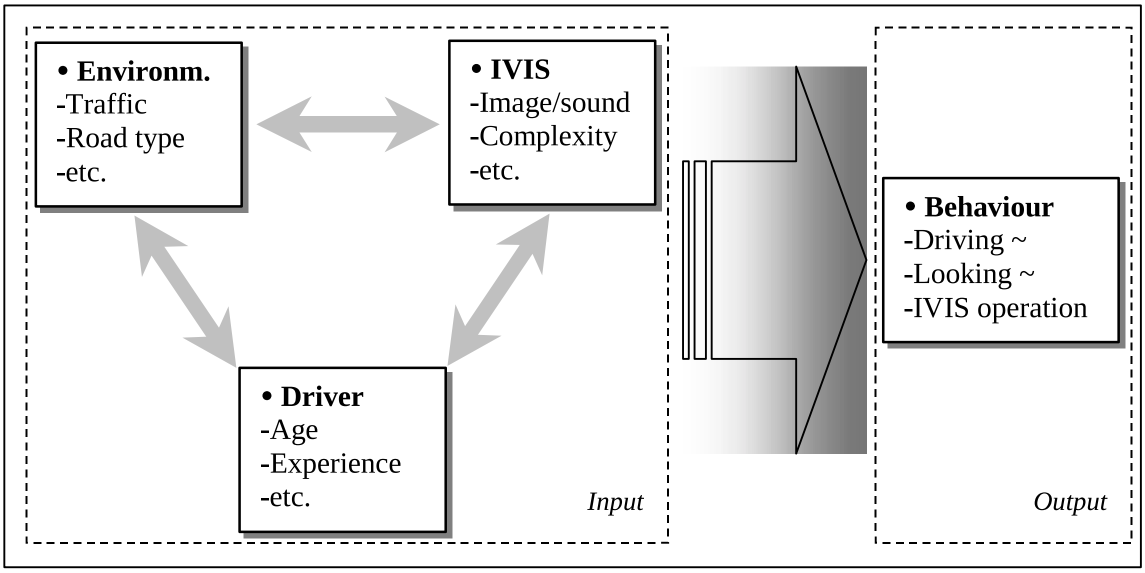


Figure 1: The interaction between Driver-Environment-IVIS

Figure 1 shows the factors (‘Input’, left) that affect behaviour (‘Output’, right), amongst which is the operation of an IVIS. The enumerations are examples of factors that affect the characteristics of environment, IVIS and driver. These three factors themselves are in mutual interaction, which is indicated by the double-headed arrows.

In sum, the variety of the three mutually interacting factors makes it an arduous task to formulate a generic benchmark for the Driver-Environment-IVIS interaction. The interaction must take account of:

- Driver-related factors: variety in driver experience and predisposition.
- IVIS-related factors: variety in IVIS types, functions and physical appearance;
- Environment-related factors: variety in traffic conditions.

Pressing questions to be faced are (see also Brookhuis et al., submitted):

- What should be measured to judge the safety of driving behaviour?
- When is one’s driving behaviour below standard?
- How is this standard determined?
- When is one’s ability to process information unacceptably low?

It seems obvious to use vehicle measures, in particular referring to the car’s longitudinal and lateral position on the road. These measures will have to be interpreted both in an absolute and in a relative sense. An example of an absolute safety criterion is following distance, that is the distance to a leading car. Within a certain (absolute) distance, nobody is humanly able to stop a car to prevent a crash. A relative criterion describes the trend of a given measure as a function of time. For example: when a driver swerves increasingly after several hours of driving, this can be reason to rate driving as inadequate. The detrimental effect of time-on-task on driving performance would be so great as to cause (literally) abnormal driving behaviour.

Vehicle measures have been introduced to explain the distinction between absolute and relative safety criteria (Brookhuis et al., in press). These types of measures tell us something about the amount of control the driver has over his/her vehicle, both longitudinally (front/rear) and laterally (left/right).

Next to vehicle measures, one may look at behavioural safety measures as well. Good examples of this kind of measure are reaction times (RT) and visual behaviour (VB). RT mainly gives an indication of the complexity of the task at hand (which may be driving, operating an IVIS, or both), whereas VB provides information about the perceptual and attentional load of both tasks. Besides RT and VB, the quality and speed of IVIS operation per se might be used to assess driving safety. An IVIS with a complex display might be operated more slowly and less accurately and be more likely to compromise traffic safety.

Key questions are what the differences that these measures show mean, and under which circumstances, with and without IVIS. What behaviour does the IVIS illicit under which circumstances and what is the accompanying safety rating?

This Deliverable is an output of Workpackage 1. This Workpackage aims to define the experimental protocol to be used in the planned experiments i.e. appropriate assessment parameters, scenario's and subjects' tasks in the study of the effects of specific IVIS. This Deliverable provides a systematic approach in the preparation of the experiments to be carried out.

1.4 Structure of this document

The definitions of metrics in the pilots concern the interaction between the driver-IVIS-environment, and the measurement tools related to this:

- The *type of systems* to be included in the project. The operation of an IVIS poses an additional task on the driver. Therefore, the effect an IVIS has on driving quality will be assessed using the so-called 'secondary task paradigm'. An experimental analogue for IVIS ('surrogate-IVIS') will be used in order to enable systematic variation of system characteristics. However, in order to gain insight into the variety of appearances of the secondary task paradigm, a literature review was carried out first (**Task 1**);
- The choice rationale for the underlying *secondary task paradigm* in the HASTE context will be more thoroughly investigated by means of five pilot studies. The results of these pilot studies on memory, attention, and vigilance are described in **Task 2**;
- The *type of traffic conditions* (**Task 3**) i.e. traffic densities, road types, weather conditions, that are to be included in the experiments. Scenarios will be devised in order to ensure cross-situation generalisability of the experimental findings;
- The *dependent variables* (**Tasks 4, 5** and **6**) that should be measured in the pilot tests. Tasks 5 and 6 go into detail about two major approaches for the assessment and interpretation of mental labour – the concept of workload and that of situation awareness;
- The *type of participants* (**Task 7**), i.e. drivers in the European Driver Population, that are to be included in the experiments;
- The implementation and interpretation of the surrogate-IVIS, as well as various other independent variables, will be discussed in their proposed experimental context in **Task 8**.

Table 1 displays the Internal Deliverables and their task numbers. This document is has a chapter for each task.

Table 1: Overview of the Internal Deliverables for WP1 of HASTE

Task Number	Task Name	Partner Responsible
1	Secondary Task Review	ITS/Leeds
2	Memory, Attention and Vigilance - Five Pilot Studies	ITS/Leeds
3	Scenario Definition	VTT
4	Driving Performance Measures	VTI
5	Workload Measures	TNO
6	Situation Awareness Measures	Transport Canada
7	Driver groups	ITS/Leeds
8	Performance Criteria	TRAIL/Delft

2 Secondary Task Review

2.1 Introduction

The operation of an IVIS could generate moments of high mental and physical workload for drivers. Whilst it is important to ensure that such in-car systems are used efficiently by drivers, interference with the driving task must be kept to a minimum. For instance, to enable optimum driving performance, the secondary task should not overly compete with the drivers' visual or mental workload. In addition, effort must be made to reduce the requirement for a manual response to the secondary task, if this is necessary for the driving task.

Although a sufficient number of studies have examined the interaction between in-car secondary tasks and driving performance (Radeborg, Briem & Heman, 1999; Verwey, 2000; Lui, 2001), there is an absence of data on the systematic relationship between increasing secondary task complexity and driver performance. In addition, the use of miscellaneous methodological techniques and the employment of different modes of input and variable response styles do not allow a satisfactory comparison across studies.

This report summarises the results of a number of studies that have examined the interaction between secondary task performance and driving. Based on the results of these investigations and on the literature concerning dual-task performance, a number of secondary tasks appropriate for the pilot stage of the HASTE project are recommended. Furthermore, the report provides a brief outline of some of the experiments on crossmodal interaction and multisensory integration, suggesting a number of dual task paradigms that may benefit from findings in this area.

However, before reviewing the list of secondary tasks, a brief description of the theoretical framework used in dual-task performance is outlined below.

2.2 Dual task methodology

The use of dual or secondary task paradigms has proved to be a fruitful technique in assessing the limitations of human operators. This approach has utilised multi-component models such as working memory (Baddeley & Hitch, 1974; Baddeley, 1986, 1990; Baddeley & Logie, 1999), and the multiple resource theory (MRT) introduced by Wickens (1984; 1992).

Working memory refers to a system that is responsible for the online processing and maintenance of information for a short amount of time. It is a system which allows us to understand and interact with our surrounding environment, assists us in acquiring new skills and drawing upon old skills to solve problems and accomplish goals (Baddeley & Logie, 1999). The model consists of a *Central Executive*, which acts very much like a supervisory attentional system (Norman & Shallice, 1986), controlling and manipulating two 'sub-systems': the *Phonological Loop* (PL) and *Visual Spatial Working Memory* (VSWM). Auditory/verbal material are assumed to have direct access to the phonological loop, while the visual-spatial characteristics of objects, such as their shape and orientation are dealt with by VSWM.

Dual task studies using the working memory model are based on the assumption that if two tasks share the same working memory resource, performance in one or both deteriorates when the two tasks are done together, compared to when each task is performed alone. The results of such studies have helped develop the model, and for instance show a clear distinction between visuo-spatial and phonological components. For example, Logie (1986) asked subjects to remember a series of words, either by using a pegword mnemonic technique (thought to rely on visuo-spatial imagery) or by rote rehearsal (believed to rely on phonological resources). Encoding of the word list was either accompanied by irrelevant speech (played to subjects via headphones) or irrelevant visual patterns (presented on the computer, which subjects observed but ignored). Irrelevant visual patterns were found to impair memory for the list of words learnt by pegword mnemonic, while learning by rote rehearsal was interrupted by irrelevant speech. Therefore, in addition to demonstrating a clear distinction between verbal/phonological and visuo-spatial working memory resources, this experiment, and many analogous examples, illustrates our ability to conduct concurrent tasks that do not rely on the same processing resources.

The multiple resource theory (MRT; Wickens, 1984, 1992) is broadly based on a similar concept, using secondary tasks as a measure of residual capacity not utilised by the primary task. This model assumes the presence of three dichotomous limited capacity resources defined by *processing stages* (early perceptual versus late central processing), *modality* (auditory versus visual encoding), and *response codes* (spatial versus verbal). As with the working memory model, studies on MRT have shown that optimal dual task performance is achieved when there is minimum conflict between resources for each task. Therefore, if a primary task incorporates a high visual load (e.g. driving), an auditorily delivered secondary task is shown to produce less disruption than a visual secondary task (see Parkes & Coleman, 1990).

While temporal and processing similarities between two tasks usually leads to dual task interference, in some circumstances, such similarities can result in the *facilitation* of dual task performance. This facilitation has been observed in movement tasks, for instance when the movement of two hands is much more successful if they follow the same pattern or rhythm (Klapp, 1981; Heuer, 1996). In such conditions, similarities in temporal or processing mechanisms can lead to a *co-ordination* between the two tasks, resulting in a more successful use of resources, as opposed to a competition for these facilities. Indeed, in the most extreme case, such co-ordination results in the performance of the two tasks as a single task (Klapp, Hill, Tyler, Martin, Jagcinski & Jones, 1985). Therefore, presentation order of stimuli for each task, and the response sequence required by each task, are also essential determinants of successful performance. Specifically, while ultimate performance for some tasks is achieved via co-ordination of responses, response scheduling is a more successful technique for other task combinations.

In summary, results of studies on multi-component models of human memory and information processing suggest that concurrent performance of a number of tasks is most successful when there is minimal overlap of resources for each task. However, even when processing resources are not shared, there are other factors that can produce task impairment within a dual task paradigm (Bourke, Duncan & Nimmo-Smith, 1996). These include:

- The priority assigned to each task by the experimenter;
- The priority assigned to each task by participants (this may be regardless of experimenter instructions);

- Participants' idea of perfect performance;
- Participants' idea of acceptable performance.

Within a driving environment, it is expected that both priority and perfect performance are primarily assigned to the driving task at all times, and that attention to the secondary task is only applied when the driving task is least demanding. Nevertheless, to make sure that a driver's attention to the road is not jeopardised by the secondary task, attempts must be made to ensure that the secondary task is presented at times of low driving workload.

The next section of this report describes a number of studies that have examined the interaction between secondary task performance and driving. Particular attention is given to tasks considered suitable to the HASTE project.

2.3 Review of secondary tasks

The interaction between secondary task performance and driving has been studied in both simulators and out on the road. Generally, the effect of secondary tasks on driving has been examined by recording variables such as driving speed, curve negotiation and braking time. Correspondingly, the effect of driving on secondary task performance is shown to be related to: drivers' age and experience; time of day (possibly related to traffic volume); road situation and road familiarity (Haingey and Westerman, 2001; Zeitlin, 1993; Verwey, 2000). For instance, Verwey (2000) examined the effect of road situation on secondary task performance, by dividing a section of track into variably loading segments (based on traffic density and visual/mental demand). Results demonstrated an impairment of the visually loading task (visual detection) at complicated road locations such as roundabouts and busy intersections.

The effect of driving on secondary task performance can be evaluated using objective measures such as (i) reaction time, (ii) number of accurate responses and (iii) number of missed responses. Workload can also be assessed using subjective measurements such as SWAT (Reid and Neigren, 1988) and NASA - TLX (Hart & Staveland, 1988), which rate the physical, temporal and mental demands of a task. Finally, a number of studies have also monitored subjects' physiological reactions during dual task performance, including effects on heart rate, blood pressure, respiration and eye-movements (e.g. Rantanen & Goldberg; 1999; Veltman & Gaillard, 1998). The objective, physiological and subjective scores achieved on a task can be determined by a list of variables, a number of which are described in the next section.

2.3.1 Manipulating Task Difficulty

As discussed above, successful secondary task performance during driving relies greatly on the means (modality) used for its presentation. However, before establishing the appropriate modality for presentation and response of a secondary task, it is important to outline the factors that are thought to effect task difficulty, although at present, there are no strict guidelines in the literature within this field. A number of determinants are listed below.

2.3.1.1 Number of stimuli

Human operators are able to process between 5 and 9 'chunks' of information at any one moment of time (Miller, 1956). Therefore, task difficulty can be manipulated by systematically increasing the number of items, or ‘chunks of items’ presented to subjects, from 2 to 9 units.

2.3.1.2 Stimulus presentation rate and time of recall

The rate at which stimuli are presented influences performance accuracy, although the exact relationship is related to the particular task demands. For example, if subjects are asked to respond to each stimulus as soon as it is presented in a simple reaction time task, comfortable performance can be achieved at a rate of one stimulus/response per second - but task difficulty increases with faster rates of presentation.

The relationship between rate of presentation and *memory* for items is less apparent. For instance, a high rate of presentation (or immediate recall of items) induces a short delay between encoding and retrieval, and can improve memory for sequentially presented lists. Alternatively, reducing the rate of presentation (or delayed recall) increases rehearsal time for each item, therefore improving memory.

2.3.1.3 Serial versus simultaneous presentation of stimuli

This is an important consideration for visually presented stimuli. For example, simultaneous presentation of a large number of stimuli may create an unnecessarily high visual load for subjects, taking their attention away from other visual tasks. Alternatively, serial presentation of visual information may be less demanding, especially if introduced at a slow rate, but if successful performance is contingent on remembering the *order* of item presentation, the task can become quite demanding.

2.3.1.4 Recognition versus recall of items

Memory for a list of words, or a visuo-spatial pattern of objects can be tested either by asking subjects to recall the items following presentation, or by re-presenting the stimuli, and testing subject’s recognition. On tests of recognition, original stimuli can be replaced by other (previously absent) stimuli, and/or the order of presentation/spatial position of items can be modified. A comparison between tests of recall and recognition has repeatedly demonstrated that recognition memory is better than recall (see Parkin, 1993), although the rationale for these findings is not well understood.

To summarise, the above parameters can be used in isolation or combination to manipulate task difficulty. The tasks described below incorporate some of these manipulations.

2.4 Summary of suitable tasks

The tasks described here test a range of perceptual and cognitive abilities. The section starts by describing a number of simple and choice reaction time tasks, before outlining experiments that require sustained attention and higher cognitive loads. In the majority of cases, an attempt is made to introduce tasks that have already been employed in studies of driving. The ability to manipulate task load/difficulty - based on the factors described above - is also considered an important criterion for task selection. In addition, tasks that can be presented both visually and auditorily are favoured, whilst we have also endeavoured to select tasks that necessitate minimal visual load, allowing the bulk of visual attention to be dedicated to the driving task. Finally, each of the described tasks is classified into those that measure capacity when administered in isolation; and those that induce processing limitations when performed in dual-task circumstances.

2.4.1 Testing attention and vigilance

In these tasks, subjects are simply asked to respond to a specific stimulus by pressing a button. In *simple* reaction time tasks, subjects respond every time the stimulus is presented. In a *choice* reaction time task, a different response is made to each particular stimulus - e.g. right button for even numbers and left button for odd numbers. Two visual attention and one auditory attention tasks are described. Since visual attention is principally required for the driving task, the visual attention tasks were chosen due to their brief attentional demand. The selected auditory attention task is thought to represent circumstances where drivers are required to attend and respond to a variety of in-vehicle devices, each of which may be distinguished by a particular sound signal.

2.4.1.1 Visual search

Subjects are required to identify a target stimulus, presented occasionally within a visual array of distracters. Performance is judged by measuring reaction time and accuracy of target detection. Speed of visual search is shown to depend on:

i. the similarity between targets and non-targets (Treisman, 1988, 1992; Duncan and Humphreys, 1992)
ii. the number of non-targets
iii. the similarity of non-targets to the target.

In these tasks, the displays usually remain on the screen for no more than two seconds, or until the subject produces a response, which is usually much shorter. Subjects' attention to the task can be attracted using an auditory signal. A number of examples are illustrated in Figure 2. Here, the target letter is a red T and task difficulty increases from left to right.

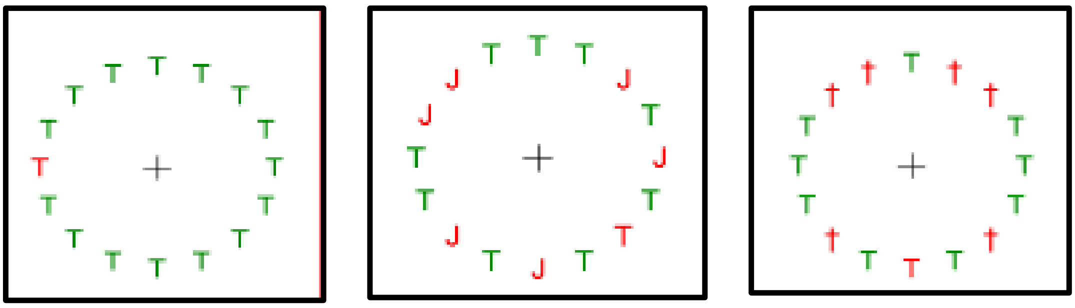


Figure 2: Example of a visual search task

Useful Field of View (UFOV) refers to an area within the visual field where stimulus information can be adequately conceived by subjects. The shape and size of UFOV are influenced by an increasing mental workload (Williams, 1995; Rantanen & Goldberg, 1999; Verwey, 2000). An example of this task includes the detection of numbers between 20 and 99, presented at a random rate on the corner of the car dashboard (Verwey, 2000). Peripheral detection is excluded by presenting the letters GG between each pair of numbers. Results show a reduction in detection accuracy with increased workload, induced by driving in 'complicated' road scenarios such as roundabouts.

It is important to not that whilst the use of such visual detection tasks can be beneficial when examining the effect of increased driving workload, they are not very informative if performed in isolation.

2.4.1.2 Auditory detection task

A series of distinguishable sounds are presented via headphones, and subjects are required to press a designated key for each particular sound. Task difficulty can be manipulated by

i. increasing the number of auditory stimuli from 2 to 5
ii. increasing presentation rate
iii. changing presentation rate from regular to random
iv. including an inhibition condition, where subjects are asked to inhibit response to a particular auditory stimulus.

2.4.2 Tests of immediate and short term memory

In this section, an attempt is made to describe a number of tasks that engage subjects' attention and require the retention of information for a short period of time. In line with the working memory model, tasks are arranged into those that require visuo-spatial resources (with minimal central load) and those that place a high load on central resources.

2.4.2.1 visuo-spatial and motor tasks

In general, tests of spatial memory and perception require a rather high degree of visual attention, and are considered unsuitable as secondary tasks during driving. Furthermore, there are currently only a small number of tasks that can validly test memory for spatial information. For instance, some tests of spatial memory are inclined to invite the use of verbal labelling resources, while others are thought to engage central working memory resources. Examples include some matrix tasks (Phillips & Christie, 1977a; b - see Figure 3), and the Corsi Blocks task respectively (De Renzi & Nichelli, 1975). In addition, spatial tests that rely on an extensive use of mental imagery are not considered suitable, as they require a high workload (see Logie, 1995). Nevertheless, a number of tasks are described which are believed to engage central resources without imposing too much central load.

Figure 3 shows an example of patterns used by Phillips and Christie (1977a; b). Subjects are shown a four by four matrix, with some filled cells. Following a retention interval (which may be filled with another task), the matrix is represented, and subjects must decide if the presentations were the same. Difficulty can be manipulated by changing the number of filled cells, and increasing the matrix size.

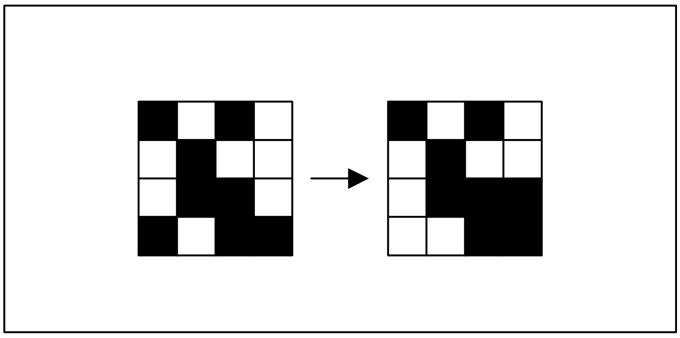


Figure 3: Memory for circular objects

Smith and Jonides (1997) used a visually presented spatial task based on the Sternberg paradigm (see below), to test the activation of working memory in neuroimagery studies. This task is considered an adequately 'pure' test of spatial resources, as it requires minimal use of verbal and central memory components. In Figure 4, a spatial memory task used by Smith and Jonides (1997) is shown. Subjects are shown a fixation cross on the computer screen, followed by three dots. The three dots are replaced by the fixation cross, and after a delay, a circle appeared on the screen. Subjects' task was to decide whether this probe item encircled one of the target items.

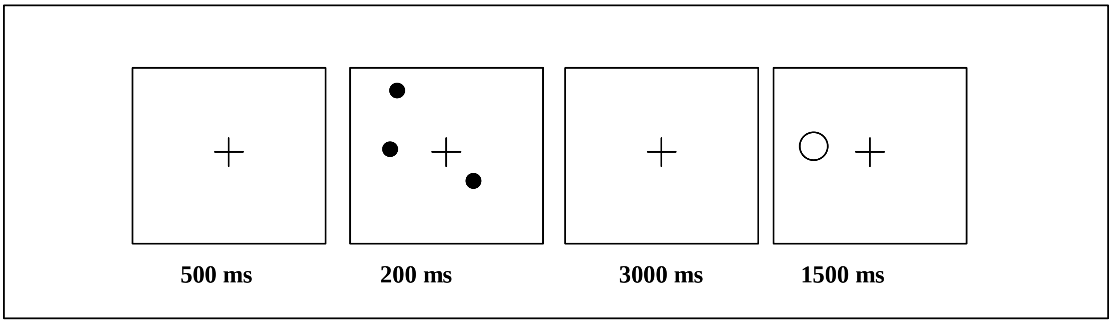


Figure 4: Spatial memory task

The task described in Figure 4 can be used as a secondary task during driving, since it requires drivers' visual attention for a short duration. Attention to the task can be achieved by using a short alerting tone. Task difficulty can be manipulated by increasing the target dots from 2 to 9, or manipulating the delay interval.

2.4.2.2 Tapping tasks

A selection of finger tapping tasks have been used in dual task paradigms to impair memory for spatial information. These include tapping a single location; patterned tapping of a 5 x 5 matrix, or tapping in a figure of 8 (e.g. Farmer, Berman, & Fletcher, 1986; Smyth & Scholey, 1994; Quinn & Ralston, 1986).

The effect of driving on tapping tasks can be examined by comparing tapping accuracy during different driving scenarios. Accuracy is monitored by

i. number of keys pressed
ii. conformity with the required pattern and
iii. maintenance of a regular tapping pace.

Care must be taken to ensure that finger tapping is not detrimental to the driving task as regards competition for manual resources, and foot tapping should be considered as an alternative. The advantage of tapping tasks is that an increase in workload can be introduced gradually by developing the task from simple to patterned to random tapping. Random tapping of keys has been equated to random number generation (Baddeley, 1966), and is thought to require high mental workload (Towse, 1998). A more detailed description of this task is provided in the next section.

2.4.2.3 Tasks imposing central load

The tasks described below are administered using any combination of auditory/visual display and verbal/manual response. This is especially beneficial since it allows an identification of the best combination for use in a dual-task paradigm during driving.

2.4.2.4 Continuous memory tasks

(CMT e.g. Veltman & Gaillard, 1998). In this task, target letters (A, AB, ABC, ABCD) are presented between non-target letters (E, F, G, H, I, J, K, L, M, N, O, P, Q, R, S, T, U, V, W, X, Y, Z). A button is pressed every time a target letter is presented, and pressed twice if the target letter has been presented more than once. Level of difficulty increases by increasing the number of target letters that must be remembered in any one chain of letters. Once a target is presented more than three times, the tally goes back to zero. The advantage of this task is that it can be presented both auditorily (see below) and visually. In addition, letters can be replaced by difficult to name shapes, for a more visually loading task, whilst difficult to name tones can be presented to assess drivers' memory for different sounds. Load can be assessed by using either a manual or a vocal response.

2.4.2.5 Paced Serial Addition Test (PASAT)

(e.g. Gronwall & Sampson, 1974; Channon, Baker & Robertson, 1993; O'Donnell, MacGregor, Dabrowski, Oestreicher & Romero, 1994). Used chiefly as a neuropsychological test, this task imposes a high cognitive load, and is thought to rely on central working memory resources. Digits are presented at a regular pace - anywhere from one every 1.2 seconds to one every 4 seconds. Subjects are required to add each digit to the one immediately preceding it, and state the sum out aloud. Therefore, the correct response for the digit sequence '*2,3,1,4,3,5*' would be '*5,4,5,7,8*'. The advantage of this task is that it can be presented both visually and auditorily, and that task difficulty can be manipulated by increasing the rate of presentation and the value of digits.

2.4.2.6 Random Number Generation (RNG)

This task was first introduced by Baddeley (1966) as a test of central executive resources, and has been used extensively in dual task studies and Alzheimer's disease patients. Subjects are asked to randomly generate a list of numbers from a specified list (e.g. any number from 1 to 10). Tests show that the 'Index of Randomness' (see Towse, 1998) is reduced if subjects are asked to perform a centrally demanding task during random number generation. As a control, subjects are also asked to utter the list in order. As outlined above, random tapping of keys has been used as an alternative to random number generation, where subjects are asked to tap a set of keys in either in a random or ordered fashion. The administration of this task during driving is reasonably simple, although manipulation of task difficulty is slightly more arbitrary.

2.4.2.7 The Sternberg Paradigm

Sternberg (1966; 1975) presented subjects with a list of items, and after a short delay, introduced a 'probe' item. Subjects' task was to decide if this probe item belonged to the original list. Results showed a linear relationship between number of items in the list, and subject reaction time, with an increase of around 38 ms for each additional list item (see Figure 5). The straight line is represented by the function: Time = 397 + 38s, where s is the size of the memory set. This paradigm can be adapted to test memory for letters or numbers, presented visually or auditorily. In addition, subjects' memory for difficult to name musical tones or difficult to label objects can also been tested.

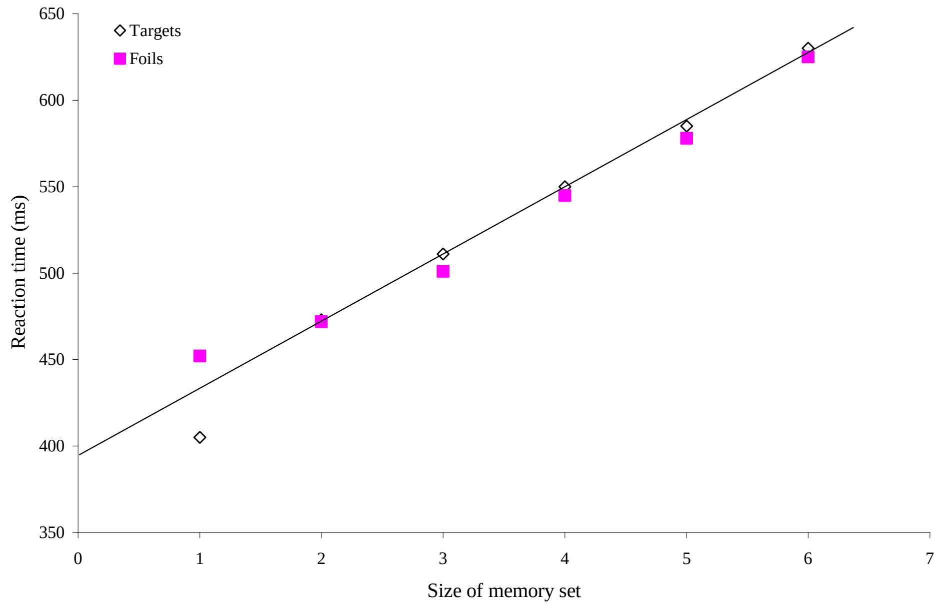


Figure 5: Time taken to recognise a probe item within a list.

2.4.2.8 Tone counting tasks

(e.g. Rantanen & Goldberg 1999). The degree of workload imposed by such tasks has been shown to influence the shape of UFOVs. Subjects hear one of three tones corresponding to the notes B, F and A, presented at frequencies of 493, 698 and 880 Hz respectively. This task is presented in two levels of difficulty: *Moderate*, where subjects are asked to respond to every fifth tone of low frequency, and *Difficult* where one key is pressed for the third low frequency tone, and a different key is pressed for the second middle frequency tone.

2.4.3 Overview

To summarise, all of the perceptual and cognitive tests described above are considered suitable tasks in assessing the interaction between driving and secondary task performance. The next section provides a short review of studies on crossmodal interaction and multimodal integration. It is hoped that the findings from these studies can be incorporated in some of the tasks described above, to investigate the effect of crossmodal integration in dual task performance.

2.5 Crossmodal interactions

Whilst most of the studies on dual task performance insist on the separation of tasks with respect to modality, our experience of objects and events in the environment frequently involves the amalgamation of visual, auditory, olfactory and tactile processes. Indeed research on multisensory perception has unveiled a strong influence of individual modalities on one another. Therefore, it is important to establish if workload within the driving environment can be reduced by employing a crossmodal approach in the presentation of information. A number of significant results in this field our described below.

The effects of vision on sound has been illustrated in classic experiments such as the McGurk effect (McGurk & MacDonald, 1976), where sound phoneme identification is determined by lip movements, and the ventriloquism effect (Bertelson, 1999) where sound localization is judged by visual events. Other examples include the influence of sound on perceived skin texture (Jousmaki & Hari, 1998), and modifying the taste of food and drinks by changing their colour (Zellner & Kautz, 1990). Many such phenomena occur because of the strong dominance of vision over other modalities (Batic & Gabassi, 1987), although there are cases when visual ambiguity is solved by the presence of auditory stimuli (e.g. Watanabe and Shimojo, 2001). In general, this ‘convergence’ of information from different modalities is assumed to be beneficial in conditions of uncertainty and ambiguity (Driver & Spence, 2000).

Integration of stimuli from different modalities is thought to be related to the temporal and/or spatial proximity of each stimulus to the other; which are then combined to provide information about some outside event. In recent years, research has devoted considerable time and interest to the crossmodal links in spatial attention between vision, audition and touch (see Eimer, 2001; Driver & Spence, 2000; Spence, Nicholls, Gillespie & Driver, 1998). These experiments typically measure reaction time to stimuli in the primary (e.g. vision) and/or secondary (e.g. audition) modality, presented to subjects in different spatial locations, e.g. to the left and right of the midline. Results suggest that in many cases, spatial attention to a stimulus presented in one modality ‘pulls’ participants’ attention to stimuli presented in other modalities. For instance, Spence et al. (1998) asked subjects to respond to tactile targets preceded (cued) by visual or auditory stimuli. Reaction time was found to be faster if targets were cued by auditory/visual stimuli in the same spatial location.

A similar finding was reported for responses to auditory and visual stimuli cued by tactile targets. Crossmodal integration has also been found to be influential in cases of unimodal *perception*. For example, Vroomen and de Geider (2000) report an improvement in the detection and localization of visual stimuli that were accompanied by a unique auditory stimulus presented within an auditory stream.

Recently, the advantage of crossmodal integration in the detection and perception of stimuli has been demonstrated for in-vehicle tasks presented during driving. Lui (2001) reports better performance and faster response times to information presented either auditorily or audio-visually, compared to when information was only visual.

Clearly, more detailed studies on the effect of crossmodal stimulus presentation within dual task studies are important, since the use of such paradigms may lead to reduced workload for drivers. The use of tactile stimuli within a crossmodal paradigm is especially recommended, since it is important to ascertain how attention towards this modality affects driving performance.

With respect to the HASTE project, and the tasks described above, the application of a crossmodal paradigm is most suitable within experiments that are recommended for testing attention and vigilance. For instance, it will be interesting to see how responses to the auditory stimuli presented in the ‘auditory detection task’ are affected by a brief presentation of visual or tactile cues. Furthermore, the effect of crossmodal presentation on the size of the UFOV also merits an investigation.

2.6 General summary

This chapter has outlined a number of suitable tasks which are designed to measure different levels of perceptual and cognitive ability in human operators. A list of factors which can control task difficulty are also proposed. Finally, the incorporation of crossmodal paradigms is recommended for some of the perceptual tasks, to compare the effect of unimodal and crossmodal secondary task presentation on driving workload.

3 Suitable secondary tasks

3.1 Introduction

This section of the report outlines the results of five pilot studies planned to assess participants' performance in tasks designed to test sustained attention and memory. While each of the tasks described below were performed as a single task in the laboratory, their employment as a 'secondary task' in combination with driving is anticipated within Workpackage 2.

The degree of attention and workload required by most IVIS varies enormously, as does the load these systems place on the driver's visual, auditory, motor/manual and higher cognitive processes. The purpose of the tasks described below was to separate the load placed on each modality/human information processing system as much as possible, whilst also controlling the level of difficulty imposed by each task. It was anticipated that by loading each processing system/modality separately, we would be able to examine the way in which each task type affects driver performance.

Visual attention and vigilance were tested using two visual search tasks, the design for which was based on Treisman's Feature Integration Theory (e.g. Treisman, 1988). The effect of increasing cognitive load was assessed using two variations of a Continuous Memory Task, and a new 'Keyboard Task' was created to examine participants' operation of an unfamiliar visual-manual task. Each task was presented at a variety of difficulty levels, by increasing the degree of information processing.

3.2 Participants

Sixteen volunteer participants (8 male, 8 female) were recruited for the five studies (Mean age 29.9 ± 13 years). Subjects were required to complete all five tasks, which were presented in a random order across participants. Volunteers completed a subject consent and information form at the start of the session, and received £10 at the end of the test session, which lasted around 60 minutes. Instructions for each task were provided before its commencement. In the following sections, the methods used and results achieved for each task will be described separately.

3.3 Testing environment

The experiments were carried out on a PC using the E-Prime software. E-Prime is a graphical experiment generator for Windows 95/98/ME E-Prime consists of a suite of applications to design, generate, run, collect data, edit and analyze the data. E-Prime includes: 1) a graphical environment allowing visual selection and specification of experimental functions; 2) a comprehensive scripting language; 3) data management and analysis tools.

3.4 Introduction to visual search tasks

According to Treisman's Feature Integration Theory (e.g. Treisman, 1988) the speed at which a visual target is identified within a display is affected by its visual similarity to other objects in the display. Visual search experiments have shown that unique features of a target object allow it to 'pop out' of the display, resulting in faster decision times. Difficulty in target identification will therefore increase as the non-target objects become more similar to the target in colour and/or shape. In addition, increasing the number of objects in a display is shown to increase reaction time to targets, but only when a target object must be recognised by a conjunction of features (i.e. colour and shape).
Based on the above theory, two visual search tasks were designed:

- Experiment 1 - Response to a target red Q amongst an array of green and red Os (see Figure 6).
- Experiment 2 - Response to an upward facing arrow, displayed amongst down, right and left facing arrows (see Figure 9).

3.5 Experiment 1: Visual search (target red Q)

3.5.1 Design and Procedure

Each display remained on the screen for 5 seconds or until subjects response, and varied in size (4, 8 or 16 objects) and difficulty level (Red Q amongst green Os, Red Q amongst red and green Os, Red and green Os only - i.e. no target condition). Each subject completed four blocks, with 104 trials in each block. Subjects were asked to press the keyboard number '1' key if they detected a red Q within the display, and the number '2' key otherwise. The displays were presented randomly within a block of trials.

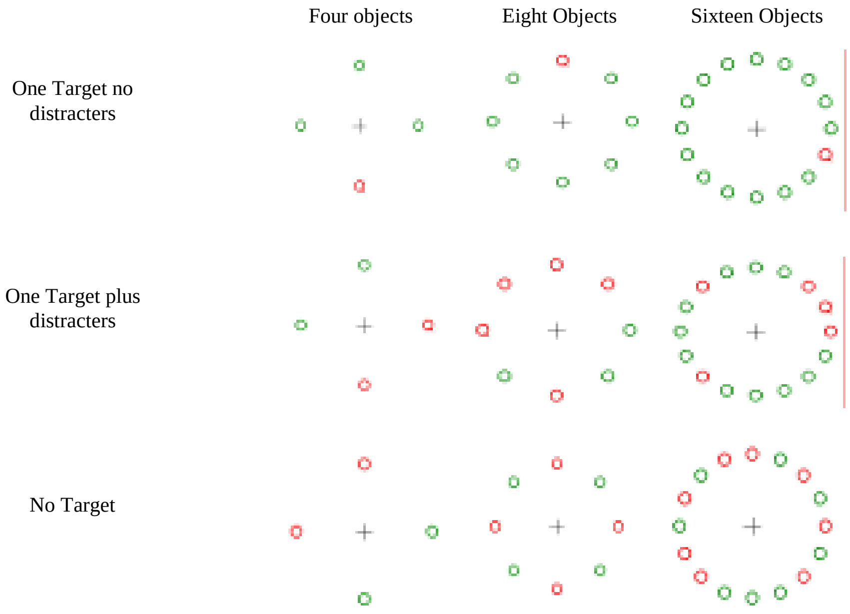


Figure 6: The displays used in visual search task 1

The number of correct responses and response reaction time (RT) for each difficulty level was compared across the three display sizes. It was anticipated that error and RT would be lowest when the red Q was placed among green 'Os' as the target would 'pop out' of the display. It was also predicted that error and RT would increase in the presence of red Os.

3.5.2 Results

The number of correct responses was not found to vary across the different levels of difficulty or across the three display sizes – with subjects performing between 95-100% in all conditions (Figure 7).

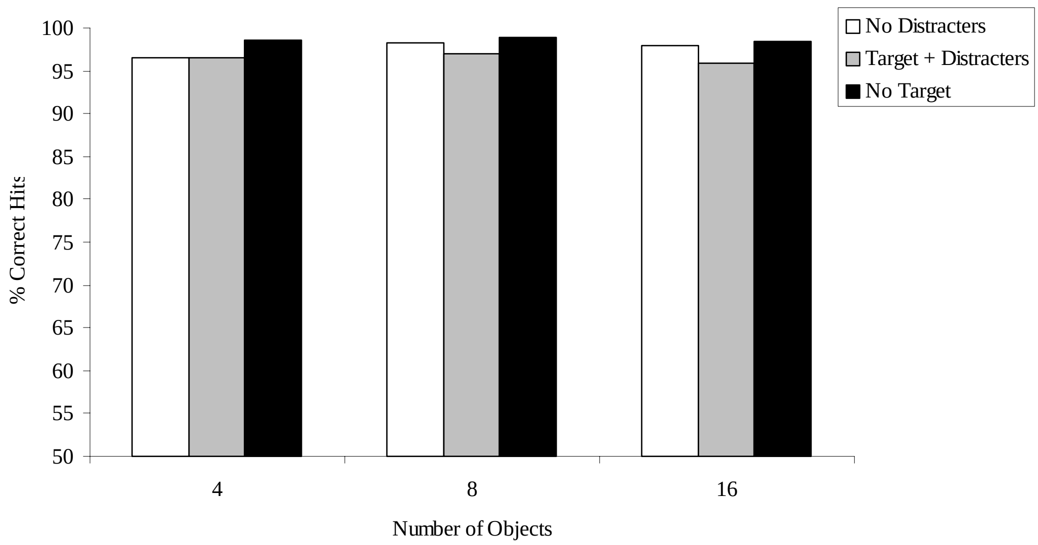


Figure 7: % correct by condition and display size

Mean RT to the target red 'Q' was found to increase in the presence of distracter red 'Os', and this increase was reliable for the 4 and 16 object displays (t (187) = 5.05, $p < .001$ and t (187) = 4.96, $p < .001$). Within the no distracter condition, mean RT increased when display size increased from four to eight objects, although this increase was not found to be reliable. Also, in the presence of distracters, the mean RT increased between eight and sixteen objects, although this difference was not found to be reliable (Figure 8).

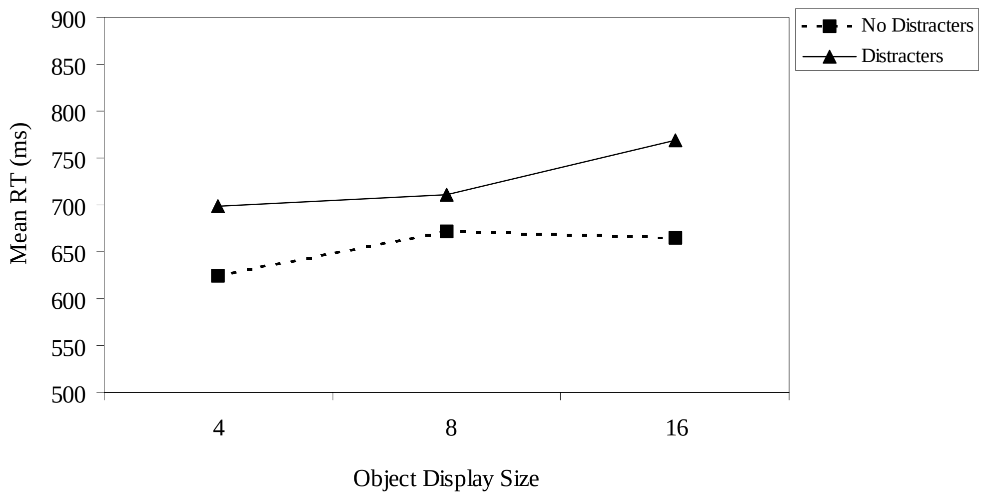


Figure 8: Mean reaction time to target Q

3.6 Experiment 2: Visual search (target up-facing arrow)

3.6.1 Design and Procedure

Each display remained on the screen for 5 seconds or until subjects response. Each display contained 16 arrows, and consisted of one of the following conditions:

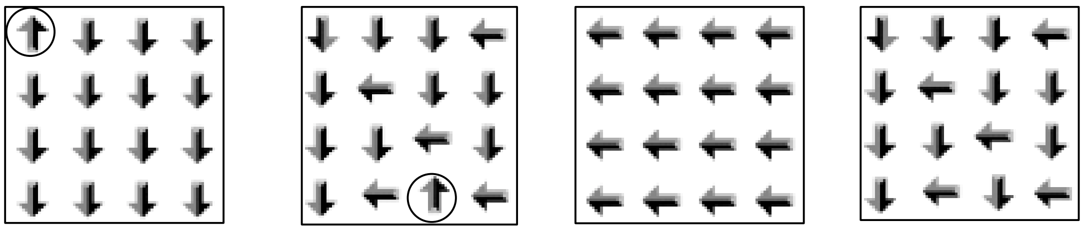
i. An upward facing arrow amongst 15 left, right or downward facing arrows.

ii. An upward facing arrow, amongst a mixed display of arrows facing up, down and right.

iii. A display of 16 up, right **or** downward facing arrows (no target).

iv. A display of 16 arrows facing up, down **and** right (no target).

Subjects completed four blocks of forty trials, and were required to press the number '1' key if they observed the target upward arrow, and the number '2' key otherwise. The displays were presented randomly within a block of trials.



(i) (ii) (iii) (iv)

Figure 9: The displays used in Experiment 2

Subjects' performance was assessed by calculating the percentage of correct responses and reaction time in each of the four conditions.

3.6.2 Results

The percentage of correct keys pressed was found to be highest when all 16 arrows were facing the same direction, and when the target upward facing arrow was present amongst 15 arrows which were all directed to the left, right or down. Error was found to increase in the mixed arrow condition, whether the target upward arrow was present or not. The presence of the mixed arrows prevented the 'popping out' of the target, increasing the number of errors (see Figure 9).

This confusion by the mixed arrows conditions was also reflected by the reaction time results which showed a reliable increase compared to when all arrows were facing one way, or when the target was amongst 15 arrows all facing the same direction (see Figure 11). Reaction time was found to be higher in the mixed arrows condition, regardless of target presence ($F(3,2682) = 579.56$, $p < .001$).

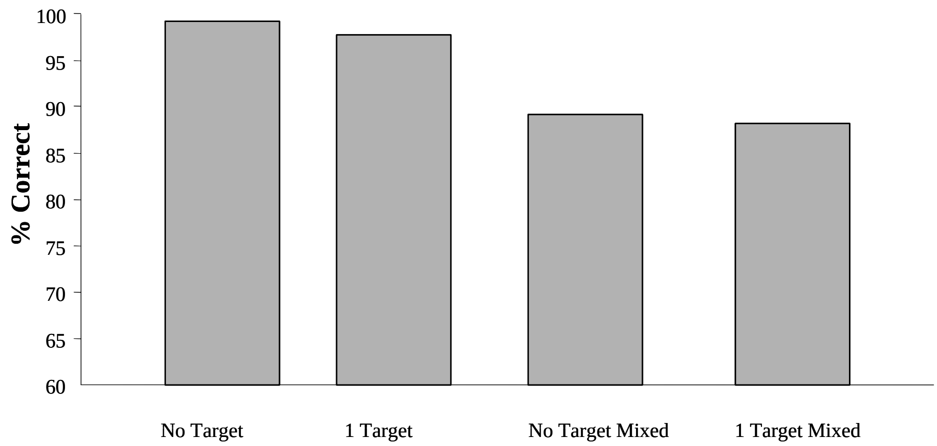


Figure 10: % keys pressed correctly

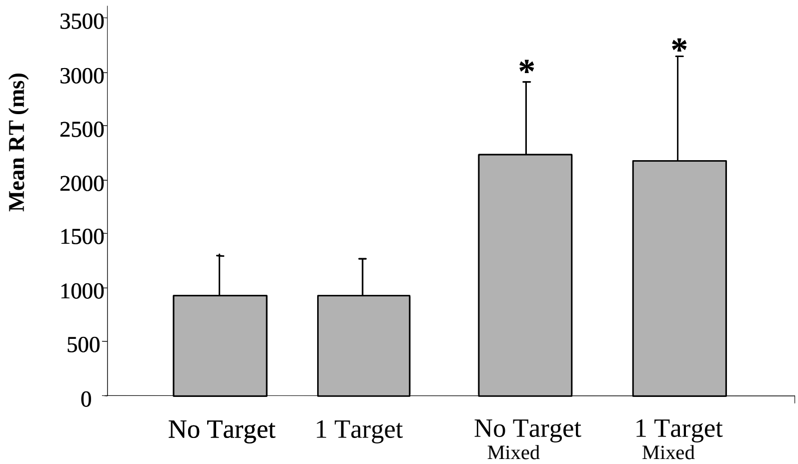


Figure 11: Mean reaction time by condition

3.7 Conclusions about the visual search tasks

Results of the visual search tasks described above generally concur with previously reported empirical findings. Subjects' reaction time to targets was found to increase in the presence of similarly featured distracter objects, whilst distracters also reduced the number of correct responses, at least in Experiment 2.

The use of one or both of the above tasks in the car can provide information about drivers' ability to distribute their visual attention between the visual search task and the driving task. It is anticipated that the number of missed responses; number of incorrect responses and reaction time will all increase when visual search is required in during driving.

In order to ensure compatibility of the visual search task with an IVIS, the tasks described above will require the following modifications:

Drivers will be alerted to the onset and termination of the visual search task with the aid of an auditory signal.

The visual search task will be presented in short bursts of 10-15 trials, across a variety of driving scenarios/road situations.

Displays will be presented at three different rates to investigate the interaction between speed of presentation, driving performance and accuracy in visual search.

Whilst many forms of IVIS require a rapid and simple response to displayed information, others can place higher levels of load on the driver, and may require the driver to remember information for short periods of time. An example of tasks that place such demands is described in the next section.

3.8 Introduction to continuous memory tasks

This experiment tested the effect of an IVIS that requires high levels of cognitive load and/or longer periods of sustained attention. Subjects completed two versions of the Continuous Memory Task (CMT) adapted from Veltman and Gaillard (1998). In Experiment 3, subjects remembered a series of target letters presented on the computer screen, and the particular strategy employed by each participant meant that this task required either visual and/or articulatory/ phonological memory resources. In Experiment 4, the letters were replaced by short bursts of sound, therefore testing subjects' memory for auditory information.

Since information provided by IVIS can be presented both visually and auditorily (together or in isolation), it was important to acquire baseline measurements of subject performance for both forms of stimulus presentation. This would allow subsequent assessment of the effect of loading each modality (visual, auditory) on driving performance.

3.9 Experiment 3: Visual continuous memory task (vCMT)

3.9.1 Design and Procedure

Subjects were asked to count the number of times a 'target letter' was presented on the computer screen. Letters were presented in succession, at a rate of one a second. The number of target letters increased from 1 to 4 in the course of the experiment. The target letters were the letters A, B, C and D and all other letters of the alphabet were used as non-target letters. Each experiment started with a practice block followed by 16 blocks of trial, in which a list of either 10 or 20 letters were presented in sequence. Subjects were required to press the number '1' key the first time they saw their target letter(s) on the screen, number '2' the second time and so on. Target letters appeared a maximum of three times in any one list, although subjects were not aware of this. It is important to note that the tally for each target letter was to be accomplished separately. It was anticipated that the number of errors would increase linearly with target letters.

3.9.2 Results

Memory for target letters was found to be best when subjects were only required to keep a count of one target letter (98% correct in the 10 letter list condition and 89% correct in the 20 letters list condition). Performance deteriorated when subjects were required to keep count of more than one target letter, although it did not vary between 2, 3 and 4 target conditions. Finally, performance was not found to vary between the two different list lengths (see Figure 12).

Mean reaction time to target letters was found to be reliably faster in the one target letter condition ($F(7,410) = 17.29$, $p < .001$ — see Figure 13). While reaction time to target letters was found to increase as the number of target letters increased from 2 to 4, this increase was not found to be reliable. Reaction time was also not found to differ for the two different list lengths.

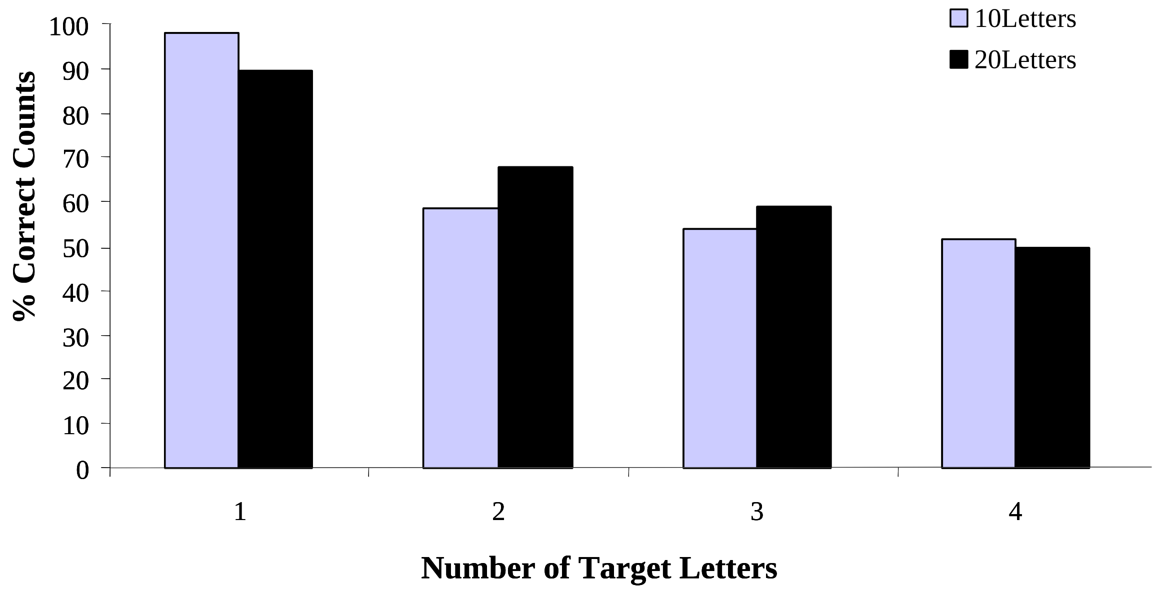


Figure 12: Performance in visual CMT

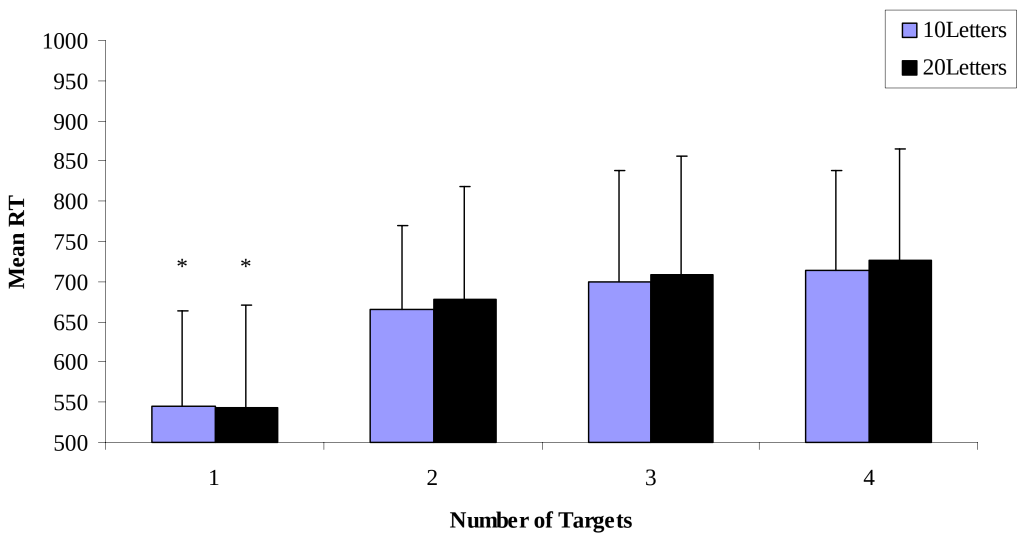


Figure 13: Mean reaction time to target letters by condition

3.10 Experiment 4: Auditory continuous memory task (aCMT)

3.10.1 Design and Procedure

The number of targets and list length presented in this version of the CMT were identical to those described in Experiment 3. However, the visual stimuli were replaced by bursts of sound that were roughly 320ms long. The target sounds were chosen for their obvious distinction from one another other, and represented the sound of a 'whip', a 'bus horn' and two versions of a 'buzzer' the non-target sound, which was randomly interspersed between the target sounds was a 320ms long burst of broadband noise. Stimulus intensity was controlled by the computer. Results examined the mean reaction time and number of correct counts to target sounds.

3.10.2 Results

As with the visual CMT, performance in the auditory CMT was quite good when subjects were required to keep count of one target sound (85% for 10 sound list, and 83% for the 20 sound list. However, performance deteriorated when more than 1 target sound was to be counted, although no significant difference in performance was seen between 2, 3 and 4 target sounds (see Figure 14). Also, as with the visual CMT, no difference in performance was seen between 10 and 20 sound lists.

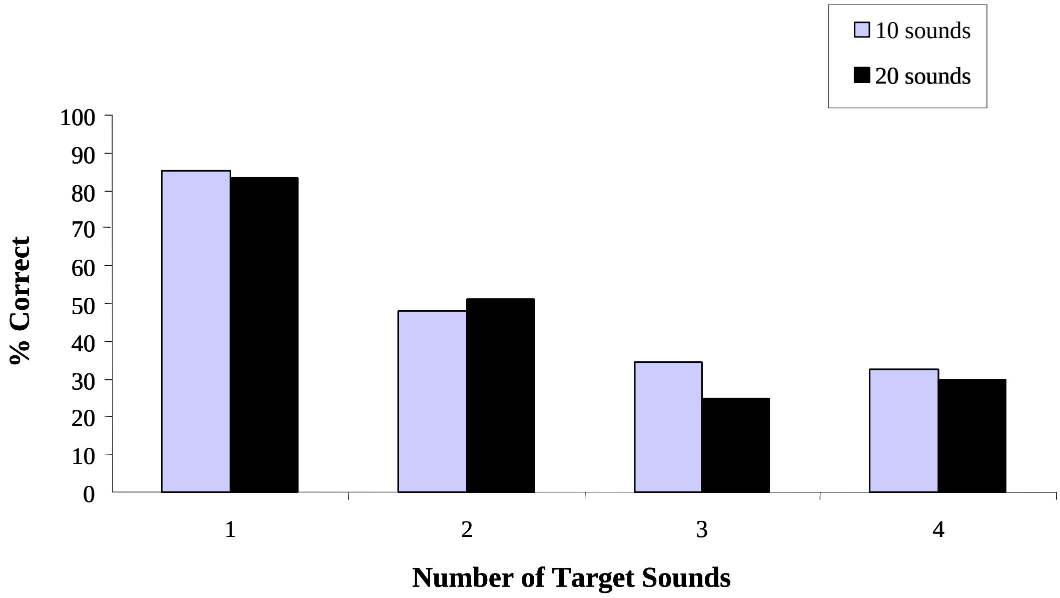


Figure 14: Performance in the Auditory CMT

Mean reaction time in the auditory CMT was found to be lowest in the one target sound condition, increasing significantly for the 2, 3 and 4 target sounds conditions ($F(7,346) = 4.07$, $p < .001$). However, post hoc t-tests failed to find a reliable difference in reaction time the 2, 3 and 4 target sound conditions (see Figure 15).

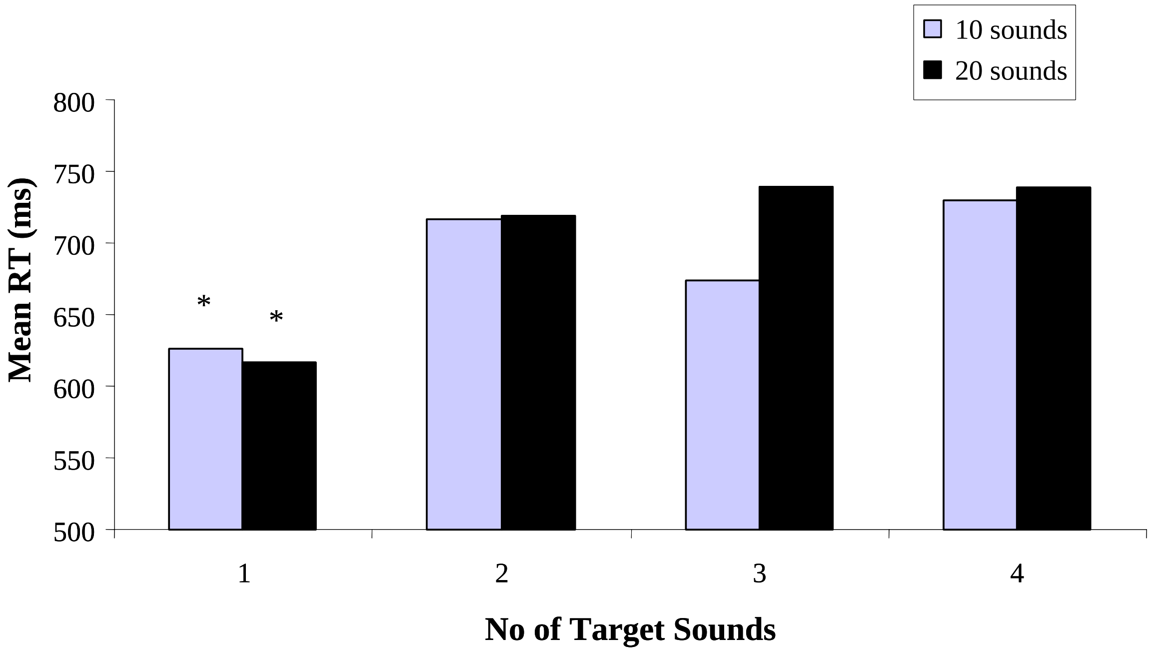


Figure 15: Mean reaction time to target sounds by condition

3.11 Conclusions about the continuous memory tasks

The pattern of results for both versions of the continuous memory tasks revealed a steady increase in reaction time and error as the load (i.e. number of targets) increased from 1 to 4 items. We believe that the absence of a significant difference between 2, 3 and 4 target conditions may have been due to the short length of the experiment and the relatively small number of subjects tested. We therefore aim to repeat the above experiments with an additional group of 16 subjects, in the hope of obtaining a more reliable difference in performance between the four target conditions.

In order to use the above experiments as a secondary task in combination with driving, we need once again to modify the tasks to some extent as described below:

i. The onset and termination of the visual CMT will need to be accompanied by an auditory signal to alert drivers towards the stimuli.
ii. Rate of stimulus presentation may need to be reduced to prevent unnecessary overload of drivers.
iii. As with the visual search tasks, the CMT tasks will be presented in short durations of 10-15 seconds, at regular intervals and during the different driving scenarios.
iv. Response keys to the task will be positioned in accordance with IVIS apparatus.

3.12 Experiment 5: Manual task

This task was designed in an attempt to quantify the visual-manual interactions drivers encounter with an unfamiliar IVIS, or an IVIS that requires a high degree of visual-manual-cognitive load.

3.12.1 Design and procedure

Subjects were presented with a new keyboard, on which the keys were replaced with unfamiliar Hebrew keys (Figure 16).

Figure 16: The keyboard presented to subjects in Experiment 3

For the first part of the experiment, the above keys were presented on the computer screen one at a time, and subjects' task was to find and depress the corresponding key on the keyboard. Each display remained on the screen for 5 seconds, or until the depression of a key. Subjects completed three blocks of 26 trials in this section of the experiment.

Following the first three blocks, three Hebrew characters were presented sequentially, at a rate of one a second. Subjects were required to remember the correct order of keys presented, and depressed the appropriate keys after presentation of the third display in the series.

3.12.2 Results

The results of this experiment are based on fifteen subjects, as one subject had difficulties performing the memory task. Subjects' performance in the last two blocks of section one is plotted in Figure 17. Results revealed an adequate level of performance for the detection of unfamiliar keys (around 64%), with a relatively small number of incorrectly depressed and missed keys.

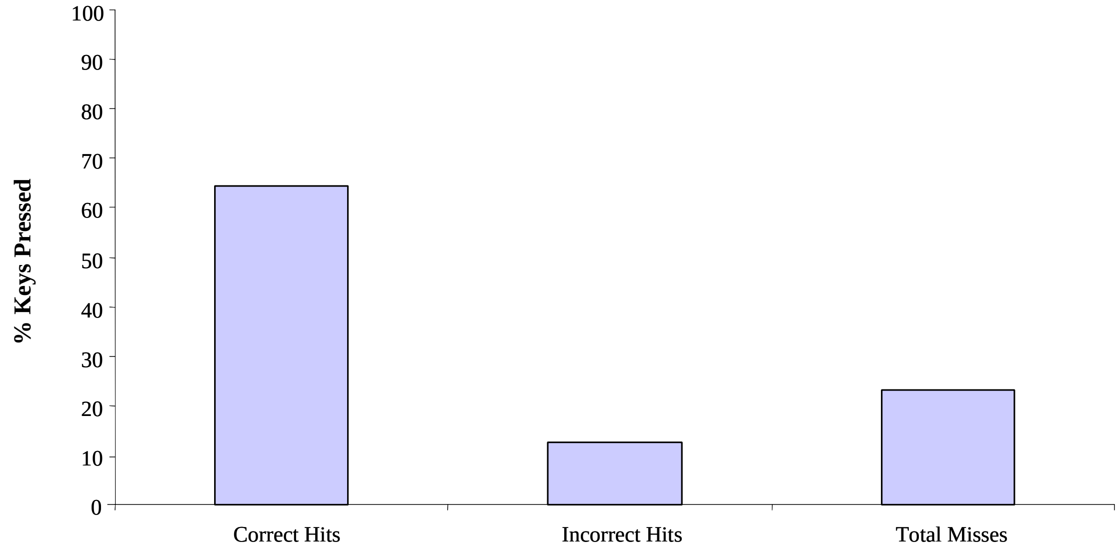
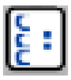
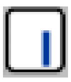
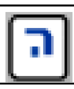
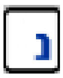


Figure 17: Response to the Manual task

Reaction time to each key varied greatly across subjects, although a sample of highest and lowest mean reaction times to keys is provided in Table 2.

Figure 18 shows subjects' performance when asked to remember the identity and order of presentation of three Hebrew characters. On average, subjects remembered the identity and order of presentation of all three keys around one third of the time. There were very few trials in which subjects failed to depress any keys, and the number of incorrectly depressed keys was also quite small.

Table 2: Reaction Time to a sample number of keys

Hebrew Key	English Key	Mean RT (ms)	Std RT
	**q**	**1462**	**454**
	**i**	**1567**	**506**
	**u**	**1593**	**422**
	**b**	**3426**	**805**
	**t**	**3443**	**792**
	**c**	**3809**	**733**

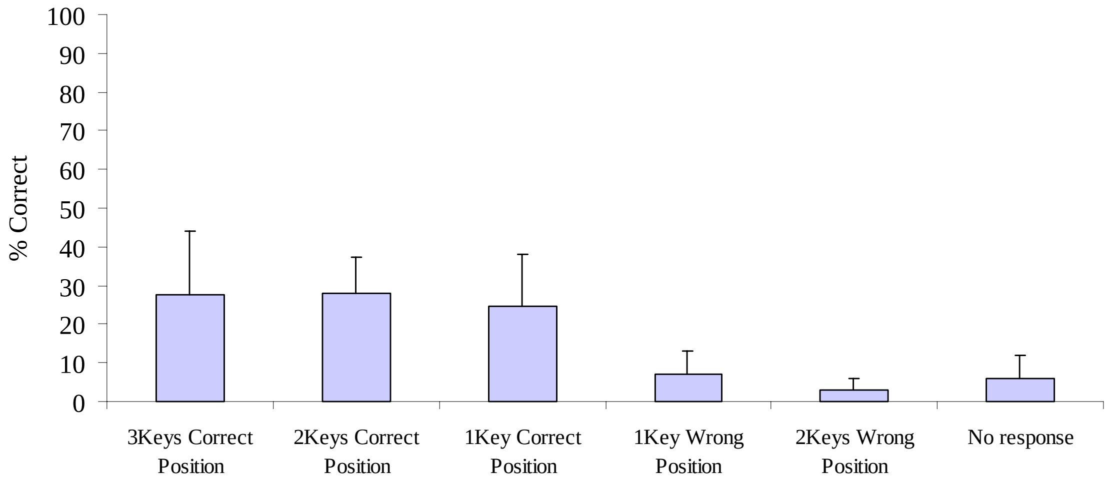


Figure 18: Memory for order of presentation of three Hebrew keys

3.12.3 Summary and conclusions

Participants were relatively competent in locating the position of the unfamiliar Hebrew keys, although the speed at which keys were located was somewhat slow (1880 ± 1300ms).

Memory for three unfamiliar keys was also rather poor, with performance fully accurate on only a third of the trials. This is in contrast to memory for familiar visually presented words, numbers or letters, which is approximately 7 items (e.g. Miller, 1956).

One reason for such poor performance in the keyboard task may have been the short duration of the task, and hence the lack of practice by subjects. This short duration was partly enforced by a desire to complete all five pilot studies in one (hour long) experimental session. We therefore hope to increase the practice time of this experiment in a forthcoming study, to further investigate competency in such highly loading visual-manual tasks.

Finally, if this task is to be used in conjunction with driving, the visual display may need to be accompanied by an auditory signal to divert drivers' visual attention from the driving task. Also, as with the CMT tasks, the keyboard may need to be replaced by a more appropriate response display.

3.13 Conclusions from the pilot studies

The pilot studies described in this section of the report have provided a good range of baseline results for a series of tasks requiring different levels of perceptual and cognitive demand. Whilst each of the experiments was accomplished in isolation for this stage of the project, it is anticipated that all tasks will be performed in combination with driving. The kind of load placed by each of the chosen tasks (i.e. simple visual attention, memory for visual/auditory information, visual-manual co-ordination/memory), is one that is usually required by different forms of IVIS, either in isolation or collectively with other loads. It is therefore hoped that examining the effect of each of the above tasks on driving performance will provide a more controlled measure of the way in which an IVIS interacts with driver performance.

4 Scenario Definition

In this work *a scenario is defined to be a snapshot view of a possible future set of circumstances*. The scenario does not describe the possible actions (policies, funding etc.) which are needed to achieve the described future set of circumstances. Those actions can be described later in this project as implementation strategies.

The terminology used is as follows:

<u>IVIS</u> – in-vehicle information and communication system.

<u>Parameter factors or groups</u> (e.g. driver, road infrastructure etc.)

<u>Parameters</u> (for example age) and parameter values either qualitatively (young, old etc.) or quantitatively (under 25 years, 65 years and older etc.).

<u>Key parameters</u> – parameters that are most critical (in one scenario combination or as an individual parameter).

In addition to parameters, the scenario might include some variables. Variables are parameter values that can adopt a range of values, which are considered, as an approximation, not to affect other aspects of the scenario. Parameters such as “penetration rate of IVIS” or “frequency of using the IVIS” are often taken as variables (over a limited range) in larger scale modelling studies. However, in this scenario the examination is done in individual level and therefore the variables are not taken into scenario definition.

4.1 Main aim of the scenario work

Workpackage 1 aims to determine the methods and metrics to be used in the studies involved. This requires the selection of parameters to measure driving performance, work/task load and situation awareness and also gives consideration to the driver types to be tested. This will lead to the determination of scenarios to be used for the testing of the systems in Workpackage 2. Ergo, in the present Workpackage a number of choices, or definitions of parameters involved are to be assessed. The following approach is selected. Scenario building is basically prepared, i.e. the scope is defined. A scenario is defined to be a snapshot view of a possible future set of circumstances. The scenario does not describe the possible actions (policies, funding etc.) that are needed to achieve the described future set of circumstances. Scenarios are created in such a way that predictions can be made about the safety effects of in-vehicle information systems. Scenarios define the test regime (traffic environment etc.) where for example the drivers’ cognitive workload, driving performance and traffic safety are evaluated. The main purpose of the scenario definition is to identify traffic scenarios in which safety problems with an IVIS are likely to occur and to point out the most important parameters in those scenarios. The scenarios do not define the experimental protocols, only the test regime. In addition to defining scenarios, the IVIS need to be classified, defined and selected.

4.2 Scenario parameter groups and parameters

4.2.1 Usual parameter groups in scenario

One way of describing the set of circumstances in which the safety problems with an IVIS are more likely to occur, is defining the critical parameters of listed factors in traffic situation. This is called scenario definition. The parameter groups used in this scenario building are shown in (Figure 19);

- Driver
  - driver
  - driver's state (of mind)
- Driver's task
  - manoeuvres
  - occasion
- Vehicle (complexity of car control)
- Road infrastructure
- Traffic conditions
- Environmental conditions

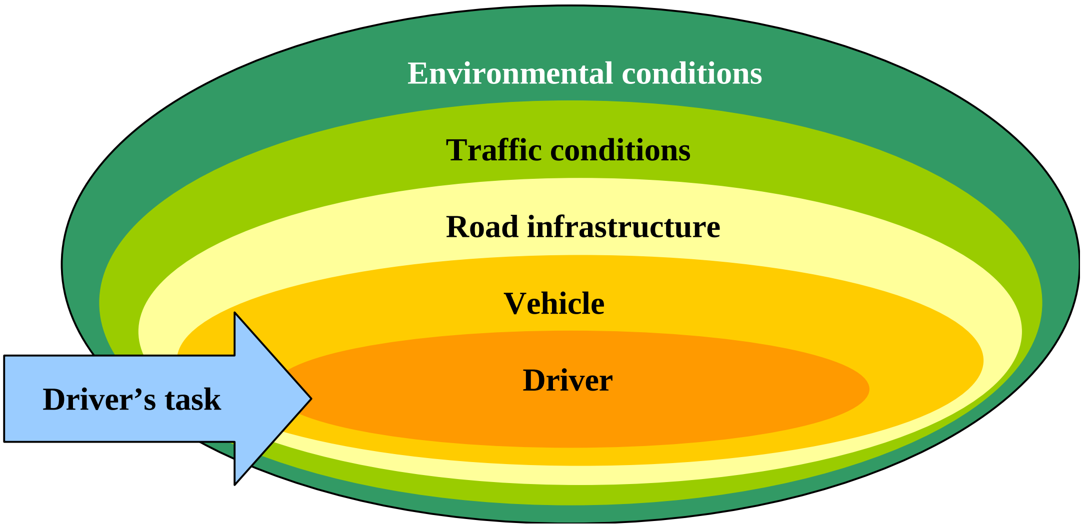


Figure 19: Usual factors in traffic scenario

To identify the overall complexity of traffic situation, in addition to the defined traffic scenarios, the in-vehicle information and communication systems (IVIS) need to be defined in a systematic and comparable manner. This is done later on this project by listing and defining different IVIS parameters and variables. When talking about in-vehicle information and communication systems (IVIS) both the device and the service are included. We are focusing in in-vehicle information and communication systems that only provide driver with information and do not intervene directly to the driving task.

Each parameter is described with parameter value. The parameter can be defined as value (e.g. number etc.) or content (qualitatively vs. quantitatively). Parameters may be descriptive in nature (e.g. driver's special needs), have a specific value (e.g. carriageway width: 3.7m), or be one of a limited set (e.g. male/female).

4.2.2 List of all possible parameters

1. Driver and driver state

- Age
- Gender
- Driver type (professional etc.)
- Special needs (disabled person etc.)
- Experience of driving
- Experience with IVIS
- Experience with technology/ other mobile systems
- Vigilance
- Health
- Medication

2. Vehicle

- Vehicle type – Cars, trucks
- Transmission (manual, automatic)
- Automatic control systems (ABS, TCS etc.)
- Other devices/ systems present (radio, RDS-TMC, other IVIS etc.)
- Ownership or experience with the vehicle - Rented/borrowed

3. Infrastructure

- Road type – urban, rural, motorway
  - *Urban, rural, motorway*
- Lanes
  - *Number of lanes*
  - *Width of the lane (narrow lanes)*
- Speed limit (50 km/h, 100 km/h etc.)
- Traffic guidance - signs and road markings
  - *Type of the guidance (information, mandatory)*
  - *Placing (horizontal and vertical)*
  - *fixed signs, VMS*
- Junctions, type and priority
  - *Type of junctions (T-junction, at-grade, grade-separated, ramp, roundabout etc.)*
  - *Density of junctions (spacing between junctions)*
  - *Priority (right-hand rule, yield, stop, signal control, etc.)*
- Pedestrian facilities
  - *Pedestrian walkways (composite with carriageway, separated)*
  - *Crossings with and without signal control, at or outside junctions*
- Continuity of the road infrastructure
- Additional parameters
  - *Pavement*

4. Traffic conditions and hazards

- Other road users
  - *Vehicle(s) (density)*
  - *Vulnerable road users (no, low density, high density)*

- Unexpected events, special hazards
  - *Stopped vehicles, animals, obstacles on roadway, etc.*

5. Environmental conditions
- Visibility (fog, darkness etc.)
- Slipperiness (friction)
- Lightning (outside, inside the vehicle)
  - *Daylight, twilight, dark (no lighting/lighting)*
  - *Bright sunlight (glare)*

6. Driver task
- Manoeuvres (turning left-right etc.)
- Familiarity of the route (familiar/ unfamiliar)
- Urgency of the trip
- Travelling company/ other people in vehicle

4.3 Selected critical parameters (key parameters)

This section describes in more detail the most important set of circumstances in which safety problems with an IVIS are more likely to occur, i.e. critical scenario parameters. In addition to this, the background information (reasons why the parameter is defined to be critical) are described.

The following parameters were selected to be the **most critical parameters** when defining set of circumstances in which safety problems with an IVIS are more likely to occur:
- ♦ **Driver age with a focus on the older driver**
- ♦ **Driver vigilance in the form of boredom resulting in low vigilance**
- ♦ **Road type as urban, rural and motorway environment**
- ♦ **Junctions as a parameter of road infrastructure**
- ♦ **Pedestrian facilities as a parameter of road infrastructure**
- ♦ **Other road users as creating the possibility of crossing patterns**
- ♦ **Special events as potential hazards**

In addition to these most important critical parameters, on page 47 some interesting additional parameters are discussed. These parameters are as follows;
- Experience of driving and novice drivers
- Driver's medication
- Other in-vehicle information systems in vehicle
- Driving lane (number of lanes, width etc.)
- Weather – driving conditions

4.3.1 Justification for inclusion of elderly drivers

In HASTE, the tasks we are concerned with can generally be thought of as dual tasks. In the literature there is a clear age effect for performance on dual tasks in general and specifically on divided attention tasks. Recent efforts have been directed towards measuring age differences in divided attention. This is because the difficulties experienced by older and cognitively impaired drivers, especially at intersections, may come from a failure to divide attention between the competing relevant stimuli present in this context.

Several studies suggest not only that divided attention capacity is reduced in later life, but that the decline in this function is greater both when motor integration skills (such as operating a vehicle) and when active visual search in the periphery are required (Ponds, Brouwer & van Wolffelaar, 1988; Brouwer, Waterink, van Wollffelaar & Rothengatter, 1991; Brouwer, Ickenroth, Ponds & van Wolffelaar, 1990).

Ponds, Brouwer & van Wolffelaar (1988) used a dual task in a simple simulator. The simulator consists of a computer screen with an steering wheel. The primary task was a visual tracking task within a driving scenario, with the participant aiming to keep the car within its lane boundaries while gusts of lateral wind attempt to push it outside. The secondary task was a visual divided attention task that required the participant to count the number of dots randomly presented within a well-defined area of their visual field and decide if 9 or any other number of dots were present. The authors report a significant decrement in dual task performance in old compared to young and middle-aged persons, largely due to poorer performance on the primary task.

Brouwer, Waterink, van Wolffelaar & Rothengatter (1991) replicated this study taking into account the motor skill integration ability of the driver. The secondary task used a verbal response mode instead of the motor mode (push of a button). The results were similar, but the age differences in divided attention were more marked when the motor rather than verbal response mode was employed. This is consistent with other evidence from the literature suggesting that ageing affects the integration of motor skills (Korteling, 1991). This study suggests that the integration of motor skills may be partly responsible for decrements in divided attention performance in tasks with motor components (such as driving).

One tool for assessing the ability to divide attention is *Multi*CAD, a test that uses a video recording of a car following situation. It first determines the angular motion sensitivity of a participant and then measures the performance on a visual divided attention task. It evaluates the ability to divide attention between a primary task (in this case detecting when the lead vehicle brakes, and then brake appropriately) and a secondary task (detecting peripherally presented safety-relevant vehicles and pedestrians who will intersect with the driver's own vehicle). The divided attention stimuli are presented either at 15^o or at 30^o of visual field.

*Multi*CAD was used by Janke & Eberhard (1998) in a study of 82 older drivers (of which 26 with probable cognitive impairment). *Weighted error scores* obtained in an on-road driving test (the number of errors weighted by their severity) were correlated with the proportion of errors and the mean time-to-braking in *Multi*CAD for the trials where the divided attention stimuli were located at 15^o and at 30^o of visual field. Only the proportion of errors in the 15^o visual field trials was significantly correlated with the weighted on-road error scores. The cognitively impaired drivers had a higher (though not significantly so) proportion of errors on all trials compared to the unimpaired subjects.

Considering the issues regarding dual task performance and divided attention performance decrements with age, it seems that this driver group is crucial for the evaluation of the issues we research in HASTE.

Specifying who is an 'old driver' can be difficult. It can be done by age cut-off (i.e. any driver over a certain age is considered an old driver) or by performance (i.e. someone whose performance on a driving (or other type of) task conforms to patterns previously observed and considered characteristic of older drivers). It is assumed that this conceptual difference is not very important for HASTE at this stage, so perhaps age-based definitions of old drivers are best. Even within this category, a distinction can be made between old in terms of performance on various psychometric tasks versus old in terms of accident risk. For the latter, it is largely thought that 75 years old is a definite point when accident risk increases dramatically. For the former, various authors use different cut-off ages and it is assumed that 60 is a good cut-off for the reasons described below.

Defining an age cut-off between middle-aged and old drivers is artificial. In previous research, this was considered to be 50 (Wierwille, 1990), 55 (Sixsmith, 1990), 60 (Barham, Alexander, Ayala & Oxley, 1993) or 65 (Carr, Jackson, Madden & Cohen, 1992). Furthermore, the elderly group is being sub-divided into *younger-old* drivers (up to 74 years old) and *older-old* drivers (75 and above). In deciding the cut-off point to denote an old driver one needs to consider both when the physical and cognitive decline attributable to ageing becomes significant (compared to younger persons) and when it begins to affect the ability to drive. In terms of cognitive decline, age associated memory impairment (AAMI) is reported to have a prevalence of 35% among adults over the age of 50 (Baddeley, 1996). Planek (1981) considers that 55 is the age at which the general decline due to ageing starts influencing the driving ability (Burns, 1997). If age-related decline begins to influence driving ability around the age of 55 then the effects of it are likely to be minor for some years yet. Therefore 60 years of age seems a better cut-off than 55.

In HASTE, we need to first clarify if psychometric performance or accident risk should guide our choice of age cut-off and then choose an appropriate age, or else go for an age cut-off somewhere between the ages suggested by the two approaches detailed above. The latter option would hinder comparisons between data obtained in HASTE and data from the previous literature, but this may not be a very important consideration for our project.

4.3.2 The effects of vigilance

Although many definitions of boredom have been offered over the years, boredom is usually associated with feelings of increased constraint, repetitiveness, unpleasantness and decreased arousal (Geiwitz, 1966). Hill and Perkins (1985) proposed a model of boredom with cognitive, affective and psychophysiological components. According to this model, individuals seek stimulus variety and if faced with repetitive, unchanging stimuli, they become bored and even frustrated. Moreover, this process is moderated by cognitive and affective factors. Consequently, boredom is only perceived if one construes the situation as monotonous. (Scerbo, 2000).

Depending on the driver's mood, personality, needs, capabilities, but also the level of vigilance, the driver performance fluctuates over time. Also fluctuating over time (as far as the moving driver is concerned) are the environmental demands on the drivers. An example of the two combined is presented in Figure 20 (Shinar, 1978 – adapted from Blumenthal, 1968).

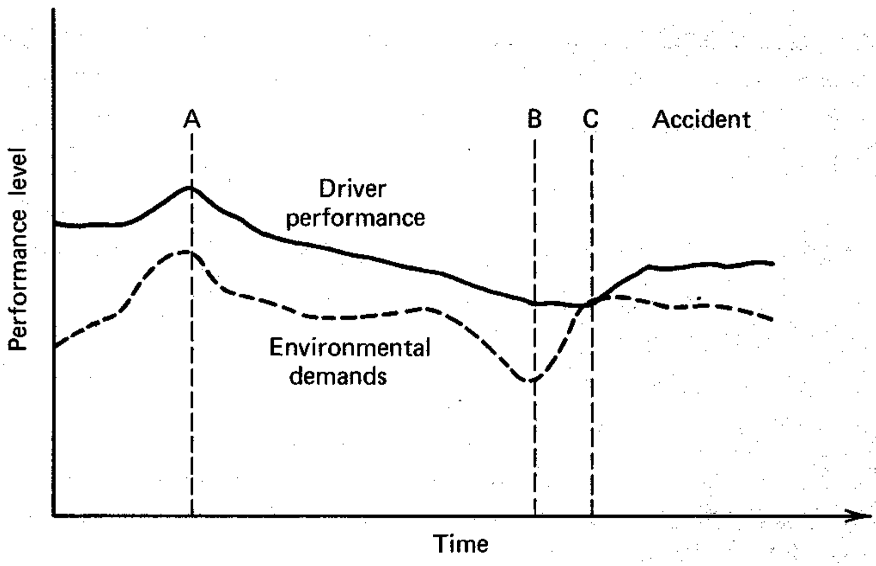


Figure 20: Combination of driver performance and environmental demand on the driver

Accidents are prevented as long as the driver's performance level is maintained above the environmental level. Thus in congested high-speed driving the environmental demand may be high, but the driver's information processing capabilities are probably more focused on the driving task (point A in Figure 20). Conversely when the demands are low, the driver may allocate much of the information processing capacity to non-driving task (point B), such as IVIS. A characteristic common to most emergency situation is that they place high demand on an unprepared driver – leading to situation depicted by point C in which the two curves cross each other and an accident results. (Shinar, 1978). Therefore also vigilance as boredom in vigilance is selected to be in HASTE one of the critical parameter.

4.3.3 Road types

The argument for including all types of road can be supported using a variet of methods.

4.3.3.1 Statistical-epidemiological argument

Based on chance alone, one would expect a distribution between accidents on urban roads, rural roads and motorways that is equal to mileage on these roads. However, statistics show that this is not the case. For example, UK statistics show an accident distribution of 51%-23%-26% between these roads, while mileage is highest on motorways. In the Netherlands private car accidents with injuries and specifically fatal accidents are predominant outside built-up areas, in particular on A-roads with a speed limit of 80 km/h, not on motorways (CBS, 1999).

4.3.3.2 Speed argument (speed accuracy trade-off)

The road infrastructure of a motorway is designed for high speed travelling (120 km/h). This is not the case for A[1] roads and urban roads. Nevertheless, the relation between distance covered and impact of a crash are both linearly related to driving speed. Thus, a driver

[1] An A road is defined as a rural road (i.e. outside built-up area) with a speed limit of 80 km/h.

distracted by an IVIS will cover more distance while distracted, and if he or she is heading for a congestion on the motorway the (speed) impact will be very high. On an A road, driving speed may be lower, but whilst steering inaccurately one may hit a tree or 'meet' a meeting car, resulting in a much higher impact than on the motorway where collisions usually are front-tail, notwithstanding the higher allowed speed. The combination of lane control and speed are thus most important on urban and A roads.

4.3.3.3 Unique characteristics of different road types

Urban roads — High informational load caused by the presence of all types of traffic participants (pedestrians, cyclists, etc.). Complex non-standardised road trajectories. Visual occlusion by buildings. Journey duration is in general short (< 1 hour).
Rural roads — Partially standardised road environment. Presence of slow traffic (farming vehicles) possible. Meeting traffic. Journey duration intermediate (30 min – 2 + hours).
Motorways — Highly standardised and thus predictable road environment. Driving conditions for longer journeys can reduce alertness because of monotony (De Waard & Brookhuis, 1991, Feyer et al., 1997). Travel time is longest, up to several hours (Parasuraman, 1986; Fell & Black, 1997, Horne & Reyner, 1999).

4.3.3.4 Definition of urban, rural and motorway environment

Road type definition is based on speed limit and/or road function. There are also differences between countries. Most similar between countries are motorways. Roads are categorised on the basis of:

- Speed limit
- Function
- Lane width
- Number of lanes per direction
- Presence of other types of traffic participants

Urban roads

There are different types of urban roads; from low to higher speed limit these could be characterised as

- 'woonerf' (the famous Dutch export article) where all traffic participants mix and the speed limit is 15 km/h
- area entrance roads; in the Netherlands and Germany these roads have a limit of 30 km/h or 50 km/h; narrow roads
- urban through roads, high traffic volume urban roads, speed limit 50 km/h (sometimes 70 km/h)

Differences in functionality also imply differences in optimal road architecture, in the associated driving behaviour, and the presence of other types of road users (e.g. pedestrians, bicyclists).

Rural roads

In principle these are all roads outside built-up areas. Speed limit can be 60, 80 or 100 km/h. Meeting traffic is common.

Motorways

Divided lanes, at least two lanes per direction plus emergency shoulder. Speed limit in general 120 or 130, with the exception of Germany. Lane width 3.6 – 4 metres.

The following three road types can be found in most European countries and differ on above listed criteria. Ideally, these types of roads should be included in evaluations:

Table 3: Road type characteristics and distinctions

Road Type	*Speed limit (km/h)*	*Total no. lanes*	*Delineation*	*Lane width*	*Road function*	*Other traffic*
Urban	50	1-2	Not necessarily	< 3 m	Local, urban transport	Cyclists, pedestrians, mopeds
Rural	80	2	Centre line	± 3 m	Interurban transport	
Motorway	120/130	4+	Centre + edge line	3.6-4 m	Long-distance transport	

4.3.4 Road infrastructure: junctions and pedestrian facilities

4.3.4.1 Junctions

Junctions are usually more hazardous than the road sections between them, i.e. links. On the main roads the accident rate (number of accidents per million vehicle kilometers) of injury accidents at junctions is on average 2.5 times higher than on links. (Kulmala, 1995). This is the case, since crossing and entering in intersection is one of the most difficult tasks of driving. It is relatively short period of time during which lot of information have to be perceived and processed, decisions have to be made and rather complex car handling have to be performed. (Hyden and Draskoczy, 1992). The accident risk at junctions is affected by different junction characteristics - for example by traffic volumes, speed limits and junction type.

4.3.4.2 Pedestrian (and bicycle) facilities

The pedestrian accident risk (number of injury accidents per kilometers) has been estimated to be three times higher and bicyclists' accident risk even six times higher than accident risk of drivers. Approximately 70% of the injury and fatal pedestrian accidents occur at intersection, where there is possibility of crossing path with vehicles. Overall the vulnerable road user accident data indicates that more complex traffic situation (intersections and urban area for example) where divided attention is needed are the most severe to vulnerable road users. Especially in urban area where it is very hard to separate the route of different road user groups (vehicles vs. vulnerable road users), but where the density of vulnerable road user is high, driver's full attention to primary task – the driving and interaction with other road users – is crucial. (Elvik et al., 1989). The distraction resulting from an IVIS may lead to degraded driving performance. Therefore we may expect to see the distraction provided by an IVIS magnified, as it were, in the risk that is handed down to these road users.

4.3.5 Other road users: possibility of crossing paths

In addition to road infrastructure, the possibility of accidents involving more than one party (i.e. non-single-accidents) is greatly affected by the number of other road users in traffic situation. Although the road infrastructure planning defines the possible crossing paths with other road users, the actual appearance (density) of other road users defines the potential conflict points (interaction with other road users) and therefore accidents in specific traffic situation. It is also quite to be expected that the risk of rear-end accidents increases with increasing volumes, as the higher the volume the more of the mileage is driven in platoons with shorter headway (Kulmala, 1995).

4.3.5.1 Car-following situations

Rear-end collisions are the result of inadequate car-following performance. A car-following situation exists when headway is below the 4-5 s range. This means that a few glances of average duration away from the leading vehicle would already use up much of the time available for action, would the leading vehicle suddenly decelerate. For this reason, the car-following situation is one in which an interaction with the IVIS is to be expected. That is, an IVIS that requires visual inspection of a display can be expected to be specifically leading to degraded driving performance in the car-following situation. This is the reason for including it in the scenarios.

4.3.5.2 Encounters with oncoming vehicles

The same reasoning applies with respect to oncoming vehicles on non-motorways. Here it is to be expected that the more frequent encounters there are, the more serious it is that attention is withdrawn from the tasks of (a) observing what the opposing vehicle is doing and (b) staying in one's own lane. Thus, an interaction may be expected to occur when the IVIS presents information of a visual nature in these situations, in the sense that the effects of the IVIS gets relatively worse the more frequent these situations are.

4.3.5.3 Vulnerable road users

The distraction resulting from an IVIS may lead to degraded driving performance. Not only can this pose a risk to oneself, but it may also endanger other people. Vulnerable road users are a particular case, since they are to a large degree dependent upon the behaviour of motor vehicles towards them. Thus, we may expect to see the distraction provided by an IVIS magnified, as it were, in the risk that is handed down to these road users. For this reason we propose to include at least one situation in the scenarios in which this could become manifest, which is a pedestrian crossing etc.

4.3.6 Special events — road type related potential hazards

As stated earlier, a characteristic common to most emergency situation is that they place high demand on an unprepared driver. Another type of emergency situation is problems with dividing attention in complex traffic situation. Below, some typical potential hazards on specific road types are described.

4.3.6.1 Urban roads

In urban roads the interaction with other road users is often continuous – the driving paths cross constantly and the density of other road users, both vehicles and vulnerable road users (pedestrians and bicycle) is often quite high. In this kind of environment, the drivers needs continuously to divide his/her attention among various objects of interest. In the visual domain, in driving we typically focus most of our attention on the road area ahead of us. However, especially in urban area, quite large amount of attentional capacity must be allocated to be periphery since critical events happening immediately off the roadway may also influence our behaviour (traffic entering your road/lane etc.) (Shinar, 1978). Examples of these sudden unexpected events typical in urban environment are 1) pedestrians crossing, 2) stopped or parked car pulling out in front and 3) difficulties to detect a bicycle which is going straight and therefore crossing the driving path when a vehicle is turning .

4.3.6.2 Rural roads

The mechanism of the single vehicle accidents on the rural roads is similar to the situation described in motorway. However, the monotony is not very often so strong on rural roads, while the variance of environmental and traffic conditions is higher (except if traffic density is very low). Therefore there is a much stronger need to adjusting speeds to the momentary situation. However, rural intersections are the sites on the rural network, where the interaction between the different road users is the most complex. (Hyden and Draskoczy, 1992). As examples of these sudden unexpected events typical in rural environment, such as 1) car turning off in front (braking), 2) oncoming vehicle is on wrong lane (overtaking etc.) and 3) sudden appearance of an animal (moose etc.) on the road can be mentioned.

4.3.6.3 Motorway

The mechanism of accidents on motorway section can most often be described as monotonous activity that lowers the level of alertness and might enhance drowsiness. In such an unalert state the changes of the environment or other road users might remain unobserved, until it may be already too late to react to them in safe. Taking high speeds also into consideration, the incongruence between the high speed and the low alertness is one of the most important contributing factors in motorway accidents (Hyden and Draskoczy, 1992). Examples of these sudden unexpected events typical in motorway environment are 1) cargo falling from a truck in front, 2) vehicle braking suddenly in front and 3) sudden lane change of another vehicle .

4.3.7 Driving tasks

Driving tasks are defined separately from the scenario parameter groups. Usually the driving task is described as a task containing three different levels of demand (Michon, 1985):

1. car control,
2. interaction with other road users and
3. striving to fulfil the personal requirement on the travel demand (travel time, comfort, route etc.).

4.3.7.1 Manoeuvre

The main <u>manoeuvres</u> of driving are:

- maintaining lane,
- following (the vehicle in front)
- turning

Driving requires also many other manoeuvres such as overtaking, merging and lane changing.

4.3.7.2 Occasion/urgency of the trip

The third level of driving demand in Michon's model includes drivers' strivings for example to fulfil the travel time. It is natural, that if a driver for example has a limited time to complete the route (time-pressure), he/she is often more willing to accept higher risks in gap acceptance, overtaking etc. On the other hand, when driving in unfamiliar environment, the driver needs to direct more attention to route selection, traffic guidance etc. and might experience some difficulties in dividing the attention to other road users and IVIS as well.

4.3.8 Additional parameters

4.3.8.1 Experience of driving with a focus on novice drivers

In comparing the accident causes observed from experienced and novice drivers, it was found that novice drivers are more likely to be involved in accidents due to improper directional control of their vehicle than experienced drivers. The driving task appeared immeasurably complex —controlling acceleration and deceleration while maintaining proper position in the lane and at the same time watching out for other cars, pedestrians, signs and signals. One large part of the driving learning process is the automation of the various motor procedures such as shifting gears, braking and steering. A more subtle learning is concerned with the perceptual information gathering task — novice drivers usually look closer in front of the vehicle and more to the right of the vehicle's direction of travel than experienced drivers. (Shinar, 1978). However, the difference in car control task between novice drivers and experienced drivers is rather a short-term difference and driving skills are improved quite rapidly because of the learning process.

4.3.8.2 Impairment through medication

Road traffic safety is a significant issue for society. Europe must develop a rational transport policy to support interventions to reduce accidents. Driver impairment is a significant source of accident risk. There is increasing evidence that drugs and medicines may impair driver functioning and increase accident risk. However, these issues will not be considered within the HASTE as the most critical parameter mainly for the reason that concerted efforts in this area are being or have been undertaken elsewhere. The particularly relevant projects are described in more detail in Appendix 1 (page 102).

4.3.8.3 Other IVIS in the vehicle

The extensive growth in the wireless communication industry over the past ten years has been accompanied by concerns for the growing potential hazards of drivers using different wireless communication devices from moving vehicle. The use of radio inside vehicle was one of the first wireless communication devices used in car. Some research studies (Tijerina et al., 1995 etc.) indicated that even manual radio turning can be disruptive to driving. After radio there has been implementation of RDS-TMC devices, mobile phones and other in-vehicle devices and because of the rapidly increasing technical development, it is possible than within few years drivers have more than two wireless communication devices in their vehicle. Therefore it is important that in the future research the overall effect of all installed in-vehicle systems (combination of different devices used while driving) is studied as well. However, HASTE focuses on evaluation of one system, not combination of different systems.

4.3.8.4 Lane (number of lanes, width etc.)

The special characteristics of lane are mostly dealing with number of lanes (are lane changes possible etc.), lane separation (possibility of on-coming vehicle), but also the lane width as an indicator of complexity of driving task such as maintaining the lane.

4.3.8.5 Weather conditions

Slippery roads — Especially in many northern countries a considerable share of all accidents occur in adverse weather and road conditions, when friction between tyres and the road is low because of rain, snow or ice. Although risk estimates for different conditions are approximate in view of the limited exposure data, the risk of injury accidents in Finland has been estimated to be over nine times higher on snowy roads and 20 times higher on icy road surface than on dry and bare roads (Rämä, 2001). However, the main human errors leading to increased risk in winter are drivers' poor ability to recognise slipperiness and to adapt their behaviour to adverse weather conditions. The use of IVIS when driving in adverse road conditions might increase the accident risk because of more complex car control. However, the behavioural changes because of IVIS and their effects on accidents on slippery road conditions are estimated to be quite small.

Fog — Weather conditions such as rain, fog, or snow usually have an adverse affect on driving quality. Major reasons for this are reduced visibility (e.g. smaller viewing distance, glaring) and reduced friction (slipperiness), leading to an increase in accidents (e.g. Shankar et al., 1995). Although these adverse weather conditions are of great importance when studying driver behaviour (whether that may be with or without an IVIS), it is hardly feasible to include them in a controlled experiment context like that of the HASTE project. For a more elaborate review however see Collins (1997).

5 Driving Performance Measures

5.1 Introduction

This task aimed to establish the likely effects of IVIS on driving performance measures and determine what performance measures to be included in the pilots in WP2. In this section, driving performance measures are described from a technical, performance evaluative and a HASTE-feasibility point of view. Driving performance measures based on in-vehicle measurement and expert ratings are also included. Eventually, a selection of measures is made on the basis of the scenarios to be used and the feasibility of the measures.

5.2 Driving performance

Usually, car driving is described as a task containing three different levels of demands (Michon, 1985): car control, interaction with other road users, and striving to fulfil the personal requirement on the travel (travel time, comfort, route). Driving performance can be defined as the reflection of how well the goals on the three levels of Michon's model is fulfilled; i.e. how well the personal requirements of the trip are fulfilled, how little risk of accidents there was, and how well the driver managed to control speed and lateral position of the car. In behavioural studies, personal requirements are not relevant since the driving task almost always is artificial and have no specific purpose. Driving performance is in those cases limited to include risk of accidents. This is also the case of the HASTE project.

On the control level and the interaction level of Michon's model, the driving task can be divided into lateral and longitudinal control, which corresponds to steering, speed and headway control.

5.2.1 Lateral control

Lateral position is controlled by steering movements. The aim of lateral control is to keep the car on the road in order to minimise risk of accidents and maximise personal requirements such as comfort. If the driver's visual attention is taken away from the traffic scene the lateral control may be affected, which may result in lateral instability. The performance of controlling lateral position may be studied through variations in lateral position (e.g. SDLP) and time margin before crossing a lane limit (TLC). The most promising technical/objective measures are described below. Objective lateral position measures are most feasible for highway and rural road driving. In urban driving where there are several intersections and lanes, the lateral position measurement may not be very well defined. This of course implies difficulties for field measuring. Lateral control may be measured by subjective ratings; either by the driver or by an observer. For this, separate protocols may be used. The method of subjective ratings have not been used to the same extent as objective measures.

5.2.2 Longitudinal control

Speed is controlled so that (1) the vehicle does not collide into lead vehicles travelling in the same direction; (2) the steady flow of traffic is not hindered; (3) travel speed is within signed speed (hopefully); (4) the driver may not lose control of the vehicle when encountering “normal” events, and (5) the travel is comfortable. The importance of estimations related to these five criterions differs between individuals, and the criterions may also contradict.

The performance of controlling speed may be studied through speed variation, speed in relation to signed speed, and time margin for correcting speed before colliding into a lead vehicle. The most promising objective measures are described below. Also longitudinal control, as lateral control, can be rated by the driver and/or by an observer.

5.2.3 Laboratory, simulator and real traffic

Experimental testing of IVIS can be carried out in field experiments and driving simulator experiments (Nilsson et al., 2000). Both approaches are associated with advantages and disadvantages and it is difficult, if not impossible, to recommend a most preferable environment for “large-scale”/ “full-scale”/ “final” human factors testing of ADAS. While it is obviously clear that the technical verification of a prototype is best carried out in road traffic, it is more questionable whether “real traffic” is the only or even the ideal environment for a human factors evaluation of ADAS. In fact, it may be most reasonable to use both field tests and simulator tests complementary in the evaluation sequence since both environments can be subject to experimental methodology.

The technical possibilities of the measurement of vehicle related parameters, such as speed and lateral position of the own vehicle and other vehicles, and events in the traffic environment are much greater in simulators and mock-ups than in real traffic. Another advantage of simulators and mock-ups is that the drivers can be forced beyond the limit of their capabilities of managing critical situations without being in any physical danger.

In this report, also fixed base simulators are referred to as simulators. In the HASTE project, a PC-based driving simulator with a PC-keyboard and a mouse as input devices will also be used. This setup is referred to as a laboratory.

5.3 Objective driving performance measures

5.3.1 Accidents

5.3.1.1 Value as performance measure

Since the purpose of the driving performance measures are to reflect the risk of accident, the most obvious measure of driving performance is the occurrence of accidents.

5.3.1.2 Technical considerations

Accidents are very rare, why number of accidents is not a feasible performance measure in field trials. In simulator studies, the risk of accidents can be manipulated. However, if accidents occur too often in the simulator, the test subjects will not find the simulation as realistic and also the test subjects will not behave as normal.

5.3.2 Absolute Lateral Position

5.3.2.1 Value as performance measure

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

5.3.2.2 Technical considerations

It should be kept in mind that the variation in lane width often is of the size of centimetres and that systems for lane tracking normally are not more accurate than a few centimetres. Lateral position is usually defined as the distance between the right hand part of the front right wheel to the left part of the right hand lane marking. When the line crosses the lateral position it becomes negative

5.3.3 Standard Deviation of Lateral Position

5.3.3.1 Value as performance measure

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

5.3.3.2 Technical considerations

Measuring lateral position in real traffic requires optical equipment for lane marking detection. Also, of course, visible lane markings are required. Weather conditions, reflections and shades may affect the quality of the measurement. When there are several lane markings, it has to be assured what lane markings are tracked. In urban environment, lane markings may be present, but what the relevant lateral position is may be poorly defined. Lateral position can alternatively be measured manually on video recordings. It should be kept in mind that the variation in lane width often is of the size of centimetres and that systems for lane tracking normally are not more accurate than a few centimetres. Care should be taken that for SDLP calculations overtaking manoeuvres should be deleted from data.

5.3.4 Time to Line Crossing

Time to Line Crossing (TLC) is defined as the time to cross either lane boundary with any of the wheels of the vehicle if speed and steering wheel angle are kept constant. As the vehicle approaches the line TLC will decrease until it reaches a minimum. Under “normal” conditions this will occur when the motion of the car is changed from going towards one line to the other. During this change the car will pass a situation where it momentarily will not move toward any of the line but follow the road perfectly this will result in an indefinite or undefined TLC. The distribution of TLC is not bell-shaped. In order to determine the safety margins we have to look for the TLC minima, which is also the case for TTC. The steering performance could then be described in terms of TLC_{min} distribution.

Used as steering performance indicators the mean and median values the local minima of TLC, and also median value of TLC (not only min-values). Also the number of TLC-minima less than one second has been used. For automatic data analysis, criterions are set for finding these minima. For example, only TLC values less than 20 second is accounted for, and minima are only calculated for TTC sections that last at least one second.

5.3.4.1 Value as performance measure

Time to Line Crossing was first proposed by Godthelp and Konings (1981) to describe steering behaviour. According to Godthelp et al, TLC reflects the time available for error neglecting, assumed a fixed steering strategy. In other words; TLC reflects a lateral control safety margin. Godthelp's proposed calculation of TLC included a complex mathematical definition, based on vehicle speed, steering wheel angle, heading angle and lateral position. In this calculation, it is assumed that the road is straight. Van Winsum et al. (1996) proposed an alternative method of calculating TLC that considered road curvature. Due to problems achieving all necessary data for exact calculation, approximations are often used based on lateral position and lateral velocity and in simulator studies also lateral acceleration in relation to the road (van-Winsum and Godhelp, 1996). The different ways of calculating TLC are not described in this report.

5.3.4.2 Technical considerations

Calculating TLC in field experiment requires at least lateral position and lateral velocity data. The requirements on the measurement of lateral position and velocity are high since the TLC inaccuracy is the product of true TLC and the weighted average of lateral position and velocity inaccuracy. Also taking lateral acceleration into account is further difficult for field measuring. In a simulator, necessary variables for TLC calculations usually are available. In simulator experiments, also the road geometry can be taken into account.

5.3.5 Standard Deviation of Steering wheel angle

5.3.5.1 Value as performance measure

The standard deviation of steering wheel angle (STS) reflects the occurrence and magnitude of steering corrections, and can therefore be used as an indicator of driving performance. However, steering is more the corrections. Also voluntary steering manoeuvres are reflected in STS. Therefore, the road should be rather straight for the STS to work well as a performance indicator.

5.3.5.2 Technical considerations

Removing low frequency components can reduce the influence of curve following on steering wheel angle. However, the steering wheel movement patterns are still rather different in curves, as shown e.g. by Gabrielsen and Sherman (1994).

5.3.6 High frequency component of steering wheel angle variation

5.3.6.1 Value as performance measure

As with the standard deviation, the magnitude of the high frequency band of steering wheel angle reflects steering corrections. However, this method aims at excluding the effect of open loop behaviour and only focus on corrections. McDonald and Hoffman (1980) support that steering corrections are reflected by high frequency components.

5.3.6.2 Technical considerations

There are several ways of calculating the magnitude of frequency components. One way is to calculate the mean value of the spectral density of the steering wheel angle over a specific frequency band. For all frequency calculations, the tolerance for artifacts is low, but this should not be a problem since steering wheel angle measurement is not difficult.

5.3.7 Steering entropy

5.3.7.1 Value as performance measure

Boer (2000) suggests that steering behaviour differs between drivers and that it may be difficult to identify closed loop and open loop steering behaviour. An attempt to compensate for this is the proposal of steering behavioural entropy as a measure of steering performance. This method includes modelling of individual steering behaviour, based on steering data for individual normal driving (no extra task, like interaction with IVIS). The basic hypothesis is that secondary task demands not only affect the magnitude and/or variance of vehicle control parameters, but also leads to more disruptive and less predictable control behaviour. One way to quantify this predictability is in terms of the *steering entropy*, as described in Boer (2000). The method has been shown sensitive to workload induced by visual as well as cognitive distraction, in simulated and real world environments

5.3.7.2 Technical considerations

This indicator requires a baseline measurement of each subject's steering behaviour, which affects the design of the study including this indicator.

5.3.8 Zero crossings of steering wheel angle

5.3.8.1 Value as performance measure

Number of zero crossings reflects the frequency of steering corrections, not the magnitude. Curve following results in a constant deviation from the neutral point (zero angle). As a result, the number of zero crossings decrease during curve following; a decrease that is not related to performance. Therefore, the road should be rather straight for this indicator to work well.

5.3.8.2 Technical considerations

Removing low frequency components can reduce the influence of curve following on steering wheel angle. Care has to be taken in the calculation of this indicator so that only driver-induced zero crossings are recognised and not artifacts caused by noise

5.3.9 Changes in steering wheel rotational direction (reversal rate)

5.3.9.1 Value as performance measure

The number of changes in steering wheel rotational direction reflects the frequency of steering corrections, not the magnitude. Compared to number of zero crossing, this indicator is independent of the neutral point, and thus works just as well for curve following as for straight driving.

5.3.9.2 Technical considerations

Care has to be taken in the calculation of this indicator so that only driver-induced changes in steering wheel angle are recognised and not artifacts caused by noise. Also, the operational definition of reversal has to has to include a angular dead zone in which angular changes are not detected. These issues are discussed in detail in Elling and Sherman (1994).

5.3.10 Number of lane exceedences (LANEX)

5.3.10.1 Value as performance measure

Number of lane exceedences per time unit has been used as a measure of lateral control, e.g. by Tijerina et al. (1999). This measure is feasible for rural road, motor way and highway driving.

5.3.10.2 Technical considerations

This indicator should not cause many problems if there is a reasonably well working lane tracker available.

5.3.11 Time or distance out of lane

5.3.11.1 Value as performance measure

Driving in the oncoming traffic lane is of course related to higher accident risk. This behaviour is caused by overtakings, but also by the driver shortening the distance of travel in left curves (for right hand traffic). The fraction of distance (or time) travelled in the oncoming traffic lane and the total distance (or time) travelled has been used (e.g. Nilsson, Alm and Janssen, 1991) as a driving performance or risk indicator. This parameter is feasible for rural road and motor way driving.

5.3.11.2 Technical considerations

As with LANEX, this indicator should not cause many problems if there is a reasonably well working lane tracker available.

5.3.12 High risk overtakings

5.3.12.1 Value as performance measure

The number of hazardous manoeuvres can be used as an indicator of driving performance. Such manoeuvres are for example overtaking at oncoming traffic and overtaking at intersection. High risk overtakings most likely do not occur to a high extent. If observed to increase due to the use of IVIS, however, the manoeuvres should be considered.

5.3.12.2 Technical considerations

Overtakings, intersections and travel in the oncoming traffic lane have to be strictly defined for automatic detection. Still, this indicator is only feasible for automatic detection in simulator studies since measurement on oncoming traffic is required, which is difficult in real traffic. An accompanying "expert" could make a note if he/she considers overtakings being hazardous. (This is not included in the expert ratings of driving performance below).

5.3.13 Time headway

5.3.13.1 Value as performance measure

Time Headway [seconds] to lead vehicle is defined as the time to collide into lead vehicle if it stops dead. Time Headway is a measure of longitudinal risk margin. The closer and faster a subject travels behind a lead vehicle, the less is the chance to manage avoiding a collision in case of the lead vehicle reduces the speed. For a small headway, the time a subject may be distracted by another task without a highly increased risk of accident, is much less than if the time headway is large. The proportion of the time headway less than one second has been used as a risk indicator for car following situations. An alternative name for Time Headway is Time Gap.

5.3.13.2 Technical considerations

Time headway is calculated as the distance to lead vehicle divided by own momentary travel speed. In simulator experiments, the distance to the lead vehicle should be defined as the distance between the bumpers of the cars. In field experiments, the distance is measured with laser equipment or radar, which cannot be controlled to lock on a bumper; rather the back of the car is focused on, which should be sufficient. A percentage inaccuracy of the measured distance results in a corresponding inaccuracy of the calculated time headway. The system for measurement of distance to lead vehicle is here suggested to be able to detect vehicles up to 70 metres away. This requirement results in the possibility to detect time headway less than two seconds at the travel speed of 110 km/h. A difficulty of time headway measurement in field studies is to know whether the radar locks on correct vehicles or not. In simulator studies the function of the radar is often idealised, which results in validity problems in relation to real radars.

5.3.14 Distance headway

5.3.14.1 Value as performance measure

Distance Headway [metres] to a lead vehicle is defined as the distance to lead vehicle, preferably defined as distance from bumper to bumper. The only difference to time headway is that vehicle speed is not considered. An alternative name for Distance Headway is Distance Gap or just Gap.

5.3.14.2 Technical considerations

See above Time Headway.

5.3.15 Time To Collision

5.3.15.1 Value as performance measure

Time to Collision (TTC) [seconds] is defined as the time to collide into lead vehicle if vehicle speeds are kept constant. TTC reflects risk margin; the less TTC, the less margin. TTC frequently reaches min-values, and in between have sections of not defined values (the lead vehicle at higher speed than following vehicle). Mean value and median value of the TTC-minima and the number of TTC-minima less than one second have been used as indicators of risk of collision.

5.3.15.2 Technical considerations

For automatic data analysis, criteria are set for finding the TTC minima. For example, only TTC values less than 5 seconds are considered, and minima are only calculated for TTC sections that last at least one second. TTC is calculated as the distance to the lead vehicle, measured from bumper to bumper, divided by the difference in momentary speed between the own vehicle and the lead vehicle. TTC is of course defined only for values larger than zero. The lead vehicle speed and distance to the lead vehicle are required. Since inaccuracy in distance- and speed data affects both nominator and denominator in the calculation of TTC, care has to be taken so that inaccuracy is kept at a minimum. As with time headway, the most crucial problem is to know what objects the radar locks on.

5.3.16 Absolute speed

5.3.16.1 Value as performance measure

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

5.3.16.2 Technical considerations

Speed measurement should cause no problems for speeds larger than about 5 km/h. It should however be considered that there may be a discrepancy between true speed and speed displayed to the driver. Since wheel diameter influences speed measurement, speed data may need correction to be accurate.

5.3.17 Standard deviation of speed

5.3.17.1 Value as performance measure

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

5.3.17.2 Technical considerations

See below Absolute speed.

5.3.18 Abrupt onsets of brakes

5.3.18.1 Value as performance measure

An abrupt and intense beginning of the deceleration caused by braking indicates the occurrence of a critical situation (Nygård, 1999). This measure can be used as a measure of driving performance; high rate of jerks indicates higher risk and less performance. However, it should be taken into account that another road user may cause the critical situation. In this case, this measure may indicate a correct action. This measure is feasible in high traffic density environments where jerks are most likely to occur.

5.3.18.2 Technical considerations

Since it should be clear what is the cause of the critical situation for this measure to be feasible as a performance measure, the measure is favourably not included in real traffic, but rather in simulator studies.

5.3.19 Reaction time

5.3.19.1 Value as performance measure

Driver reaction time (RT) to a change in traffic light colour and to the appearance of unexpected events, such as obstacles and sudden firm braking of a lead vehicle, is a straightforward measure of speed control performance. Reaction time is preferably measured from the onset of the event to the driver reaction onset as release of accelerator or onset of evasive manoeuvre. It should be considered that the driver might not have the foot on the accelerator if an ACC is used.

5.3.19.2 Technical considerations

Data on the unexpected events and the use of primary controls are required for RT-calculation. Automatic measurement of unexpected events in field trials may be very difficult. Measurement of the driver reactions to unexpected events is common in simulator studies.

5.3.20 Post Encroachment Time (PET)

5.3.20.1 Value as performance measure

Crossing lanes or roads where the driver has to give way is risky if the time gaps between vehicles (in the lane to be crossed) are small. Imagine an intersection where one vehicle drives across the intersection and the approaching vehicle passes behind it without having to alter its trajectory. The risk of potential collision can be measured as the time separation between when successive vehicles occupied the same space. This time is the PET.

5.3.20.2 Technical considerations

It is difficult to measure time gaps between vehicles in field trials. It could be done manually for crossings in intersections if the situations are recorded on video. This does however not result in very accurate data. In simulator studies, this is not a problem.

5.4 Use of other controls

5.4.1 Use of the direction indicator

5.4.1.1 Value as performance measure

The use of the direction indicator in situations where the direction indicator should be used can be used as a performance measure. Such situations are when performing turns and overtaking other vehicles. For this measure to be feasible, the study scenarios should include overtakings and turns.

5.4.1.2 Technical considerations

Measurement of the use of the direction indicator and information on distance to intersection/turn are needed for automated detection in real traffic. Video studying or accompanying observer may of course note such behaviour.

5.4.2 Switching between full and dipped beam (headlights)

5.4.2.1 Value as performance measure

Switching from full to dipped beam at the occurrence of oncoming vehicles can be used as a performance measure. This measure is feasible for non-urban driving and low-medium traffic density.

5.4.2.2 Technical considerations

Automatic detection of oncoming traffic in real traffic is difficult, but could be noted by an accompanying observer or by studying video recordings.

5.4.3 Gear selected

Driving in correct gear may say something about vehicle control. You could assess it by measuring RPM. Disadvantage is that if a car has automatic gear it –of course- doesn't say that much.

5.5 Expert ratings of driving performance

Expert ratings of driving performance have not been used to the same extent as objective/technical measures by far. This is in spite of that humans most likely are more sensitive to changes in risk margin than computers. On the other hand, experts tend to be humans with their personal prejudice, they might be biased by the looks of the driver, etc. The difficulty lies in producing rating methods that are reliable. However, there are some attempts to use expert ratings for evaluating fitness to drive, which can be equalled to driving performance. The rating protocols described below may not be fully suitable for ratings of driving performance. They do, however, contain parts that seem highly relevant.

5.5.1 Wiener Fahrprobe and Lund Observer Protocol

The Wiener Fahrprobe by Risser (1985; see Appendix 1.2.C) is a method for rating driving performance of novice drivers. The method requires two accompanying persons who are trained on the ratings; one who rates standardised driving performance variables such as yielding behaviour and speed choice, and one who performs non-standardised observations of e.g. hazardous situations and occurrence of conflicts with other road users. If only one observer can observe, the non-standardised part is preferably excluded. It can be concluded that the protocol focuses on a tactical level (Michon's driver model). Since this method requires at least one observer in the car, this method is hard to apply correctly in a simulator or laboratory (low cost simulator). It is reasonable that if observations of video recordings are performed rather than inside the vehicle, the sensitivity and reliability of the method is reduced. The Lund Observer Protocol is a reduced version of the Wiener Fahrprobe.

5.5.2 Test Ride for Investigating Practical Fitness to Drive (TRIP)

The TRIP protocol (see also Appendix 1.2.B) was developed in Holland, and has mainly been used by to assess driving skill of drivers with severe functional limitations (see Appendix B). The assessment has been performed by driving licence examiners, re-educated for the job of assessing driving skill (or skill potential) including tactical compensatory behaviour. The protocol contains several parts that aim on driving performance on the operational and the tactical levels of Michon's driver model.

5.6 Ratings of own driving performance

5.6.1 Driver Behaviour Questionnaire

Reason (1990) investigated unsafe driving behaviour using a questionnaire known as the Driver Behaviour Questionnaire (DBQ). The DBQ was submitted to 520 British drivers covering a broad age range. In a subsequent study, Parker et al (1995) administered a shorter version of the DBQ to more than 2.000 drivers. In both studies, the data was submitted to principal components factor analysis and three unique factors were extrapolated, together explaining roughly 1/3 of the variance. One of the factors, mistakes, was related to misjudgements of various kinds. The second factor, violations, comprised items that related to violations and breaching of formal and informal traffic codes. The third factor, slips and lapses, was related to erroneous behaviours, characterised as relatively harmless.

The three-factor structure has been possible to recover in further studies on independent samples of drivers. However, a similar study on Swedish drivers by Åberg & Rimmö (1998) found somewhat different results. In addition to the original 44 British items 60 new ones were included. Four distinct factors were excluded, together explaining 44% of the variance. Both the original violations and mistakes factors were identified, but the slips and lapses factor was further subdivided into two distinct factors, an inattention factor and an inexperience factor.

In several studies, the violation factor has been found to predict road traffic accidents retrospectively and prospectively. Some findings indicate that the mistake factor may also be important as an accident predictor. The slips and lapses factor, however, appears to be unrelated to accident involvement.

ADA systems could be very beneficial for many of the different errors that are committed in traffic, e.g. navigation, speed adaptation or collision warning.

The DBQ is not well suited for being used in experimental studies. Behaviours that could be classified as aberrations cannot be expected to appear with sufficient frequency to be of great use in experimental studies. It may also be difficult to classify a particular behaviour as being a driver aberration or not. Also, since they partly include covert behavioural phenomena, they are difficult to classify. Other methods may be necessary.

5.6.2 Driving quality scale

Brookhuis, Uneken and Nilsson (1999) presented a scale for self-rating of driving performance, the Driving quality scale, which is well suited to be used in experimental studies on driving behaviour (Figure 21). Drivers are asked: "How well did you drive during the trial, compared to normal?"

| ⇐ I drove extremely well

| ⇐ I drove as usual (normal)

| ⇐ I drove extremely badly

Figure 21: Driving Quality Scale

5.7 Selection of driving performance measures

Here, a set of mandatory driving performance measures is suggested. Also suggestions for optional measures are listed. Some measures will not be included in the final experimental designs, since the feasibility of some measures depends on the traffic conditions. There may also be some changes in the sets of measures if factors on feasibility and knowledge on sensitivity arises within WP2. In Table 4 and Table 5, the mandatory and optional measures are listed. Also suggestions on required accuracy are included in the tables (in parentheses). These numbers are qualified guesses, and should not be held as absolute limits. They are included to give some guidance on required accuracy and to emphasize the importance of accurate measurements.

Table 4: Mandatory driving performance measures

Measure	Field	Simul.	Lab.	Requirements for field trial
Objective measures				
Reversal rate (-)	OK	OK		Steering wheel angle (e<0.5^o)
LANEX (-)	OK (not urban)	OK	OK	Lateral Position (e<10cm)
LP (e<5cm)	OK (not urban)	OK	OK	Lateral Position (e<5cm)
SDLP (e<5cm)	OK (not urban)	OK	OK	Lateral Position (e<5cm)
TLC (e~<0.5s) (mean of mins)	OK (not urban)	OK	OK	Lateral Position (e<5cm)
Speed (e<2km/h)	OK	OK	OK	Speed (e<1km/h)
STD Speed (e<2km/h)	OK	OK	OK	Speed (e<1km/h)
Time headw. (e<0.02s) (mean of mins)	OK	OK	OK	Distance to vehicle (e<0.5m) Speed (e<1km/h)
Dist. headway (e<0.5m) (mean of mins)	OK	OK	OK	Distance to vehicle (e<0.5m)
TTC (mean of mins) (e<0.25s)	OK	OK	OK	Distance to vehicle (e<10cm) Speed diff (<5km/h)
Reaction time to unexpected events (e<5ms)		OK	OK	
Subjective ratings				
Lund observer protocol (one observer)	OK	OK	OK	

Table 5: Optional driving performance measure

Measure	Field	Simul.	Lab.	Requirements for field trial
Objective measures				
High freq comp of steering wheel angle (e<?)	OK	OK		Steering wheel angle (e<0.5^o)
Steering entropy (e<?)	OK	OK		Steering wheel angle (e<0.5^o)
Abrupt onsets of brakes (-)	OK	OK		Long. acc (e<$0.1m/s^2$)
Time spent out of lane (e<0.1s)	OK	OK	OK	Lateral Position (e<10cm)
PET (e<10ms)		OK	OK	

5.8 Basic parameters

Below, the parameters required for calculation of the above listed driving performance measures are described.

Table 6: Required driving parameters

Measure	Laboratory	Simulator	Field
Travel time	OK	OK	OK
Distance travelled	OK	OK	OK, separate or original pulse generator. GPS has proven to produce accurate two dimensional position data, which can be used for distance calculation.
Speed	OK	OK	Original speed encoders generate about one pulse per km/h. Speed measurement should therefore be OK over 5 km/h. Separate encoders increase accuracy. GPS using the Doppler effect seems promising.
Longitudinal acceleration	OK	OK	OK, separate accelerometer, or calculated based on accurate speed data. CAN-bus data may be used, if data is available.
Lateral position	OK	OK	Difficult; requires optical equipment. Lines may be poor, and it is difficult to know if the correct lines are accounted for. GPS does not seem to have sufficient accuracy for lateral position measurement at the moment.
Lateral speed	OK	OK	Difficult; accurate lateral position data required.
Absolute lateral acceleration	OK	OK	OK, separate accelerometer. CAN-bus data may be used, if data is available.
Road relative lateral acceleration	OK	OK	Difficult, accurate lateral speed data is required.
Steering wheel angle	No	OK	OK, separate sensor. CAN-bus data may be used, if data is available.
Use of brake	OK	OK	OK, original sensor. CAN-bus data may be used, if data is available.
Brake force	No	OK	OK, but separate force-/pressure sensor in brake system is required. CAN-bus data may be used, if data is available.
Use of accelerator	OK	OK	OK, but separate sensor is required. CAN-bus data may be used, if data is available.
Use of secondary controls	No	OK	OK, if CAN-bus data can be used or relays can be connected to.
RPM	-	OK	OK, original or separate RPM-meter. CAN-bus data may be used, if data is available.
Distance to lead vehicle	OK	OK, but it has to be clear what vehicles to be accounted for.	Difficult; requires radar or equivalent equipment. Difficult to know if the radar actually locks on correct vehicles.
Lead vehicle speed/ relative speed	OK	See “Distance to lead vehicle”	Difficult; radar and laser Doppler can detect speed. If this is not the case, accurate distance to lead vehicle is required.

5.9 Calculations

Some measures require signal filtering, mathematical modelling (e.g. steering entropy) and non-linear mathematical operations (like removal of artefacts or identification of local minima). These operations are preferably executed in Matlab or any corresponding program. Most mathematical operations that are required are very simple to implement and should not reduce the feasibility of the measures.

6 Workload Measures

6.1 Introduction

Workload is a complicated phenomenon. It can be put into different categories, such as visual workload (how much sources does the driver have to look at?), motor workload (how much does the driver have to do with hands and feet?) and mental workload (how much information does the driver has to process?). Although there are several methods for measuring workload, they are usually aggregated over time. This makes traditional subjective measures of workload sometimes difficult to interpret, because it is unclear whether the driver refers to overall workload or peaks in workload (De Waard, 1996).

In this section several ways of measuring driver workload are discussed and after that prioritised with respect to the feasibility for use in a HASTE testbed.

6.2 Workload Measures

6.2.1 Primary task performance

When workload is predictable, the driver will generally try to control workload by making the primary driving task easier to perform. For example driving in a more complex environment usually leads to a higher workload, which results in the choice of a lower speed. This compensation of changing ones behaviour in order to have better control over the driving task is also called risk compensation. The idea of risk compensation is that every person has a preferred level of risk. If risk on one situation increases (for example driving through a complex environment) you compensate this with increasing safety on another situation, for example you decrease your driving speed. In sum, increased task complexity can be adapted for by either increasing the amount of invested effort or by adjusting one's strategies.

Complexity-related mental effort has also been described as 'computational effort' (Kahneman, 1973; Mulder, 1980). This type of effort has to do with the cognitive and knowledge demands of the (driving) task. Therefore it has been reasoned that computational effort is related to working memory and attentional processes. (e.g. Baddely & Hitch, 1974).

Next to being related to (driving) task complexity, the amount of effort a certain task takes is determined by 'how hard a person has to try'. There are both internal and external circumstances which affect the efficiency with which a task can be executed. For example, if a person is fatigued, under the influence of medicines (internal factors), or performing in a noisy environment (external factor) this person has to compensate for these adverse working conditions. A person working in such conditions may very well be able to execute a task effectively, but it will most probably be done at a much higher cost (i.e. less efficiently) Therefore, this type of effort has been called 'compensatory effort' (Gaillard, 1992, Hockey, 1979, 1993). Related to this concept is the level of task commitment or involvement one has. The (un)willingness to invest extra effort in a task is reflected in the primary task performance, especially when subjective task difficulty increases.

Therefore, if we measure the primary task performance, in this case several parameters of driving performance, we can identify workload. When studies are done in laboratories, primary task performance is often measured in terms of response times or the number of errors made. When we go outside the laboratory, it really depends on what the task is. For instance, when we take driving a vehicle as a primary task, sensitive measures may be speed (if something becomes more demanding people have the tendency to reduce their speed), the standard deviation of the lateral position (swerving inside the lane) and standard deviation of the steering wheel movements. An example of a speed measurement is the time required to finish a specific route. Jordan & Johnson (1993) as well as Fairclough et al. (1991) found that people needed more time to finish a specific route is they had to interact with the radio or had a conversation compared to subjects who did not have to do these extra tasks. The application of time-to finish-a-route as a measure for mental workload is rather rough. Direct speed measurements are more sensitive, as long as circumstantial conditions, such as traffic density, is compensated for.

Standard deviation of lateral position is a sensitive performance measure (e.g. Hicks & Wierwille, 1979; Green et al., 1993; O'Hanlon, 1984). De Waard (1996) only found a increase in the standard deviation of lateral position (more swerving) as an effect of alcohol and prolonged driving, but not as a result of more mentally loading tasks such as handling a car phone. Green et al., (1993) also warn for the fact that the standard deviation of lateral position is also depending on lane width, since drivers have the tendency to swerve more inside their lane if the lane is wider.

The standard deviation of the steering wheel movements are related to the standard deviation of lateral position, but it is closer to one of the main causes of swerving that is steering behaviour. This measure is most accurate for straight road driving and is based on the idea that if a task gets more loading, less attention will be paid to keeping the most accurate position in the lane.

The problem with primary task performance is that a decrease in primary task performance is normally used as an indication of an increase in workload. However, this decrease may also be a result of time-on-task (if one gets tired, performance may decrease even though the workload of the task does not change). Also, an increase in primary task performance does not always have to indicate a decrease in workload. Here, it may also be that practice improved task performance. Another problem with using primary task performance as an indication of workload is that is the task is very easy, and people have a lot of capacity left, the measure is not sensitive to small increases or decreases of workload. Therefore, is the task is not very loading, secondary tasks are a very useful tool to assess workload

6.2.2 Secondary task performance

Secondary tasks can be used as indicators of the task load of a primary task. The mechanism of compensating for a loading task by diverting from other activities is clearly visible in driving: when a driver is involved in a heavy discussion, but suddenly the preceding vehicle starts braking, the driver will stop talking. This mechanism is exactly what is measured with secondary tasks. By giving drivers an extra task that is not directly related to the driving task (e.g. counting words or doing subtractions) and measuring the performance on this task, an indication for workload may be obtained. When the workload of the primary task (driving) is low, performance on the secondary task will be better. When performance on the secondary task drops, this means that more workload is experienced in the other task involved (driving in this case).

Secondary tasks applied in this way have a few drawbacks that reduce their usability. The most important disadvantage is the intrusion on primary task performance. Since secondary tasks generally compete for attentional demands or resources with primary task execution, they may result in poorer driving performance. Also when the primary task becomes increasingly complex, people often are not able anymore to execute a secondary task at all, which causes this method to loose its sensitivity under these conditions. Another point is that secondary task performance may be affected by strategic resource allocation in which the driver may choose to allocate more attention to the secondary task instead of the primary task or vice versa. These considerations suggest that, in order to be useful, the secondary task should not compete for resources with the primary task. The Peripheral Detection Task (PDT) method exemplifies this. Since the attentional demands of peripheral detection are low, performance on the PDT is regarded as a useful indicator of workload.

6.2.2.1 Peripheral Detection Task (PDT)

Peripheral detection as a method for estimating workload has become more popular during the last years in driver behaviour research. It is based on the idea that the functional field of view is reduced with increased workload or, alternatively, that attention becomes more selective with increased workload (Miura, 1986). It has been implemented in several ways, but one method consists of presenting a light stimulus for one second at a horizontal angle between 11° and 23° with an inter-stimulus interval of 3 to 5 seconds. The stimulus can be perceived in the peripheral field of view and does not require foveal vision. The driver responds to the stimulus by pressing a response button attached to the index finger. The percentage of missed signals and average reaction time increase with higher workload. This method is useful for measuring workload over a longer period of time (as in the case with the subjective measures) as well as for measuring variations and short lasting peaks in workload.
In a number of studies this method has shown sensitivity to small variations in workload. Some examples are workload as a function of traffic and road environment, driving experience or IVIS complexity (e.g. Van Winsum & Hoedemaeker, 2000, and Van Winsum et al., 1999).

Furthermore, PDT has a functional correspondence with roadside objects. The horizontal angle at which the stimuli are presented to the driver corresponds with the location of pedestrians or road signs. If more PDT stimuli are missed because of increased workload, it may be assumed that under similar circumstances also more road signs, pedestrians or other relevant objects may be missed because of attentional narrowing. Because of this, the measure appears to be valid. Similar findings have been reported in different studies under similar circumstances. Thus, the method appears to be reliable.

6.3 Visual performance workload measures

The primary source of information available to the driver is visual. Information gathered by looking at objects and events enables the driver to perform decision-making tasks, as well as to control and navigate the vehicle in the road traffic environment. Thus, high visual attention requirements of in-vehicle displays are considered to be detrimental to driver safety because it is at the expense of attending to the roadside view. The problems associated with visual attention are restricted to visual displays and displays that require visual-manual interactions (such as menus). Other types of IVIS that require, for instance, voice input or output may be associated with other categories of workload. Glance frequency and glance duration are often measured by off-line video analysis. These are time-consuming. Alternatively, eye-tracking equipment can be used for on-line or off-line analysis.

6.3.1 Visual occlusion

A classical measure of assessing the amount of visual information that a driver needs is the method of visual occlusion. By measuring how much information drivers get from their visual environment, an indication for visual workload is found. With visual occlusion, the driver's field of view is visually obscured by spectacles with LCD glasses. Every time the driver feels it is necessary to visually scan the road, he can open the glasses for a short preset period of time (e.g. 0.5 s). The method of visual occlusion is based on the idea that if the driver needs to get more information from the visual environment (whether on the road or on an in-vehicle system), the fiel of view will be openend more often. The visual occlusion technique is often used by TNO Human Factors (Godthelp, 1981, De Vos et al., 1996; Hoedemaeker & Kopf, in press) by using spectacles with LCD glasses, called the PLATO device (Portable Liquid-Crystal Apparatus for Tachistoscopic Occlusion). An index for visual workload is then derived from the frequency at which drivers open the PLATO glasses while they try to perform the driving task.

6.3.2 Basic ocular measures

Various basic ocular measures are indicators of visual demand and workload. Previous research has shown that high attentional workload produces attentional focus narrowing, which is indicated by spatial variability reduction of gaze direction (Underwood and Radach, 1998; Recarte and Nunes, 2000). Ocular measures have proven to be sensitive to visual functional-field size reduction observed in cognitive workload tasks (Recarte and Nunes, 2000). Examples of measures used are, fixation duration or rate, standard deviation of gaze direction, saccade size and velocity, root mean square road center/velocity (of both head and gaze direction), power spectrum measures, large visual angle reversal rates, saccade and fixation rates, smooth pursuit rates, scan paths (e.g. mirror checking), staring (i.e. fixated gaze without blinks), etc. Head movement (especially rotation) may be a powerful new indicator of workload.

Table 7: Basic ocular measures

Basic ocular measures
Fixation-based measures
Duration
Frequency/rate
Fixation point variation (std)
Smooth Pursuit-based measures
Duration
Frequency/rate
Size/dispersion
Saccade-based measures
Duration
Frequency/rate
Size
Peak velocity
Peak acceleration
Pupil size

Staring (i.e. fixated gaze without blinks)
Root Mean Square Road Center
Mean Square Control Velocity
[large, medium, small] visual angle rates
Human rater/transcription based measures (this is possible to automate in driving simulators)
glances to objects in scene camera (e.g. cars, signs, pedestrians)

Table 8: Head movement based measures (XYZ head rotation and position)

Head movement based measures (XYZ head rotation and position)
Variance
Mean
Root Mean Square Road Center
Mean Square Control Velocity
Power spectrum measures (Fourier analysis)
[large, medium, small] visual angle rates

6.3.3 Glance-based measures

Visual demand imposed by secondary tasks is traditionally quantified in terms of *glance-based measures* (Figure 22), according to the standard offline video analysis method described in ISO 15007-2 and SAE J-2396. As described in Victor *et* al., (in press), unintrusive eye-/head tracking systems can be used to automate this method, which may easily be adapted for online use. Real-time calculation of glance behaviour, traditionally done by frame-by-frame video transcription, is now possible and being further developed at VTD, thus enabling established safety thresholds to be monitored in real-time (e.g. Zwahlen diagram, EU recommended 4 glances or 2 second single glance duration, AAM recommended 20 second total glance duration, AAA recommended 10 second total glance duration, 15 second rule (Green, 1999) etc.).

Recent work at VTD has shown that glance-based measures are reliably computed automatically in the analysis software developed at VTD using faceLAB data output. To this end, signal-processing algorithms for noise reduction, data quality management, signal segmentation first into fixations, saccades and smooth pursuits and then into glances, transitions and tasks, have been developed. Cluster identification and intelligent vehicle/world model target association using a hierarchical classification scheme, i.e. in/off-road then larger to smaller in-vehicle targets (e.g. dash - centre console – radio - dials), is employed.

Glance-based measures according to the ISO/SAE specifications include:

- Glance frequency
- Average glance duration
- Total glance duration
- Total task time
- Time off road scene ahead
- Transition times
- Percentage of time spent on different targets
- Fixation probabilities

- Link value probabilities
- Scan paths (e.g. mirror checking)

Table 9: Hierarchical classification of targets

On-road	***Off-road***
Center	Instrument cluster
Left	Speedometer
Right	Tachometer
Near	Center Console
Far	Radio
Sides	Climate
Mirrors	Telephone
	Palm pilot, etc

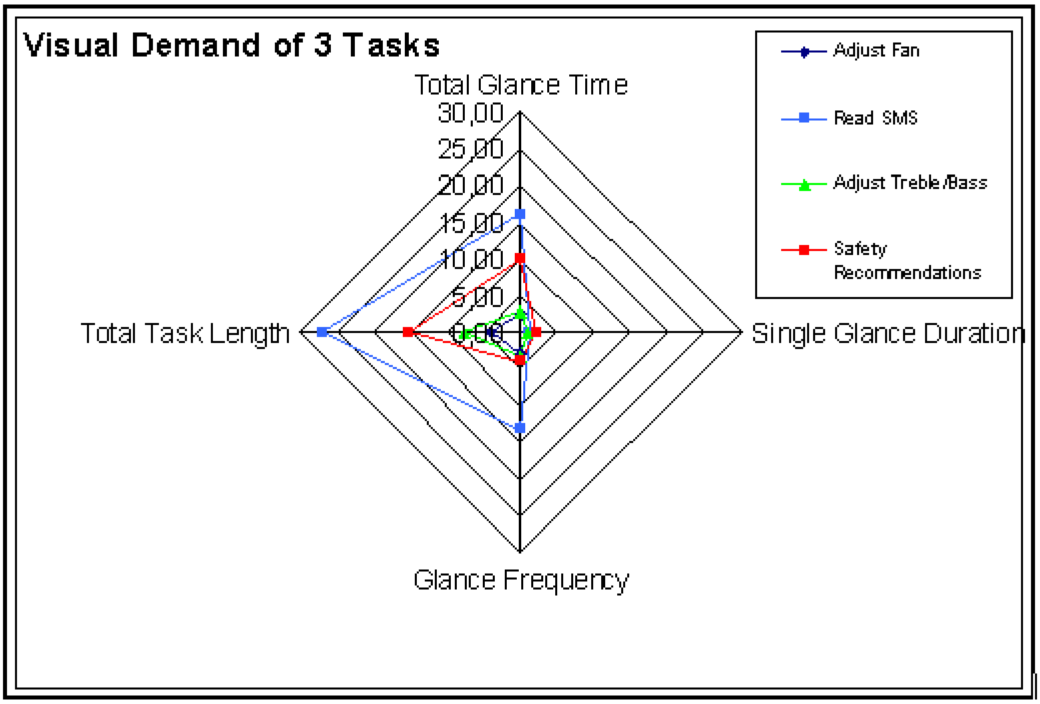


Figure 22: Three tasks are evaluated according to the three safety recommendations

6.3.4 Glance frequency

Glance frequency is often described in terms of the mean number of glances (MNG) to an IVIS. A glance is defined as a series of eye fixations on the same target area. MNG varies strongly between different in-car tasks (Rockwell, 1988; Wierwille et al., 1988; Kurokawa & Wierwille, 1990; Taoka, 1990). Depending on the complexity of the task, typically between 1 and 7 glances are needed to acquire and process the information. Because it is related to the overall complexity of the display, it is a highly sensitive measure of visual attention or visual workload (Verwey et al., 1996).

For example, drivers generally apply the strategy to sample the in-car display for preferably one second at maximum, then return to the forward view, then sample the in-car display again, etc (Wierwille, 1993a, 1993b). When visual load increases (for example, because of information complexity or poor legibility) more glances are made. The increase in glance frequency with higher visual load is then a coping mechanism of the driver intended to prevent the negative effects on safety. Thus, the validity of the measure for driver safety is limited although the increase in glance frequency should be taken as a sign that the driver is actively counteracting negative effects of driver safety in response to increased visual load. Because the range of the number of glances and the number of glances per task are consistent between different studies, this measure can be considered as highly reliable.

6.3.5 Glance duration

Glance duration is defined as the time that the eyes are taken off the roadway to attend to a target (Rockwell, 1988). It is often expressed as average glance duration (AGD). The application of glance duration in measuring workload has both its usefulness and limitations under different conditions.

A number of studies have indicated that AGD is fairly constant across a number of in-car tasks with a definite upper limit that the driver does not like to exceed. For instance, Wierwille et al. (1988) found that AGD varied between 0.60 and 1.70 s for different in-car tasks. Because of this, the sensitivity of this measure for variations in visual load is limited (i.e. range of variability is limited).

Lateral deviation of the vehicle from normal lane position increases with the time spent looking away from the external view (Zwahlen & Balasubramanian, 1974; Zwahlen & DeBald, 1986; Nieminen & Summala, 1994; Pohlmann & Traenkle, 1994). Since this increases the chance of driving off the road, longer glance durations to in-vehicle IVIS are considered to be dangerous. Thus, the validity of glance duration in this aspect is high.

Glance duration may increase as a function of visual load. However, it may also be higher if there is sufficient time to process the in-car information. This suggests that glance duration as such has limited reliability.

Given the relatively low sensitivity of glance duration per se, it needs to be combined with glance frequency (Fairclough et al., 1993). Guidelines to prevent driver overload by visual information should preferably be expressed in terms of total glance time (the summation of successive glance times to process a single piece of information). It is generally the case that if the number of glances increases because of visual load, the single glance duration increases as well, although much less so.

Finally measurement artefacts and other problems are quite common in measuring glance behaviour. Glasses often inhibit the eyes from being rightly or accurately detected by the measurement device. Postural movements are a second source of technical artefacts: they cause the eyes to be out-of-focus, thereby inhibiting measurement. Both of these artefacts limit the practicability of devices that measure glance behaviour to a certain degree.

6.4 Subjective workload

Subjective measures of workload are proven to be easy to administer, relatively non-obtrusive and sensitive to variations in workload. However, a possible drawback of these methods is that they depend on subjective judgments of subjects that may be affected by their expectations and biases, especially when they are not properly instructed on how to answer the questions.

6.4.1 NASA-TLX

The NASA-TLX (Hart & Staveland, 1988) is a multidimensional scale that measures the subjective evaluations of six factors of workload: mental demand, physical demand, temporal demand, performance, effort and frustration level. Subjects are asked to rate the load on a scale from 0-100 and to rate the relative importance of the different factors. It is one of the most frequently used subjective measures of workload. NASA-TLX is useful for measuring workload over a longer period of time, but is not suited for detecting peaks or short lasting increases in workload.

NASA-TLX has also been proven to be sensitive to differences in workload in a number of studies of car driving. Studies have shown that it is more sensitive to workload than other multidimensional subjective workload scales, such as the MCH and SWAT discussed below (Hill et al., 1992). However, there is no evidence that it is more sensitive than unidimensional scales such as the RSME, which are easier to administer.

It is assumed that a higher mental workload is detrimental to driver safety, although the experimental evidence on the relationship between NASA-TLX performance and driver behaviour and critical incidents is lacking. The validity of this measure still needs to be established in research. Nonetheless, since NASA-TLX is consistently sensitive to increases in workload in different studies it is regarded as a reliable method of measuring workload.

6.4.2 Subjective Workload Assessment Technique (SWAT)

The SWAT is a multidimensional workload assessment scale that measures load on three dimensions: time stress, mental effort and psychological stress (Reid & Nygren, 1988). In its original form, SWAT consists of a card-sorting task wherein 27 rating scale combinations have to be rank-ordered. Based on this, a single scale of workload is constructed. This is followed by a procedure for event scoring. A major drawback of this technique is the amount of time required for the card sort. Because of this, attempts have been made to simplify the procedure and experimental studies have shown that the simplified version gives results that are comparable with the complex procedure.

Comparative studies have revealed that although SWAT is a reliable measure, it is not as reliable as the NASA-TLX. Furthermore, although SWAT has likewise been shown to be a sensitive instrument, it is not as sensitive as NASA-TLX, but is more sensitive than the MCH scale. And as with other measures of subjective workload assessment, the validity of SWAT still needs to be fully established in research.

6.4.3 Modified Cooper Harper Scale (MCH)

The MCH is a unidimensional scale consisting of ten items that add to a single score (Cooper & Harper, 1969). The MCH is not suited for measuring short-lasting variations in workload while driving. And although the MCH has been shown to be sensitive to variations in task difficulty, it is not as sensitive as the multidimensional NASA-TLX or the unidimensional RSME. Its validity still needs to be fully determined in experimental research. Comparative studies of different subjective workload measures have even shown that the reliability of

MCH is lower compared to the other more popular methods of subjective workload estimation.

6.4.4 Rating Scale Mental Effort (RSME)

The RSME (Zijlstra & Van Doorn, 1985) is another unidimensional scale in which ratings of *invested effort* are indicated on a vertical line (from 0-150). Along the line are a number of anchor points that are labelled with a verbal descriptor of effort ('almost no effort', 'extreme effort', etc.). It is easy to administer both after and during driving.
In comparison with other techniques for measuring workload, RSME is one of the most sensitive measures of workload (Verwey & Veltman, 1995). A higher degree of invested effort is considered as an attempt of the driver to keep performance on a certain level in response to increased task demands. The technique has a high reliability since it consistently results in higher workload ratings as a function of task load. Conceptually speaking though, increased mental effort may not necessarily be related to decreased driver safety. However, it can be assumed that at a certain level of task demands, performance of the primary task may deteriorate if a certain amount of effort is exceeded. Because of this, high levels of invested effort are considered detrimental for driver safety. This suggests that a high validity is assumed with RSME. Then again, this needs to be demonstrated in experimental research.

6.5 Physiological workload measures

Physiological measures are mainly suitable for measurements over longer periods of time. Over the past few decades, characteristics of the rhythm of the cardiovascular system has been widely used in order to measure psychophysiological state and mental effort, both in the laboratory (Mulder, 1987, 1988) and in field studies (Mulder, 1992; Wiethoff, 1997). The autonomous nervous system (ANS) controls the internal organs. It is autonomous in the sense that the innervated muscles are not under voluntary control. The ANS is composed of two neural branches, the parasympapthetic (PNS) and the sympathetic nervous system (SNS). For the simplicity (but see: Berntson et al., 1991) these both branches can be perceived as antogonists, with the PNS having an inhibitory or energy-conserving effect and the SNS having an excitatory or energy-releasing effect. In a cardiovascular context, the PNS is often referred to as the 'vagal system'. The heart is influenced by the autonomous nervous system, and through this connection the heart is related to physical and emotional states as well as to cognitive activities. Basically the rationale for using heart rate to assess mental workload is the following. The body meets the increased oxygen demands that accompany increased activity by increasing the heart rate. Therefore, in the absence of physical effort an increase in heart rate (oxygen demand) can be attributed to an increase in mental effort.

Heart rate and heart rate variability are discerned as different measures that can be used as an index of workload in the driving context. In general, heart rate provides an index of overall workload or activation, whereas heart rate variability is more useful as an index of cognitive or mental workload (Wilson & Eggemeier, 1991). Profiles of heart rate and heart rate variability may be helpful in linking changes in the pattern to specific loading events in the driving environment.

6.5.1 Heart Rate

Heart rate is known to be related to the amount of physical activity, respiration, thermal regulation, and muscle preparation for movement (Lysaght et al., 1989). Kalsbeek and

Ettema (1963) found that fluctuations in the interbeat interval (IBI, the time between two peaks in the electro-cardiogram) were reduced during more demanding mental task performance. De Waard (1996) found that driving significantly increased the average heart rate compared to measurements in rest. But also in different driving conditions, he found this measure to be useful. He found that driving through a weaving section and using the car phone resulted in an increase of average heart rate in comparison to a baseline measurement during driving. Time-on-task (related to sleepiness) or decreased vigilance on the other hand resulted in a decrease of average heart rate. Fairclough et al. (1991) found that the average heart rate was higher while performing a secondary task presented through a hands-free phone, compared to the same task presented by an experimenter in the passenger seat. The authors give two possible explanations for the effect. Either additional effort is required in the phone condition due to lack of conversation, or unfamiliarity with cellular phones activated the subjects.

With the use of heart rate profiles, it is possible to monitor heart rate at a more continuous level. De Waard (1996) used a window of 30 s, moving it over the heart rate pattern with steps of 10 s, resulting in a smoother heart rate profile. He found differences in heart rate linked to specific road segments where subjects were driving. Driving over dual carriageways clearly reduced heart rate, and driving around roundabouts increased heart rate. De Waard et al. (1999) found that with the profile technique, a clear decrease in heart rate was found when drivers experienced an emergency situation on the automated highway. This decrease is explained as a surprise reaction to a novel situation.

6.5.2 Heart rate variability

Heart rate variability (HRV) more specifically sensitive to mental effort than HR. Several studies have found evidence that mental effort investment suppresses HRV, i.e. makes the heart beat more regularly (e.g. Mulder, 1980; Vicente et al., 1987). Total heart rate variability can be mathemically divided into its constituting sources of physiological variance. By means of this mathematical procedure, which is called spectral analysis, a distinction has been made in low (0.04 Hz), mid (0.10 Hz) and high (0.15 Hz+) frequency variations that, together with numerous other non-cardiovascular related variations, make up for the total HRV. Especially the mid frequency band, which is also called the 0.10 Hz component, is deemed to be related to mental effort. More specifically, in conditions where taskload is such that memory demands are neither too low nor too high, this 0.1 Hz suppression is thought to be indicative for increases in computational demands (computational effort). In driving Van Winsum et al. (1989) found navigation based on a map to be more loading than navigation by vocal messages, as measured by a decrease of the 0.10 Hz component. De Waard (1996) found that HRV profiles provide a reliable reflection of mental effort associated with different tasks. Waiting for a traffic light coincides with increases in variability, while driving on a roundabout corresponds to decreases in HRV.

6.5.3 Respiration

Although the way someone breathes is as a stand-alone measure not often used in research to measure workload, some aspects of breath are important for workload (Veltman, 1991). With an increase in mental workload, the pattern of respiration changes: people start to breathe faster and deeper. Respiration is also related to heart rate variability. Therefore, information about the breathing pattern is also important when trying to interpret heart rate variability. The main interaction between the respiratory and the cardiovascular system is through a direct link between the heart and the vagal nerve. Every inhalation causes a temporary decrease of the vagal inflow of the heart. This phenomenon, which is called respiratory sinus arrythmia (RSA), causes cyclic, respiration frequency-dependent variations in mainly the high frequency band of the HRV. If respiration frequency is kept at a constant rate, RSA 'contamination' of HRV can be filtered out, thereby increasing the reliability of spectral measures for the assessment of mental effort.

6.5.4 EEG/ERP

The EEG (Electro Encephalo Gram) is used as a reflection of the activity of the brain. In general, the EEG has a low frequency and a high amplitude under sleeping conditions. With increasing mental activity, the frequency increases and the amplitude decreases. One way to get an idea of the mental workload by means of the EEG is by means of Event Related Potentials (ERPs). Here, the EEG is related to a specific stimulus. In this, the P300 is important, which is a specific peak with a positive amplitude that is found about 300msec after the presentation of a stimulus, and which thought to be related to the proces of updating of memory contents. The amplitude of the P300 is used as a measure for the amount of attention that was required in order to evaluate the stimulus (Donchin, 1981), and can therefore be used as a measure for mental workload.

6.5.5 Eye measurements

Especially when one is working with visual displays (for instance visual in-vehicle displays) eye measurements can be valuable. As was discussion before, the number of glances and duration of glances to the display can be used, but also more physiological eye measures can be informative. The diameter of the pupil, direction of sight and the number of eye lid closures are indications for mental workload. The diameter of the pupil is known to decrease if mental workload increases. When measuring this, one has to be under well controlled experimental conditions, since for instance changes in light intensity will also change the pupil diameter. The direction of sight is an indication of what is considered to be important at that moment and the number of eye lid closures can be used as an indication for mental workload since it is shown that people close their eyes less if visual workload increases. Also, the eye is closed for a shorter period of time with increasing visual workload.

6.5.6 Hormones

Hormones that are sometimes used in workload research are adrenaline, nor-adrenaline and cortisol. Adrenaline is sensitive to mental as well as physical workload. Nor-adrenaline is especially associated with physical activity (Meijman, 1989). For an indication of mental workload, the amount of adrenaline is divided by the amount of nor-adrenaline. Adrenaline and nor-adrenaline can be identified in both urine and blood samples. Urine are disadvantageous because measurements can only be found every 2 hours. It is therefore not a sensitive measure to peaks in workload. If blood samples are used, more measurement moments are possible. A permanent cathether placed in the antecubital vein (bend of the elbow) makes it possible to take blood samples periodically in relatively non-intrusive way. However, it remains a problem this kind of measurements per se remains intrusive and also not always straightforward to interpret.

Cortisol and ACTH excretion can be perceived as the (sub)chronic continuation of the (nor)adrenergic response of the stress coping reaction. Cortisol is found in higher concentrations in the body if people are scared or have other negative emotions (Veltman, 1991). This can be the case of people execute a difficult task and they feel they do not have any control over the task (Frankenhaeser & Johansson, 1986). Cortisol can be found in the blood but also in saliva, which is a direct measure of the amount of cortisol in the blood (Vining et al., 1983). In research done by bus drivers on long rides to Spain, research showed that the amount of cortisol in the urine was higher just before driving back to the Netherlands (after one night sleep) then when going to Spain. This led to the conclusion that they were not as fit on the way back.

6.6 Prioritisation of the workload measures

Of the above mentioned workload measures, a prioritisation list is proposed. The list runs from very feasible to use in a future HASTE testbed to very unfeasible. It is suggested to try to include the first half (number 1 to 5) in the HASTE tests. Numbers 6 to 10 seem to be very unpractical to use.

1. Primary task performance
2. Subjective workload measures: NASA TLX and RSME
3. Peripheral Detection Task
4. Glance frequency and glance duration
5. Secondary task performance
6. Heart Rate
7. Occlusion
8. Subjective workload measures: SWAT and MCH
9. Heart Rate Variability + respiration
10. Hormones + EEG

7 Situation Awareness Measures

7.1 Situation Awareness in the driving context

The most quoted and widely accepted definition of Situation Awareness (SA) comes from Endsley (1988), who describes SA as "a person's state of knowledge or mental model of the situation around them". This definition is further expanded to include "the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning and the projection of their status in the near future". The concept has been most widely employed in environments where there are moderate to high rates of information, multiple tasks to perform, information that must be extracted from multiple sources, and there is a need to consider and plan for future actions. To date much of the SA work has been done in the field of aviation; however, the demands of the driving context are such that the investigation of SA in this context is appropriate. Matthews, Bryant, Webb and Harbluk (2002) have recently proposed a model for Situation Awareness in the driving context.

Endsley (1988) has characterised three fundamental aspects of SA, which are referred to as "levels." *Level 1 SA* comprises perception of the elements in the environment based on the processes of search, detection, recognition and identification of the relevant status, features, and attributes of the environment that are pertinent to the goals in hand. In the driving context, such features would be the location and dynamics of one's own and other vehicles. *Level 2 SA* comprises comprehension of the current situation, which is achieved through the integration and synthesis of information acquired through Level 1. The individual goes beyond the simple knowledge that the information is present in the situation and seeks to determine its significance and meaning. For example, a driver may detect a car pulling out from their right but in that context understands the significance of that event, i.e., that the driver of the vehicle intends to merge in the traffic. A driver who did not understand the event might change lanes abruptly expecting a crash. Finally, *Level 3 SA* entails projection of the future status of the situation, based on the ability of an individual to anticipate or envisage the future status or actions of the elements in the environment. A driver who is functioning at this level of SA might be involved in the prediction of traffic behaviour and updating route planning due to traffic conditions.

In the driving domain, problems of poor situation awareness could arise at any of the three levels. At Level 1, there may be detection errors arising from inattention, internal distraction or inappropriate lookout. At Level 2, drivers may fail to comprehend the meaning or significance of the detected information. At Level 3 there may be failures to appropriately extrapolate the current situation to the future and to plan appropriately. In sum, these SA errors tell us that drivers are not perceiving, not incorporating relevant info, or not projecting and planning appropriately.

A review and analysis of the literature on Situation Awareness and its applicability to the driving context is available in Matthews, Bryant, & Webb (1999).

7.2 Methods for assessing Situation Awareness

A variety of methods have been proposed to measure SA, each with advantages and disadvantages.

7.2.1 Direct measures

This first category of measures can be considered "direct" in the sense that the driver is asked directly about the information that is assumed to make up situation awareness. A comparison of the measures can be seen in Table 10.

7.2.1.1 SAGAT

The Situation Awareness Global Assessment Technique (SAGAT) has been developed by Endsley to assess all three levels of SA posited in the model. In this procedure a trial is suspended and all displays are blanked while the participant answers the relevant questions. Once the participant has responded to the questions the trial is continued. There are a number of beneficial aspects to this procedure. It has the benefit of asking the questions when the relevant information is still fresh for the participant. Probes can be made at various times during the trial. The questions can address a wide variety of aspects of the situation addressing all three levels of SA. The one potential drawback is that the ongoing event must be stopped for the questions to be asked. Endsley (2000) has conducted studies to address this concern and concluded that pauses in the order of 5 minutes do not have an appreciable impact. Endsley (1995; 2000) has provided recommendations for the implementation of the SAGAT technique.

7.2.1.2 Post event probes

In this approach, subjects are probed for recall after the event. Although convenient to administer, there are a number of concerns with this method. First, if the test session is long, subjects may have forgotten the relevant information. It is also possible that responses will be affected by inferences that the subject has made in the intervening period between the event and when the response is collected. Second, if the nature of the task environment is highly dynamic (as it often is in driving), then it is unlikely that post event probes will capture this complexity.

Table 10: Overview of two direct measures for Situation Awareness

	Interrupt task and probe (Freeze Technique)	***Post event recall probe***
Diagnosticity	High: with sufficient probes	Low
Practicality	Medium: requires large number of trials Appropriate timing of probes	High
Comments	Requires careful analysis of goals and tasks. Decision of what & when to test	Unsuitable for situations with highly dynamic and complex information content

Method	SAGAT	
Plus	Addresses 3 levels of SA Can include confidence	
Minus	Must stop to respond	Memory problems Single rating for total event

7.2.2 Subjective ratings

In contrast with performance or probe-based measures, subjects in this approach are required to generate numerical or scale ratings based on their subjective evaluation of certain aspects of their situation awareness. While this approach is relatively easy to administer, it relies on the person's subjective evaluation of SA. A major drawback is that subjective ratings assume that the subject knows what SA is and can accurately assess its level. The ratings are also subject to influences from other factors such as performance. A comparison of the measures can be seen in Table 11.

7.2.2.1 SART

The Situations Awareness Rating Technique (SART, Taylor, 1995) was devised for use in the aviation domain. It provides measurement of three basic dimensions of performance: demand (D), supply (S), and understanding (U) of the situation. Each of these factors is in turn broken down into contributing dimensions as follows:

Attention demand: complexity, variability, and instability
Attentional supply: arousal, concentration, division of attention, spare mental capacity
Understanding: information quality, information quantity, and familiarity

The major limitation of this approach is that situation awareness is confounded with workload in the use of measures of attentional demand and supply.

7.2.2.2 SA-SWORD

This approach uses a modified version of the Subjective Workload Dominance (SWORD) technique was developed to obtain subjective ratings of SA of the information provided by displays (Vidulich & Hughes, 1991). This method requires subjects to make comparative evaluations of the relative amount of SA provided by different display formats and consequently has limited applicability in assessing a driver's situation awareness of the larger driving environment.

7.2.2.3 Observer ratings

In this type of procedure, trained observers rate the degree of the driver's situation awareness. The primary limitation of this type of technique is that it is assumed that the observer would have knowledge of the driver's SA and that this could be inferred from the driver's behaviour.

Table 11: Overview of two subjective measures for Situation Awareness

	By subject	*By observer*
Diagnosticity	Low	Low
Practicality	High	High

Comments	Unsuitable for situations with highly dynamic and complex information content	Unsuitable for situations with highly dynamic and complex information content
Method	Rating scales: SART SWORD	
Plus	Addresses 3 levels of SA Can include confidence	
Minus	Assumes subject knows what total SA is Ratings influenced by performance	Assumes knowledge of subject's SA; Problem of inferring this from subject's behaviour

7.3 Indirect / performance measures and physiological methods

7.3.1 Indirect / performance measures

It is possible to introduce changes in the environment and then observe drivers' responses to those changes. This approach has the benefit of not interrupting the ongoing task in that the manipulations and responses can be made online. These sort of indirect or performance measures can be assessed through measures of hazard detection, accuracy of detection, or latency to respond measures. It is, however, difficult to interpret non-responses when these types of manipulations are made as one is unsure whether the event was not detected or detected and a response not given.

7.3.2 Physiological measures

A number of physiological procedures have been proposed to assess SA such as EEG and eyetracking. All of these are quite intrusive and require considerable instrumentation to carry out. The results can be difficult to relate to SA and interpret (see Table 12)

Table 12: Overview of indirect and physiological measures for Situation Awareness

	Indirect (Online manipulation and observation)	*Physiological measures*
Diagnosticity	High	Low
Practicality	Medium: requires a controllable test environment	Low
Comments	Does not require task interruption	
Method	Hazard detection, accuracy, latency measures	EEG, eye tracking
Plus	May access info that is non-verbal; or not easily accessed	

Minus	Interpretation of non response End result of many processes	Only know if information perceived, not comprehended, significance

7.4 Evaluation of SA measures

The SAGAT technique was assessed to be the preferable approach for measuring SA. Concerns, however, were raised with respect to stopping the session to ask the questions. It was suggested that the SAGAT technique be used but that the questions be asked during periods of low workload (e.g., while stopped at traffic lights, during low workload straight sections of road) to avoid the problems associated with stopping the session entirely.

8 Driver Groups

8.1 Introduction

One important division of drivers that should be made is between professional drivers and private drivers. The latter group is by far the biggest, depending on the definition, i.e. the separation between the groups is somewhat artificial. There is a fairly large group of drivers that do not drive by profession but drive a lot because of their profession. Since it is difficult, if not impossible to establish a criterion to differentiate between more or less professional nature of driving by this specific group, professional drivers are defined as drivers that drive by profession only, and include taxi drivers, and drivers of HGV and coaches. HASTE will focus on the non-professional drivers, in part because they are the vast majority and in part because they will be the main naïve users of IVIS (professional drivers are more likely to receive special training in system use).

A further subdivision might be made on:
- age
- experience, or mileage driven
- gender
- special characteristics, needs etc.

8.2 Age

Depending on the purpose of dividing people in age groups, many combinations are sensible. For the purpose of the HASTE project a starting point is a broad middle range packed together, from say 25 years old until at least 50 years of age, but perhaps 55 or even 60 years might do as well. In this broad range people are in the lower class of accident involvement, still perform quite well and have no serious needs or deficits. Between 18 and 25 years old most people obtain their driving license, and gain their first driving experience. Also in the TRAINER project and the ADVISORS project one of the criteria that is used to define drivers as novice driver is age 18-24. Together with low mileage, accident statistics show that in the first two years of driving people are generally more accident prone than in the next 30 to 40 years, and age is an important subdivider for participants. For reasons of parsimony, the age category after 60 years of age is left in one piece, although there are a few arguments for further subdivision into a young-old cohort (60-70), a medium-old cohort (70-80) and an old-old cohort (80+). Elderly drivers form an increasing proportion of the driving population, but also have an increased risk of an accident per mile driven (Waller, 1997).

In summary, when dividing into age groups we suggest to focus on three groups:
- 18-24 years old
- 25-60 years old
- 60+ years old

8.3 Experience

Driving experience, or mileage driven, is normally closely tied up with age though not necessarily. The majority of people in Europe obtain a driving license between 18 and say 21 years old, however, there is always a relatively small, but still substantial cohort of people that wait until later, for many reasons. Therefore experience should be treated as a separate factor here. Although experience can be captured in the variable years of driving experience (the TRAINER project defines novice drivers by the combination of experience < 1 year and age < 25 years), it is more common to use mileage. After getting a driving license, experience starts to build up, "one really starts to learn driving". As a rule of thumb for experience we choose a subpartition in *total* mileage driven in the order magnitude 10 times categories of 1000 miles:

- <10.000 miles in total = inexperienced
- >10.000 & < 100.000 miles = relatively experienced
- >100.000 = experienced
- >1.000.000 = very experienced

8.4 Gender

Driving and behaviour characteristics differ between male and female drivers, as well as factors like accident involvement, hospitalisation, traffic rule violations etc (e.g. Zhang, Lindsay, Clarke, Robbins, Glenn, & Yang, 2000; Laapotti & Keskinen, 1998), although these gender effects are not always confirmed (Norris, Matthews, & Riad, 2000, Lourens, Visser, & Jessurun, 1999). For this reason a discrimination between male and female may be useful in the project's pilot studies. It might well be that in the near future, when ITS or ADA systems will be personalised anyway, this differentiation turns out to be useful and well accepted.

8.5 Special characteristics or needs

Drivers with special needs (DSN) are normally defined as people that are physically handicapped in one way or the other. For this reason we prefer to use the term "drivers with special characteristics" which may include drivers that are unfamiliar with the local systems and situations, foreign drivers, drivers with extreme personality characteristics (e.g. extreme risk avoiding or seeking), drivers with an illness related condition, and of course the drivers with a physical handicap (DSN). As a first test, only if a device has characteristics that may be important for DSN (e.g. a monitoring device and people with sleep apnoea), or if there are potential conflicts for DSN in interacting with the device, these groups should be included.

8.6 Participants in the tests

If tests are to be performed with "the average driver" participants should meet the following criteria:

- Age: 25-50 years
- Gender: Both male and female
- Total driving experience: between 10,000 – 1,000,000 kilometres

If subgroups are to be tested, the following groups deserve attention:
Older drivers, aged 60+
Novice drivers, aged up to 24, total driving experience < 10,000 km, licence < 1 year

9 Performance Criteria

9.1 Theory

To investigate the effects of information systems in the experiments in terms of safety as planned, the "impact" or "influence" of the supplementary task that an information system always produces (perception and processing) has to be assessed. At least two viable approaches can be adopted, i.e. determining criteria for deciding whether the influence is on the safe side or not, and using validated secondary tasks for comparison with the influence of the information system at hand.

9.1.1 Criteria for driving safety

A first attempt to assess criteria for measures of task performance to determine whether driving is safe, has been published (Brookhuis & De Waard, 2001; Brookhuis & De Waard, in press). Criteria have been proposed to determine impaired driving, that are based on the effects of illegal levels of alcohol, intoxication, visual occlusion, driver inattention and prolonged journey time on driving behaviour. Criteria are characterised in terms of absolute levels, sort of golden yardsticks by nature, and relative change (Brookhuis & De Waard, in press).

Many researchers have defined impaired driving as a statistically significant increase or decrease of a particular measure of driving. For example, in studies of in-vehicle displays a significant increase of vehicle lateral deviation would be assessed as an impairment effect resulting from attention distraction. In fatigue research, a significant decrease of steering reversal rate may be interpreted as impairment, since the fidelity of steering control is reduced. This is a natural method for the categorisation of impaired driving in an experimental setting since most experiments contain control conditions or control subjects.

The inclusion of a control condition allows the experimenter to assess the *relative* impact of an independent variable on individual driving behaviour. Because "impaired" driving in this case is always compared to a control or a baseline condition, the only criterion of significant change is dictated by statistical testing. This categorisation may be contrasted with those measures that form *absolute* criteria to define impaired driving in general. For example, following a vehicle at 0.1 second time headway is unsafe and therefore "impaired" for everybody, as the minimum (physical) response time in a laboratory environment is at least approximately 0.2 seconds. In other words, absolute criteria are those fixed values that define the absolute red line of demarcation for impaired driver behaviour (Brookhuis, 1995).

Naturally there is a degree of inter-dependence between relative and absolute criteria. The position of "normal" or "baseline" driving is crucial in determining the relationship between both criteria. For example imagine a "cautious" driver who usually follows at 2 seconds time headway, contrasted with a "risky" driver who has a normal following headway of 1 second. Obviously, the difference that distinguishes baseline driving from impaired driving in absolute terms is much smaller for the risky driver than for the cautious driver. In other words, the risky driver leaves a much smaller range for impairment (which may be termed an *impairment margin*). The width of the impairment margin describes the degree of differentiation that separates impaired driving from normal driving. In turn, this separation defines the degree of overlap between the two distributions. The amount of overlap is important as it describes (a) the discriminative properties of the categorisation and (b) the potential for false alarms versus undetected impairment when designing system criteria around these data. The power of the technique is dependent on these phenomena.

9.2 Secondary tasks

As to primary and secondary tasks, there are, basically, three different theoretical formulations regarding the operator's capability to perform two different tasks simultaneously. According to "single-channel theory" (Broadbent, 1958; Welford, 1967, Kahneman, 1973), an operator can, in somewhat simplified terms, perform only one task at a time. Certain information systems may have negative effects on road safety if the system draws the driver's attention away from the actual traffic scene. According to "multiple resources theory" (Navon & Gopher, 1979, Wickens, 1984), the human cognitive system has separate resources or channels for different types of tasks. A human operator can, consequently, perform two different tasks simultaneously, provided that these tasks use different resources of the operator, such as vision and hearing. According to "connectionist control architecture" (Schneider & Detweiler, 1988), an operator can function both according to "single-channel theory" and "multiple resources theory". The operator's experience of the tasks involved, separately and combined, will determine if the two tasks interfere with each other. Drivers with long experience of driving can, according to this view, use the information presented by an IVIS better than inexperienced drivers.

Some IVIS applications, such as those working primarily through intervention may, on the contrary, have minimal effects on mental workload or may even reduce mental workload. The most commonly used mental workload measurement techniques are measures of task performance, self-reports, and physiological measures.

The first and most obvious method to measure driver mental workload is to measure the level of (primary) task performance, e.g. lateral and longitudinal control. Performance based measurement techniques are directly dependent on the capability of the operator to perform the task at target level. However, there is more to mental workload than performance alone. Demand appraisal, or self-setting of task goals, is very important in car driving. A further distinction can be made between primary task and secondary task measurement techniques.

All primary task measures are basically measures of speed or accuracy. Examples are lateral position deviation and driving speed. However, one primary task measure alone is not sufficient to draw conclusions regarding driver mental workload. When another task is included, secondary task measures can be taken. Two paradigms can be applied to dual-task performance. Within the "loading task paradigm" the instruction is to maintain performance on the secondary task. It is assumed that secondary task performance is maintained, even if decrements in primary task performance occur. Here, primary task performance measures can be used as indicators of mental workload. Within the "subsidiary task paradigm" the instruction is to maintain performance on the primary task. Here it is assumed that primary task performance is maintained, even if decrements in secondary task performance occur. Secondary-task performance varies with difficulty and indicates "spare capacity", provided that the secondary task is sufficiently demanding.

There are problems in using a secondary-task technique. According to "multiple resources theory" the largest sensitivity in secondary-task measures is achieved if the overlap in resources used is high. Consequently, spare capacity of the same resource should be required, since time-sharing is less efficient when the same resources are used. This overlap is at the same time a threat to undisturbed primary-task performance. Other problems are the omission of secondary-task performance when primary-task demands are very high, the operator's allocation policy, and lack of operator acceptance. The choice for a secondary task can be quite difficult in tasks approaching everyday performance, such as car driving.

One task is the PAced Serial Addition Task (PASAT, Gronwall & Sampson, 1974), which is in itself very sensitive to any external influence. In this task, every x-seconds a number between 1 and 6 is randomly selected and presented, either orally or visually. The subject is requested to add the last two numbers, and tell the result. Varying the inter-stimulus interval enables to fine-tune the task to subjects' capabilities and preferred levels of difficulty.

Another secondary-task measure, which has been used in a number of studies, is the PDT ("peripheral detection task"). It has advantages over other secondary task methods by demanding a minimum of attention: 1) peripherally presented stimuli do not require a controlled visual search; they activate attention automatically, and 2) detection of simple stimuli requires less processing capacity than identification.

Up to now the PDT-method has been used in two simulator studies and two field studies. The simulator studies, related to the design of driver support systems, have demonstrated the sensitivity of the method both to critical traffic scenarios and warning types (Van Winsum et al., 1999; Burns et al., 2000). Olsson (2000) successfully accomplished the transfer of the method to real driving. In a later study, also in real traffic (Harms et al., 2001), a decrement in professional drivers' PDT-performance was observed for two out of three possible modes of a navigation system as compared with memory based driving in a built-up area.

"Yet another secondary task is the so-called Sternberg task (Sternberg, 1966; 1975) In this task subjects are presented with a list of items, which is after a short delay, followed by a 'probe' item. Subjects' task is to decide if this probe item belongs to the original list. Results show a linear relationship between number of items in the list, and subject reaction time, with an increase of around 38 ms for each additional list item. This paradigm can be adapted to test memory for letters or numbers, presented visually or auditory. In addition, subjects' memory for difficult to name musical tones or difficult to label objects can also been tested". (In: Merat, 2002)

Some of the problems with secondary-task techniques can be overcome by using embedded secondary-task measures, a sub-task performed as a part of the whole task but which has lower priority than the primary-task, and for that reason primary-task intrusion is expected to be limited. Examples of embedded tasks are the number of radio communications that occur during a flight, or in driving, the frequency of rear-view mirror scans or car-following performance (see Brookhuis, De Vries & De Waard, 1991).

9.3 Towards an experimental design

The text below is divided into a section that addresses the independent variables, and one that describes the dependent variables. In the latter section the focus is not so much on the specific variables (as is the case in the independents section), but on the interpretation of the data in terms of absolute versus relative measures. The independent variables section mainly covers box ①, ② and ④ of Figure 23 whereas for the dependent variables section box ③ of the same figure is important.

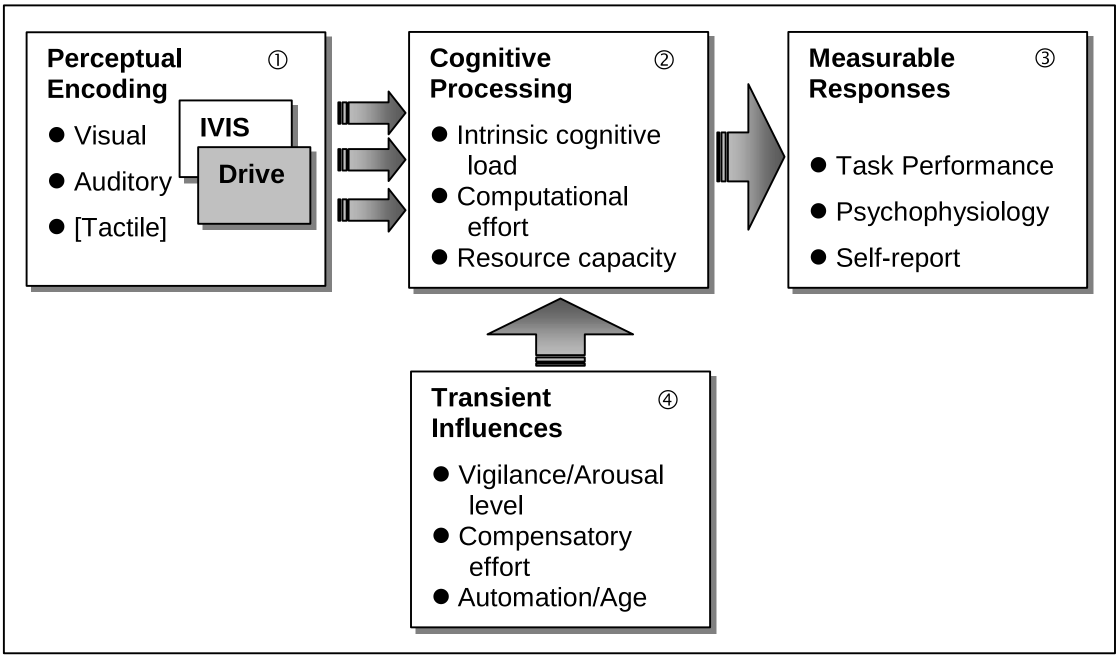


Figure 23: Modified version of Wickens' (1984) multiple resource model of attention

In Figure 23 The relationship between perceptual and cognitive load. The grey arrows from box ① to box ② represent the sensory channels. The little boxes inside box ① represent the primary driving task (grey) and the secondary IVIS task.

9.3.1 Independent variables

In order to gain insight into the relationship between performance and task load, it is important to define both terms accurately. Driving performance (which is treated in the next section) is the set of behaviours that is the resultant of, amongst others, the task load (box ① of Figure 23. Therefore, the concept of task load has to be differentiated. In the first place there are environmental factors (i.e. outside the vehicle) that determine the complexity of the driving task. Important factors in this respect are traffic density, darkness, and road type.

Next to environmental factors, there are driver-related factors. Someone's 'state-of-mind' determines what behaviour will be displayed in a certain situation. Important factors are driver state (which is in turn determined by numerous factors, for example vigilance level, time of the day, medication, health, personality etc, this refers to box ④ of Figure 23), driver age and driver experience. The factors time of the day (related to vigilance level), and use of medication generally affect the efficiency of the task performance. In this case the person has to compensate for the discrepancy between the actual and the required bodily condition. Another way of formulating this is that "one has to try harder to perform the task". This process of compensating for negative influences on task performance has been described as compensatory effort (Gaillard, 1992, Hockey, 1979, 1993). In Figure 23 box ④ refers to the process of compensatory effort.

Increased driving experience, which usually goes in parallel with age, can be described as a progressive automation of mental tasks. This enables these automated processes to be offloaded from controlled processing (e.g. Schneider et al., 1984). Automation also refers to box ④ of Figure 23.

The most important independent variable in the HASTE context though, is an additional in-vehicle factor, i.e. the IVIS. Basically the question is in what context an IVIS is either beneficial or deleterious for driving behaviour. It should be noted that neither context (personal and environmental) nor IVIS is a unitary concept in this statement. The term 'IVIS' is a collective noun for a wide variety of driving systems. Therefore, not only personal and environmental factors have to be systematically varied, but also the IVIS itself has to have various appearances in order to be an accurate reflection of what IVIS are, and will be on the market.

In order to meet this 'representativeness criterion' of the IVIS, three basic features have to be systematically varied. These are (1) the sensory modality of the IVIS output, (2) the degree of attention-demand, and (3) the degree to which a particular IVIS is perceptually and/or cognitively loading. IVIS output can be presented in various sensory modes, namely the visual, the auditory, and the tactile modality. Both multiple resource theory (Wickens, 1984; see also Figure 23) and experimental findings dictate that a secondary task is less intrusive/distracting for primary task performance if different sensory modalities are used. So next to practical reasons for meeting the aforementioned representativeness criterion (i.e. in practice IVIS output actually *is* presented in various sensory modalities) there are also psychological considerations to incorporate IVIS output as an independent variable.

Another important feature that discriminates between various IVIS is the degree to which they demand attention. In other words: how capacity-demanding is this particular IVIS in relation to the driving task. So again re-phrased: What is the capacity distribution between the primary task on the one hand and the secondary task on the other? This question can be answered by making use of the secondary task paradigm. Ideally, a driver should maintain his/her driving performance level at all times. If it is doubtful whether an IVIS is beneficial or not, the driver should stick to the driving task whenever this is required so, and spend only *spare* mental capacity on additional tasks. So, in terms of the secondary task paradigm the ideal driver behaves according to the subsidiary secondary task paradigm (see above). In the trade-off between primary and secondary task, the only secondary task performance will degrade in this ideal situation.

Having established this 'ideal' reference, it is necessary to study what (in terms of attentional distribution between primary and secondary task) behaviour a certain IVIS evokes in reality, and how this behaviour compares to that evoked by the ideal situation. In terms of experimental design this comparison can be accomplished by creating a condition *with* instruction ('maintain driving task at all times' - subsidiary task, ideal behaviour) and one *without* instruction (natural behaviour).

Thirdly and finally, a distinction between the perceptual and the cognitive demands of the IVIS in question has to be made. Naturally this dichotomy is rather artificial for perceptual and cognitive load are always present simultaneously. IVIS do differ though in the *emphasis* that lies on either perceptual or cognitive load.

Box ① of Figure 23 shows the perceptual modes that are relevant here: visual, auditory and, to a less extent, tactile information. Although crossmodal interactions (mutual weakening or reinforcement) have been described (McGurk & MacDonald, 1976), two tasks performed concurrently generally experience little interference when their sensory modalities differ. In general, the more similar the resource demands between the tasks, the greater the interference between them. According to Wickens (1984) each sensory modality has its own specific attentional resource, which would explain both performance trade-offs in unimodal dual tasks (utilisation of the same resource) and relatively unaffected performance in bimodal dual tasks (utilisation of different resources).

An important factor that affects the amount of cognitive processing is the complexity of a task. Task complexity can be defined as the knowledge and the cognitive processes that are intrinsically required for effectively executing a task. In Figure 23 the intrinsic task complexity refers to box ②.

In terms of experimental design box ① and ② both represent an independent variable. For the perceptual factor, the three conditions are visual, auditory and tactile IVIS output. For the cognitive factor three or more increasing levels of task complexity will have to be devised. This may be accomplished by varying memory or attention demands of the experimental task. Because of their strong interrelation, the factors perceptual load and cognitive load will in following text be referred to as one composite factor 'load'.

In what appearance could this factor 'load' be experimentally implemented? Firstly, in order to meet the first representativeness criterion, the testing method/paradigm must be modality-independent. In a broader sense, this criterion should also hold in future scenarios, i.e. the modality-independence also meets the technology-independence demand as stated in the T.A.

Secondly, because most impairment-safety relationships are exponential in form, it seems wise to opt for a fair number of task complexity (e.g. six) levels. This ensures a higher resolution when investigating the steep slope of the impairment-safety function.

This can be achieved by either changing the pacing (speed) or the memory demands of the task. Although the three sensory channels provide their own natural trichotomy, it remains question whether the perceptual quality should also be systematically varied.

Thirdly, by choosing a well-known paradigm (e.g. Sternberg task, PASAT task), the approach is more solid, less esoteric, and comparisons with existing literature can be made.

In sum, three constraints with respect to IVIS appearance and testing are: modality independence and technology independence, adjustable cognitive load (task complexity), and cross-comparability with existing literature.

9.3.2 Dependent variables

The classes or sets of simulated IVIS which are formulated above are intended to make predictions about real IVIS. The rationale for this is described in the text below. In this text, the simulated IVIS shall be referred to as 'surrogate IVIS'.

9.3.3 Rationale for real vs. surrogate IVIS comparison

The rationale for the use and interpretation of experimental results induced by a surrogate IVIS can be explained with a medical metaphor. A medical condition (syndrome) can be characterised by a set of concurring markers (symptoms). If a set of symptoms X is detected, this would point directly to syndrome X'. A precondition for this reasoning to be valid is that the symptoms, and/or their conjunction are specific — only then an accurate diagnosis is possible.

By carefully observing the reaction patterns (sets of potential symptoms) on the dependent variables, it could be established what variables are most indicative for ('is a set of symptoms X of') a particular surrogate IVIS (syndrome X'). If a comparable reaction pattern (effect sizes and directions) is also found with a real IVIS, this would imply comparability with the surrogate IVIS.

In short, the reasoning is: IF the goodness-of-fit between reaction patterns of a *particular* surrogate IVIS and a *particular* real IVIS if sufficiently great THEN comparison between both IVIS is justified. Statements of the form 'This (real) IVIS is of the surrogate class III type' can then be made. This rationale, which dictates that conclusions which are valid for the surrogate-IVIS (S-IVIS) are also valid for a real IVIS, provided that both IVIS types evoke the same reaction patterns, can schematically looks like this:

S-IVIS ⇒ REACTION PATTERN X
IVIS ⇒ REACTION PATTERN X
Therefore: S-IVIS ⇔ IVIS

The various conditions (classes) of surrogate IVIS are shaped by their sensory modality and their ratio of perceptuo-cognitive load. If *any* combination of the factors 'environment' or 'driver' with that particular class of IVIS results in unacceptably impaired driving behaviour, then the outcome of the in-vehicle IVIS pass-fail procedure is "Fail" (or at least a precaution of the form "You must not use this IVIS in darkness." should be made).

But how is "unacceptably impaired driving behaviour" defined? One important notion is the distinction between relative and absolute impact on driving behaviour (see above). How can one tell that driving is impaired when a certain measure has changed significantly over time?

Another notion is that each of the dependant variables also have to be looked at in mutual conjunction. A *single* unacceptable value on a single variable may be enough to render the driving behaviour unacceptably impaired, e.g. with involuntary lane departure. In this case one absolute value is sufficient to qualify driving performance as impaired. By analogy of logic gate functions this pass/fail criterion may be referred to as an 'OR' decision.

It can also be that one 'symptom' (increased value) is not serious on its own, but only in coherence with *multiple* other performance markers (i.e. box ③ of Figure 23). In this case the pass/fail decision is based on a multivariate decision. Multiple signs of on itself only slightly impaired driving behaviour together add up to a pattern of unacceptable driving behaviour . Therefore the pass/fail decision may be referred to as an 'AND' decision.

10 Conclusions

The following sections summarize the overall conclusions of this document, in the form of a set of recommendations for the experimental work to be conducted in HASTE. It is anticipated that there will be further refinement of the experimental protocol, as it is formalised and applied. There may also be practical and technical considerations that affect the implementation of the design as conceived here.

10.1 Secondary tasks

The literature review examined a number of candidate visual and cognitive tasks that could potentially be used as surrogates for an IVIS in the subsequent experiments in WP2 which are intended to examine driver performance under task load. Ideally tasks should have a number of features:

- They should have clear modality (visual, auditory)
- Tasks requiring cognitive processing should be distinguishable insofar as possible from those that require only visual attention
- They should be manipulable in terms of task difficulty

As well as visual and cognitive tasks, suitable manual tasks should be available to represent menu search and use via buttons, keys or touchscreens.

The review identified a number of suitable tasks which are designed to measure different levels of perceptual and cognitive ability in human operators. A list of factors which can control task difficulty have also been proposed. Finally, the incorporation of crossmodal paradigms, e.g. visual tasks supplemented by auditory signals, is recommended for some of the perceptual tasks, to compare the effect of unimodal and crossmodal secondary task presentation on driving workload. But the review also revealed that there was not a readily available set of tasks which could be manipulated to created various levels of difficulty.

Subsequently a set of pilot studies were carried out to identify whether the tasks could be developed so as to provide a good range of levels of perceptual and cognitive demand. Whilst each of the experiments was accomplished in isolation (i.e. as a primary task) for this stage of the project, it is anticipated that all tasks will be performed in combination with driving (i.e. as a secondary task). The kind of load placed by each of the chosen tasks (i.e. simple visual attention, memory for visual/auditory information, visual-manual co-ordination/memory), is one that is usually required by different forms of IVIS, either in isolation or collectively with other loads. Promising candidate tasks have been identified and further work is being carried out to refine the selection and range of difficulty as well as to confirm that the tasks are suitable for use while driving.

10.2 Scenarios

The review of situations and circumstances in which safety problems when interacting with an IVIS are likely to occur has identified a set of most critical parameters. They are:

- Driver age with a focus on older driver
- Driver vigilance in the form of boredom resulting in low vigilance
- Road type as urban, rural and motorway environment
- Junctions as a parameter of road infrastructure

- Pedestrian facilities as a parameter of road infrastructure
- Other road users as creating the possibility of crossing patterns
- Special events as potential hazards

Based on this review, *combinations* of circumstances can be prepared in order to create the test roads and events for the experiments in Workpackage 2. The work has also identified some parameters that will not be examined, is some cases because they are thought not to have an interaction effect with the use of an IVIS, in others because they were the focus of research being conducted elsewhere and could not for resource reasons be addressed properly within HASTE.

10.3 Driving performance measures

The focus of HASTE is on the influence of use of an IVIS on safety, i.e. on performance of the primary task of driving. "Performance" is here defined to include both the control level and the tactical level of the driving task. The selection of appropriate parameters and tools for measuring driving performance is clearly critical for the project. A preliminary set of mandatory driving performance measures has been identified. They are:

- Steering wheel reversal rate
- Lane exceedences
- Lateral position
- Standard deviation of lateral position
- Time to line crossing
- Speed
- Standard deviation of speed
- Time headway
- Distance headway
- Time to collision
- Reaction time to unexpected events
- Subjective ratings in the form of the Lund observer protocol

An additional set of optional measures has also been selected. Measurement problems have started to be addressed, so that data collected in the various sites can be compared.

10.4 Workload measures

The review looked at the whole range of workload measures, from performance on the primary tasks, through the application of secondary tasks to measure primary task load, to visual performance workload measures and subjective workload rating scales, and finally to physiological measures. A prioritisation of the various measures was made and the following are proposed as being desirable and practical:

11. Primary task performance
12. Subjective workload measures: NASA TLX and RSME
13. Peripheral Detection Task
14. Glance frequency and glance duration
15. Secondary task performance

10.5 Situation Awareness

Situation Awareness indicates driver understanding of the traffic situation. Loss of SA, will very likely lead to safety problems and interaction with an IVIS could potentially lead to such loss. The various measures of SA have been reviewed and the SAGAT technique was assessed to be the preferable approach for measuring SA. SAGAT normally involves freezing or obscuring the situation and asking subjects to report on the situation they have just been in. Concerns, however, were raised with respect to stopping the session to ask the questions. It was suggested that the SAGAT technique be used but that the questions be asked during periods of low workload (e.g. while stopped at traffic lights, during low workload straight sections of road) to avoid the problems associated with stopping the session entirely. This could permit the method to be used both in simulators and on real roads.

10.6 Participants

It has been concluded that, if tests are to be performed with “the average driver” participants should meet the following criteria:

- Age: 25-50 years
- Gender: Both male and female
- Total driving experience: between 10,000 – 1,000,000 kilometres

Some subgroups deserve particular attention: older drivers, aged 60 and over, and novice drivers, aged up to 24 with an annual distance travelled less than 10,000 km and holding a licence less than one year.

10.7 Test procedures

The final section reviews some of the issues and problems in selecting dependent variables, creating IVIS tasks, conducting the experiments and analysing the results. Further work on refining the IVIS tasks and piloting to be conducted ahead of the full experiments will, no doubt, help to clarify some of these issues. But not all can or should be resolved in advance. One major advantage of using a rich methodology is that, when analysing the results, a number of alternative approaches can be applied. It is not feasible or desirable to apply fixed benchmarks for safety or unsafety in advance. The work here has not resolved all the issues, nor was it intended to do so. The final specification of the tests and of the analytical approaches to be applied to the data will have to be resolved in the experimental work on driver performance and safety while using various IVIS.

11 References

Åberg L. & Rimmö P-A (1998): Dimensions of aberrant driver behaviour. *Ergonomics*, 41, 39-56.

Baddeley & Logie, (1999). Working Memory: the Multiple-Component Model. In *Models of Working Memory, Mechanisms of Active Maintenance and Executive Control.* A. Miyake & P. Shah (Eds). Cambridge, University Press, p. 28-62.

Baddeley, A.D. (1966). The capacity for generating information by randomization. *Quarterly Journal of Experimental Psychology*, 18, 119-129.

Baddeley, A.D. (1990). *Human Memory: Theory and Practice.* Lawrence Erlbaum Associates: Hove.

Baddeley, A.D. & Hitch, G.J. (1974). Working Memory. In G. Bower (Ed.), *The Psychology of Learning and Motivation* (pp. 47-90.). New York: Academic press.

Baddeley, A.D. (1986). *Working Memory*. Oxford University Press: New York.

Baddeley, A.D., Wilson, B.A., Watts, F.N. (Eds). (1996). *Handbook of memory disorders*. John Wiley & Sons, Chicester.

Baddely, A.D. & Hitch, G.J. (1974). Working memory. In: Bower, G. (Ed.) *The psychology of learning and motivation.* New York: Academic Press.

Barham, P.A.J., Alexander, J.J., Ayala, B.E., Oxley, P.R. (1993). *What are the benefits and safety implications of route guidance systems for elderly users?* European Commission DRIVE II Project V2031, Cranfield University.

Batic, N. & Gabassi, P.G. (1987). Visual dominance in olfactory memory. *Perceptual and Motor Skills*, 65, 88-90.

Berntson, G.G., Cacioppo, J.T. & Quigley, K.S. (1991). Autonomic determinism: The modes of autonomic control, the doctrine of autonomic space, and the laws of autonomic constraint. *Psychological Review*, 98, 459-487.

Bertelson, P. (1999). Ventriloquism: a case of crossmodal perceptual grouping. In *Cognitive Contributions to the Perception of Spatial and Temporal Events.* Ashersleben, G., Bachmann, T. & Musseler, J. (Eds). Amsterdam: Elsevier, p. 347-362.

Blumenthal, M. (1968). Dimension of the traffic safety problem. *Traffic Safety Research Review*, 1968, 12, pp. 7-12.

Boer, E.R. (2000). Behavioural Entropy as an Index for Workload*. Proceeding of the 44th annual meeting of the Human Factors and Ergonomic Society (HFES 2000)*. San Diego, CA. 2000.

Bourke, P.A., Duncan, J., Nimmo-Smith, I. (1996). A general factor involved in dual-task performance decrement. *Quarterly Journal of Experimental Psychology*, 49A, 525-545.

Broadbent, D.E. (1958). *Perception and communication*. London: Pergamon.

Brookhuis, K., Uneken, E. & Nilsson, L. (1999): ADVISORS common measures. Document RUG ID5_1_1.DOC, ADVISORS.

Brookhuis, K.A. & De Waard, D. (2001). Assessment of drivers' workload: performance, subjective and physiological indices. In P. Hancock & P. Desmond (Eds.) *Stress, workload and fatigue: theory, research and practice* (pp.321-333). New Jersey: Lawrence Erlbaum.

Brookhuis, K.A. (1995). Driver Impairment Monitoring System. In M. Vallet & S. Khardi, *Vigilance et Transports. Aspects fondamentaux, dégradation et préventation.* Lyon, France: Presses Universitaires de Lyon.

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

Brookhuis, K.A., de Waard, D., & Fairclough, S.H (submitted). *Criteria for driver impairment.*

Brouwer, W.H., Ickenroth, J.G., Ponds, R.W.H.M., Van Wolffelaar, P.C. (1990). Divided attention in old age. In: *European Perspectives in Psychology,* Wiley and Sons: New York N.Y.

Brouwer, W.H., Waterink, W., Van Wolffelaar, P.C., Rothengatter, T. (1991). Divided attention in experienced young and old drivers: lane tracking and visual analysis in a dynamic simulator. *Human Factors*, 33, 573-582.

Burns, P.C., Knabe, E, & Tevell, M. (2000). *Driver behavioral adaptation to collision warning and avoidance information.* IEA/HFES, International Ergonomics Association. San Diego. August.

Burns, P.C. (1997). *Navigation and the ageing driver*. PhD thesis, Loughborough University U.K.

Carr, D., Jackson, T.W., Madden, D.J., Cohen, H.J. (1992). The effect of age on driving skill. *Journal of the American Geriatrics Society*, 40, 567-573.

CBS (1999). *Doodlopende wegen.* (http://www.cbs.nl/nl/publicaties/publicaties/-algemeen/index/1999/no.3/doodlopende-wegen.pdf)

Channon, S., Baker, J.E., & Robertson, M.M. (1993). Working memory in clinical depression: An experimental study. *Psychological Medicine*, 23, 87-91.

Collins, D.J. (1997). *An examination of driver performance under reduced visibility conditions when using an in-vehicle signing and information system (ISIS).* (Thesis). Available on the web at: http://scholar.lib.vt.edu/theses/available/etd-32082539751141/ unrestricted/etd.pdf

Cooper, G.E. & Harper, R.P. (1969). *The use of pilot rating in the evaluation of aircraft handling qualities*. Moffett Field, CA. National Aeronautics and Space Administration, Ames Research Centre, NASA Report TN-D-5153.

De Renzi, E. & Nichelli, P. (1975). Verbal and nonverbal short term memory impairment following hemispheric damage. *Cortex*, 341-353.

De Vos, A.P., Godthelp, J., Theeuwes, J. & Verwey, W.B. (1996). *The influence of a Heading Control system on driver workload.* TNO-report TM-96-C048. TNO Human Factors, Soesterberg.

De Waard, D. & Brookhuis, K.A. (1991). Assessing driver status: a demonstration experiment on the road. *Accident Analysis and Prevention, 23*, 297-307.

De Waard, D. (1996). *The measurement of drivers' mental workload*. Dissertation Rijksuniversiteit Groningen.

De Waard, D., Hulst van der M., Hoedemaeker, M. & Brookhuis, K.A. (1999). Driver behaviour in an emergency situation in the Automated Highway System. *Transportation Human Factors*, 1(1), pp 67-82.

Donchin, E. (1981). *Surprise!….Surprise*? Psychophysiology 18, 493-513.

Donges, E. (1978). A Two-Level Model of Driver Steering Behavior. *Human Factors.* Vol. 20. No 6. pp 691-708.

Driver, J. & Spence, C. (2000). Multisensory perception: Beyond modularity and convergence. *Current Biology*, 10, R731-R735.

Duncan, J. and Humphreys, G.W. (1992). Beyond the search surface: Visual search and attentional engagement. *Journal of Experimental Psychology: Human Perception and Performance,* 18, 578-588.

Eimer, M. (2001). Crossmodal links in spatial attention between vision, audition and touch: evidence from event-related brain potentials. *Neuropsychologia*, 39, 1292-1303.

Elling, M. and Sherman, P. (1994). Evaluation of Steering Measures for Drowsy Driving. In: *Proceedings of the International Symposium on Automotive Technology and Automation*. Aachen, Germany, October-November.

Elvik, R., Mysen A. & Vaa, T. (1989). *Trafikksikkerhetshåndbok.* TOI Transportokonomisk institutt. Norks senter for samferdselsforkning.

Endsley, M.R. (2000). Direct measurement of situation awareness: Validity and use of SAGAT. In M.R., Endsley & D.J. Garland (Eds.) *Situation Awareness Analysis and Measurement* (pp. 147-173). New Jersey: Erlbaum.

Endsley, M.R. (1995). Measurement of situation awareness in dynamic systems. *Human Factors*, 37, 65-85.

Endsley, M.R.(1988). Design and evaluation for situation awareness enhancement. In: *Proceedings of the Human Factors Society 32nd Annual Meeting* (pp.97-101). Santa Monica, CA. Human Factors Society.

Fairclough, S.H., Ashby, M.C. & Parkes, A.M. (1993). In-vehicle displays, visual workload and usability evaluation. In: A.G. Gale, I.D. Brown, C.M. Haslegrave, H.W. Kruysse & S.P. Taylor (eds.), *Vision in Vehicles - IV* (p. 245-254). Amsterdam: Elsevier, North-Holland.

Fairclough, S.H., Ashby, M.C., Ross, T & Parkes, A. (1991). Effects of handsfree telephone use on driving behaviour. *Proceedings of the 24th ISATA international symposium on automotive technology and automation*. Croydon, England: Automotive Automation Limited. pp 403-409.

Farmer, E.W., Berman, J.V., & Fletcher, Y.L. (1986). Evidence for a visuo-spatial scratch-pad in working memory. Special Issue: Human memory. *Quarterly Journal of Experimental Psychology Human Experimental Psychology*, 38, 675-688.

Fell, D.L., & Black, B. (1997). Driver fatigue in the city. *Accident Analysis and Prevention, 29*, 463-469.

Feyer, A.-M., Williamson, A., & Friswell, R. (1997). Balancing work and rest to combat driver fatigue: an investigation of two-up driving in Australia. *Accident Analysis and Prevention, 29*, 541-553.

Frankenhaeser, M. & Johansson, G. (1986) Stress at work: psychobiological and psychosocial aspects. *International Review of Applied Psychology* (Sage, London, Beverly Hills and New Delhi) Vol. 35, 287-299.

Gabrielsen, K and Sherman, P. (1994). Drowsy Drivers, Steering Data and Random Processes. In: *Proceedings of the International Symposium on Automotive Technology and Automation.* Aachen, Germany, October-November.

Gaillard, A.W.K. (1992). Werkbelasting is nog geen stress. In: Keuss, P.G.J., Ten Hoopen, G. & Mannaerts, A.A.J. (Eds.) *Stress: werkbelasting, lichamelijke reacties, mentale belasting, slaapgebrek en meetmethoden.* Amsterdam/Lisse: Swets & Zeitlinger.

Geiwitz,P.J.?1966??Structure of boredom. Journal of Personality and Social Psychology , 3 , pp.592-600.

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

Green, P., Lin, B. & Bagian, T. (1993) *Driver workload as a function of road geometry: a pilot experiment* (Report UMTRI-93-39) Ann Arbor, MI, USA: The university of Michigan Transportation Research Institute.

Gronwall, D.M.A. & Sampson, H. (1974). *The psychological effects of concussion.* Auckland, New Zealand: Auckland University Press.

Haingey, D. & Westerman, S.J. (2000). Mobile (cellular) phone use and driving: a critical review of research methodology. *Ergonomics*, 44(2), 132-144.

Harms, L. & Ekrot R. (2001) Measuring Distraction from Navigation Instructions in Professional Drivers - A Field Study with the Peripheral Detection Task <<???>>

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* (p. 139-183). Amsterdam: North Holland.

Heuer, H. (1996). Dual-task performance. In O. Neumann and A.F. Sanders (Eds.), *Handbook of Perception and Action*, (pp. 113-155). Academic Press: London.

Hicks, T.G. & Wierwille, W.W. (1979). Comparison of five mental workload assessment procedures in a moving-base driving simulator . Human Factors, 34, 429-439.

Hill, S.G., Iavecchia, H.P., Byers, J.C., Bittner, A.C., Zaklad, A.L. & Christ, R.E. (1992). Comparison of four subjective workload rating scales. *Human Factors, 34,* 429-439.

Hockey, G.R.J. (1993). Cognitive energetical control mechanisms in the management of work demands and psychological health. In: Baddeley, A. & Weiskrantz, L. (Eds.). *Attention: selection, awareness and control* (pp. 329-345). Oxford: Clarendon Press.

Hockey, G.R.J. (1979). Stress and the cognitive components of skilled performance. In: Hamilton, V. & Warburton, D.M. (Eds*.). Human stress and cognition: an information processing approach* (pp. 141-178). Chichester: Wiley.

Hoedemaeker, M. & Kopf, M. (in press). Visual sampling behaviour when driving with adaptive cruise control. *Proceedings of the conference Vision in Vehicles IX,* August 2001, Australia.

Horne, J., & Reyner, L. (1999). Vehicle accidents related to sleep: a review. *Occupational and Environmental Medicine, 56,* 289-294.Horne, J., & Reyner, L. (1999). Vehicle accidents related to sleep: a review. *Occupational and Environmental Medicine, 56,* 289-294.

Hyden C. and Draskoczy M. (1992). Safety objectives for RTI application. DRIVE Project V1062 – Multilayered Safety Objectives. Bulletin 104, Lund.

Janke, M., Eberhard, J. (1998). Assessing Medically Impaired Older Drivers in a Licensing-Agency Setting. *Accident Analysis and Prevention, 30(3),* 347-361.

Jordan, P.W. & Johnson, G.I. (1993) *Exploring mental workload via TLX: the case of operating a car stereo whilst driving.* In A.G. Gale, I.D. Brown, C.M. Haslegrave, H.W. Kruysse & S.P. Taylor (Eds.), Vision in Vehicles-IV (pp. 255-262) Amsterdam: North-Holland.

Jousmaki, V. & Hari, R. (1998). Parchment skin illusion: Sound biased touch. *Current Biology,* 8, R190.

Kahneman, D. (1973). *Attention and effort.* Englewood Cliffs, N.J.: Prentice Hall. Mulder, G. (1980). *The heart of mental effort.* Thesis. Groningen: University of Groningen, The Netherlands.

Kalsbeek, J.W.H. & Ettema, J.H. (1963). Scored regularity of the heart rate and the measurement of perceptual load. *Ergonomics,* 6, pp 306

Klapp, S.T., Hill, M., Tyler, J., Martin, Z., Jagacinski, R. & Jones, M. (1985). On marching to two different drummers: Perceptual aspects of the difficulties. *Journal of Experimental Psychology: Human Perception and Performance,* 11, 814-828.

Klapp, S.T. (1981). Temporal compatibility in dual motor tasks II: Simultaneous articulation and hand movements, *Memory and Cognition,* 9, 398-401.

Korteling, J.E. (1993). Effects of age and task similarity on dual task performance. *Human Factors,* 35, 99-114.

Kulmala, R. (1995). Safety at rural three- and four-arm junctions. Development and application of accident prediction models. VTT Publications 233. VTT Offsetpaino, Espoo.

Kurokawa, K. & Wierwille, W.W. (1990). Validation of a driving simulation facility for instrument panel task performance. *Proceedings of the Human Factors Society 34th Annual Meeting* (p.1299-1303). Santa Monica, CA: The Human Factors Society.

Logie, R.H. (1986). Visuo-spatial processing in working memory. *Quarterly Journal of Experimental Psychology Human Experimental Psychology*, 38A, 229-247.

Lourens, P.F., Visser, J.A.M.M., & Jessurun, M. (1999). Annual mileage, driving violations and accident involvement in relation to drivers' sex, age and level of eductation. *Accident analysis and prevention, 31*, 593-597.

Lui, Y-C. (2001). Comparative study of the effects of auditory, visual and multimodality displays on drivers' performance in advanced traveller information systems, *Ergonomics*, 44(4), 425-442.

Lysaght, R.J., Hill, S.G., Dick, A.O., Plamondon, B.D., Linton, P.M., Wierwille, W.W., Zaklad, A.L., Bittner, A.C. & Wherry, R.J. (1989). *Operator workload: Comprehensive review and evaluation of operator workload methodologies*. Technical Report 851. United States Army Research Institute for the Behavioral and Social Sciences.

Matthews, M.L., Bryant, D.J., & Webb, R.D.G. (1999). Situation Awareness and Driving: Literature Review and Analysis. Transport Canada Contract Report. March 1999.

Matthews, M.L., Bryant, D.J., Webb, R.D.G., & Harbluk, J.L. (2002). Model for Situation Awareness and Driving. Transportation Research Record, 1779, 26-32.

McDonald W.A. & Hoffman E.R. (1980). Review of relationships beweeen steering wheel reversal rate and driving task demand. Human Factors. No 22. pp 733-739.

McGurk, H. & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, 264, 746-748.

Meijman, T. (1989) *Neuroendocrine activatie: Mobilisatie en compensatie*. In T. Meijman (Ed.) Mentale belasting en werkstress: een arbeidspsychologische benadering. Assen/Maastricht: Van Gorkum, the Netherlands.

Merat, N. (2002). *HASTE WP1 – Development of in-vehicle tasks - A Review of Secondary Task Methodology.*

Michon JA (1985). A critical view of driver behaviour models. What do we know, what should we do? In. L Evans & RC Schwing (eds): Human behaviour and traffic safety, Plenum Press, New York.

Miura, T. (1986). Coping with situational demands: A study of eye-movements and peripheral vision performance. In A. G. Gale, M. H. Freeman, C. M. Haslegrave, P. Smith, & S. P.Taylor (Eds.), *Vision in vehicles* (pp. 126–137). Amsterdam: North-Holland.

Miller, G.A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 63, 81-97.

Mulder, L.J.M. (1992). Measurement and analysis methods of heart rate and respiration for use in applied environments. *Biological Psychology*, 34, 205-236.

Mulder, L.J.M. & Mulder, G. (1987). Cardiovascular reactivity and mental workload. In: Kitney, R.I., Rompelman, O. (Eds.). *The beat-by-beat-investigation of cardiovascular function.* Oxford: Clarendon Press.

Mulder, G. (1980). *The heart of mental effort.* Thesis. Groningen: Rijksuniversiteit Groningen, The Netherlands.

Mulder, L.J.M. (1988). *Assessment of cardiovascular reactivity by means of spectral analysis.* Proefschrift. Groningen: Rijksuniversiteit Groningen.

Navon, D. & Gopher, D. (1979). On the economy of the human processing system. *Psychological Review*. 86, 214-255.

Nieminen, T. & Summala, H. (1994). Novice and experienced drivers looking behavior and primary task control while doing a secondary task. *Proceedings of the Human Factors and Ergonomics Society 38th Annual Meeting* (p. 852-856). Santa Monica, CA: The Human Factors Society.

Nilsson L, Harms L & B. Peters (2000). The effect of Road Transport Telematics. Manuscript Prepared for Barjonet et al. (eds) Traffic Psychology Today (hopefully in press).

Nilsson L., Alm H. & Janssen W. (1991). Collision Avoidance Systems - Effect of different levels of task allocation on driver behaviour. *VTI-särtryck*. Reprint from DRIVE-Project V1041 (GIDS).

Norman, D.A. & Shallice, T. (1986). Attention to action. Willed and automatic control of behaviour. R.J. Davidson, G.E. Schwartz & D. Shapiro (Eds.). *Consciousness and self-regulation*, (pp. 1-18). New York: Plenum Press.

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

O'Donnell, J.P., MacGregor, L.A., Dabrowski, J.J., Oestreicher, J.M., & Romero, J.J. (1998). Construct-validity of neuropsychological tests of conceptual and attentional abilities. *Journal of Clinical Psychology,* 50 (4), 596-600.

O'Hanlon, J.F. (1984) Driving performance under the influence of drugs: rationale for, and application of, a new test. British Journal of Clinical Pharmacology, 18, 121S-129S.

Olsson, S. and Burns, P.C. (2000*). Measuring Driver Visual Distraction with a Peripheral Detection Task.* Obtained from August 2000 The National Highway Traffic Safety Administration Driver Distraction Internet Forum on the World Wide Web. http://www-nrd.nhtsa.dot.gov/departments/nrd-13/driver-distraction/Papers20006.htm-#A6

Parasuraman, R. (1986). Vigilance, monitoring, and search. In: K. Boff, L. Kaufman, & J. Thomas (Eds.) *Handbook of Perception*. Volume 2. Cognitive Processes and Performance. (pp. 43.1-43.39). New York: Wiley.

Parker D., Reason J.T., Manstead A.S.R. & Stradling S.G. (1995): Driving errors, driving violations and accident involvement. Ergonomics, 38, 1036-1048.

Parkes, A.M., & Coleman, N. (1990). Route guidance systems: A comparison of methods of presenting directional information to the driver. In E.J. Lovesay (Ed.), *Contemporary Ergonomics*, (pp. 480-485). Taylor & Frances: London.

Parkin, A.J. (1993). *Memory: Phenomena, experiment and theory.* Oxford, Blackwell.

Perkins R.E. & Hill A.B. (1985). Cognitive and affective aspects of boredom. Br.J.Psychol., 76,(Pt 2), 221-234.

Phillips, W.A. & Christie, D.F.M. (1977a). Components of visual memory. *Quarterly Journal of Experimental Psychology*, 29, 117-133.

Phillips, W.A. & Christie, D.F.M. (1977b). Interference with visualisation. *The Quarterly journal of experimental psychology*, 29, 637-650.

Planek, T.W.(1981). The effects of ageing on driving abilities, accident experience and licensing. In H.C.Foot, Chapman, A.J., Wade, F.M. (Eds.), *Road Safety: Research and Practice*, London: Praeger Publishers.

Pohlmann, S. & Traenkle, U. (1994). Orientation in road traffic. Age-related differences using an in-vehicle navigation system and a conventional map. *Accident Analysis and Prevention, 26(6)*, 689-702.

Ponds, R., Brouwer, W.H., van Wolffelar, P.C. (1988). Age differences in divided attention in a simulated driving task. *Journal of Gerontology*, 43, 151-156.

Quinn, J.G. & Ralston, G.E. (1986). Movement and attention in visual working memory. *Quarterly Journal of Experimental Psychology,* 38A, 689-703.

Radeborg, Briem & Heman, (1999). The effect of concurrent task difficulty on working memory during simulated driving. *Ergonomics,* 42(5), 767-777.

Rämä, P. (2001). Effects of weather-controlled variable message signing on driver behaviour. VTT Publications 447. Otamedia Oy, Espoo.

Rantanen, E.M. & Goldberg, J.H. (1999). The effect of mental workload on the visual field size and shape. *Ergonomics,* 42(6). 816-834.

Reason, J. (1990). *Human error*. Cambridge University Press, Cambridge.

Recarte, M.A., & Nunes, L.M. (2000). Effects of verbal and spatial-imagery tasks on eye fixations while driving. *Journal of Experimental Psychology: Applied, 6*(1), 31-43.

Reid, G.B. & Nygren, T.F. (1988). The subjective workload assessment technique: a scaling procedure for measuring mental workload, in P. A. Hancock and N. Meshkati (Eds), *Human Mental Workload,* Amsterdam: Elsevier, p.185-218.

Risser, R. (1985). Behaviour in Traffic Conflict Situations. *Accident Analysis & Prevention* Vol.17, No.2, April 1985, S. 179-197.

Rockwell, T.H. (1988). Spare visual capacity in driving - revisited: new empirical results for an old idea. In: A.G. Gale, M.H. Freeman, C.M. Haslegrave, P. Smith & S.P. Taylor (eds.), *Vision in Vehicles - II* (p.317-324). Amsterdam: Elsevier, North-Holland.

Scerbo, M.W., Freeman, F. G., & Mikulka, P.J. (2000). A biocybernetic system for adaptive automation. In R. W. Backs, & W. Boucsein (Eds.), *Engineering Psychophysiology. Issues and Applications* (pp. 241-253). Mahwah, N. J.: Lawrence Erlbaum.

Shankar, V., Mannering, F. & Barfield, W. (1995). Effect of roadway geometrics and environmental factors on rural freeway accident frequencies. *Accident Analysis and Prevention,* 27(3),371-389.

Schneider, W. & Detweiler, M. 1988. The Role of Practice in Dual-Task Performance: Toward Workload Modeling in a Connectionist/Control Architecture. *Human Factors,* 30(5): 539-566.

Schneider, W., Dumais, S.T. & Shiffrin, R.M. (1984). In: R. Parasuraman & D.R. Davies (Eds.), *Varieties of attention* (pp. 1-27). New York: Academic Press.

Shinar, D. (1978). Psychology on the road. The human factor in traffic safety. John Wiley& sons, New York.

Sixsmith, J. (1990). *Driving experiences and new technology: evaluation and expectations of older drivers*. European Commission DRIVE Programme, Occassional paper no. 30.

Smith, E.E., & Jonides, J. (1997). Working memory: A view from neuroimagery. *Cognitive Psychology,* 33, 5-42.

Smyth, M.M. & Scholey, K.A. (1994). Interference in immediate spatial memory. *memory and cognition*, 22, 1-13.

Spence, C., Nicholls, M.E.R., Gillespie, N. & Driver, J. (1998). Cross-modal links in exogenous covert spatial orienting between touch, audition and vision. *Perception and Psychphysics,* 60, 544-557.

Sternberg, S. (1975). Memory scanning: New findings and current controversies. *Quarterly Journal of Experimental Psychology,* 27, 1-32.

Sternberg, S. (1966). High speed scanning in human memory. *Science,*153, 652-654.

Taoka, G.T. (1990). Duration of drivers glances at mirrors and displays. *ITE Journal (October 1990)*, 35-39.

Taylor, R. M., Selcon, S. J. and Swinden, A. D. (1995). Measurement of situational awareness and performance. In: Fuller, R., Johnston, N. and McDonald N. (eds) Human Factors in Aviation Operations. Avebury: Aldershot.

Tijerina, L., Gleckler, M., Stoltzfus, D., Johnston, S., Goodman, M.J. & Wierwille, W.W. (1999). A Preliminary Assessment of Algorithms for Drowsy and Inattentive Driver Detection on the Road. *DOT Report: HS 808.*

Towse, J.N. (1998). On random generation and the central executive of working memory. *British Journal of Psychology,* 89, 77-101.

Treisman, A.M. (1988). Features and objects: The fourteenth Bartlett memorial lecture. *Quarterly Joumal of Experimental Psychology,* 40A, 201-237.

Treisman, A.M. (1992). Spreading suppression or feature integration? A reply to Duncan and Humphreys (1992). *Journal of Experimental Psychology: Human Perception and Performance*, 18, 589-593.

Tsang, P.S. & Wickens, C.D. (1988). The structural constraints and resource control of resource allocation. Human Performance, 1, 45-72.

Underwood, G. & Radach, R. (1998). Eye guidance and visual information processing: Reading, visual search, picture perception and driving. In: Underwood, G. (ed.) *Eye Guidance in Reading and Scene Perception* (pp. 1-27). Oxford: Elsevier.

Van Winsum, W. & Hoedemaeker, M. (2000). *A road test of a prototype satellite system for in-vehicle menu control*. TNO-report, TM-00-C003. Soesterberg, The Netherlands.

Van Winsum, W., Martens, M. & Herland, L. (1999). *The effects of speech versus tactile driver support messages on workload, driver behaviour and user acceptance*. TNO-report, TM-99-C043. Soesterberg, The Netherlands.

Van Winsum, W. & Godthelp, H. (1996). Speed Choice and Steering Behaviour in Curve Driving. Human Factors. Vol 38. No 3. pp 434-441.

Van Winsum, W., Van Knippenberg, C. & Brookhuis, K. (1989). Effect of navigation support on drivers' mental workload. *Current issues in European transport, vol.1, Guided transport in 2040 in Europe*. London: PTRC Education and Research Services. pp 69-64.

Veltman, J.A. & Gaillard, A.W.K. (1998). Physiological workload reactions to increasing levels of task difficulty. *Ergonomics,* 41(5), 656-669.

Veltman, J.A. (1991). *Mental workload among pilots: Theoretical background and literature review* (Report IZF 1991 A-33, in Dutch) Soesterberg, the Netherlands: Institute for Perception TNO.

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

Verwey, W.B., Brookhuis, K.A. & Janssen, W.H. (1996). *Safety effects of in-vehicle information systems* (Report TM-96-C002). Soesterberg, The Netherlands: TNO Human Factors Research Institute.

Verwey, W.B. & Veltman, J.A. (1995). *Measuring workload peaks while driving. A comparison of nine common workload assessment techniques* (Report TM-1995-B4). Soesterberg, The Netherlands: TNO Human Factors Research Institute.

Vicente, K.J., Thornton, D.G. and Moray, N. (1987). Spectral analyses of sinus arrhythmia: a measure of mental effort. *Human Factors*, 29, pp 171-182.

Vidulich, M.A. & Hughes,E.R. (1991). Testing a subjective metric of situation awareness. In Proceedings of the Human Factors Society 35th Annual Meeting (pp.1307-1311). Santa Monica, CA. Human Factors Society.

Vining, R.F., McGinley, R.A., Maksvytis, J.J. & Ho, K.Y. (1983). *Salivary cortisol: a better measure of adrenal cortisol function than serum cortisol*. Annual Clinic Biochemistry 4, 20, 329-335.

Vroomen, J.& de Geider, B. (2000). Sound enhances visual perception: Cross-modal effects of auditory organization on visual perception. *Journal of Experimental Psychology*.

Watanabe & Shimojo (2001). When sound affects vision: Effects of auditory grouping on visual motion perception. *Psychological Science*, 12(2), 109-116.

Welford, A. T. (1967). Single-channel operation in the brain. *Acta Psychologica*, 27, 5-22.

Wickens, C.D. (1992). *Engineering Psychology and Human Performance.* 2nd Edition, HarperCollins: New York.

Wickens, C.D. (1984). Processing resources in attention. In: R. Parasuraman & D.R. Davies (Eds.), *Varieties of attention* (pp. 63-102). New York: Academic Press.

Wierwille, W.W. (1993a). An initial model of visual sampling of in-car displays and controls. In: A.G. Gale, I.D. Brown, C.M. Haslegrave, H.W. Kryusse & S.P. Taylor (eds.), *Vision in Vehicles - IV* (p. 271-280). Amsterdam: Elsevier, North-Holland.

Wierwille, W.W. (1993b). Visual and manual demands of in-car controls and displays. In: B. Peacock & W. Karwowski (eds.), *Automotive ergonomics* (p. 299-320). London: Taylor & Francis.Wiethoff, M. (1997). *Task analysis is heart work*. Dissertation university of Groningen, Delft university Press, The Netherlands.

Wierwille, W.W. (1990). A review of the age effects in several experiments on instrument panel task performance, *SAE Transactions, 900190*.

Wierwille, W.W., Antin, J.F., Dingus & Hulse, M.C. (1988). Visual attention demand of an in-car navigation display system. In: A.G. Gale, M.H. Freeman, C.M. Haslegrave, P. Smith & S.P. Taylor (eds.), *Vision in Vehicles - II* (p.307-316). Amsterdam: Elsevier, North-Holland.

Wiethoff, M. (1997) *Task analysis is heart work. The investigation of heart rate variability: a tool for task analysis in cognitive work.* Groningen University: PhD Thesis. Delft: Delft University Press. ISBN 90-407-1607-2.

Williams, L.J. (1995). Visual field tunnelling in aviators induced by memory demands, *The Journal of General Psychology*, 122, 225-235.

Wilson, G.F. & Eggemeier, F.T. (1991). Psychophysiological assessment of workload in multi-task environments. In: D.L. Damos (Ed.), *Multiple-task performance*. London: Taylor & Francis. pp 329-360.

Zeitlin, L.R. (1993). Subsidiary task measures of driver loading: a long term field study, *Transportation Research Record*, 1403, 23-27.

Zellner, D.A. & Kautz, M.A. (1990). Colour affects perceived odour intensity. *Journal of Experimental Psychology: Human Perception and Performance*, 16, 391-397.

Zhang, J., Lindsay, J., Clarke, K., Robbins, G. & Mao, Y. (2000). Factors affecting the severity of motor vehicle traffic crashes involving elderly drivers in Ontario. *Accident Analysis and Prevention* 2000 Jan; 32(1):117-25.

Zijlstra, F.R.H. & Van Doorn, L. (1985). *The construction of a scale to measure perceived effort*. Delft, The Netherlands: Department of Philosophy and Social Sciences, Delft University of Technology.

Zwahlen, H.T. & DeBald, D.P. (1986). Safety effects of sophisticated in-vehicle information displays and controls. *Proceedings of the Human Factors Society 30th Annual Meeting* (p. 256-260). Santa Monica, CA: The Human Factors Society.

Zwahlen, H.T. & Balasubramanian, K.N. (1974). A theoretical and experimental investigation of automobile path deviations when driver steers with no visual input. In: *Driver Performance, Transportation Research Record 520* (p. 25-37). Washington, DC: Transportation Research Board.

Appendix 1: Driving and Medication

Road traffic safety is a significant issue for society. Europe must develop a rational transport policy to support interventions to reduce accidents. Driver impairment is a significant source of accident risk. There is increasing evidence that drugs and medicines may impair driver functioning and increase accident risk. However, these issues will not be considered within the HASTE mainly for the reason that concerted efforts in this area are being or have been undertaken. The following *projects* are particularly relevant.

CERTIFIED (completed 2001)

The main objectives of this project were:

- To review information regarding impairment and accident risk from drug and medicine use in a road transport context.
- To review existing impairment testing method in relevant domains;
- To develop concepts for roadside impairment testing based on psychomtric and behavioural measures;
- To provide preliminary data on concept validity;
- To recommend verification methodology to evaluate testing protocols (based on user, legal and operational requirements);
- To identify key information and issues relevant to transport policy and implementation strategy for roadside testing in Europe.

A number of deliverables from this project are now available on the project website: http://www.psyc.leeds.ac.uk/certified/

IMMORTAL (2002-2004)

IMMORTAL specifies a research programme concerning the accident risk associated with different forms of driver impairment and the identification of ‘tolerance levels’ applied to licensing assessment and roadside impairment testing (including drug screening).
The main objectives of IMMORTAL are to:

- Investigate the influence of chronic and acute impairment factors on driving performance and accident risk;
- Recommend criteria (‘tolerance levels’) for high risk categories of impairment
- Provide key information to support formulation of European policy on licensing assessment and roadside testing.

The results will provide comprehensive knowledge concerning the influence of acute and chronic impairing factors that may be used in policy decisions. The results will also support recommendations on how to (a) examine chronically impaired people seeking (re)licensing, and (b) assess driver for acute impairment (at roadside).

SAVE (1996-1998)

SAVE (TR1047) was a Transport Telematics EU project that aimed to develop an integrated system capable of detecting driver status problems such as those listed above, which may indicate an imminent danger of an accident or other emergency.
The project work analysed drivers needs, expectancies and acceptability of the proposed

system. In addition workshops with experts and users panels were organised to supplement the user requirements work. The first stage of the development programme developed of subsystems to detect drunk driving, fatigue and critical incidents in real time, which were called Driver Impairment Monitoring (DIM). DIM were based on a number of indirect techniques, including:

- driver profile identification and recognition of serious deviations from it,
- estimation of the changes in the vehicle position on the road,
- detection of relevant road safety aspects violation,
- observation of the driver's face, looking particular at eye blinks, to detect drowsiness or inattention
- measurement of steering wheel grip force.

All signals of the various subsystems were post-processed and prioritised by an Integrated Monitoring Unit (IMU) which assessed the cause and severity of the impairment and provided an optimised interface between the driver and the system. For more information visit: http://www.iao.fraunhofer.de/Projects/SAVE/save/saveinit.htm

AWAKE (ongoing)

The objective of AWAKE is to increase traffic safety by reducing the number and the consequences of traffic accidents caused by driver hypovigilance. In order to achieve this objective, AWAKE intends to develop an unobtrusive, reliable system, which will monitor te driver and the environment and will detect in real time hypovigilance, based on multiple parameters. The system will achieve enhanced reliability and minimised false alarm rate, by supporting continuous, instead of discrete, event-related driver monitoring, strong system personalisation to driver characteristics and traffic situation awareness. In case of hypovigilance, the system will provide an adequate warning to the driver, with various levels of warnings, according to the estimated driver's hypovigilance state and also to the estimated level of traffic risk. This system will operate reliably and effectively in all highway scenarios. For more information visit: http://www.awake-eu.org/index2.htm

In addition to these past and ongoing projects, there has been substantial other work in the field that has looked specifically at the separate effects of various medication and illegal substances on driving performance. It is not feasible to include medication as an option in the HASTE project for time and resource reasons, instead a watching brief will be made on the ongoing projects of relevance.

Appendix 2: Preliminary translation of Observer Protocol (University of Lund, M. Hjälmdahl)

<table>
<tr><th colspan="3">Variable</th><th>Section 1</th><th>Section 2</th><th>Section…</th></tr>
<tr><td colspan="2" rowspan="3">Yielding behaviour</td><td>OK</td><td></td><td></td><td></td></tr>
<tr><td>Hesitant</td><td></td><td></td><td></td></tr>
<tr><td>Short gap / dangerous</td><td></td><td></td><td></td></tr>
<tr><td colspan="2" rowspan="6">Behaviour towards unprotected road-users at crossing</td><td>Give way early</td><td></td><td></td><td></td></tr>
<tr><td>Give way late</td><td></td><td></td><td></td></tr>
<tr><td>Unprotected RU forces driver to give way</td><td></td><td></td><td></td></tr>
<tr><td>Unprotected RU waits at road-side</td><td></td><td></td><td></td></tr>
<tr><td>Forces unprotected RU to stop</td><td></td><td></td><td></td></tr>
<tr><td>Puts unprotected RU in danger</td><td></td><td></td><td></td></tr>
<tr><td colspan="2" rowspan="4">Use of indicator</td><td>OK</td><td></td><td></td><td></td></tr>
<tr><td>Too early</td><td></td><td></td><td></td></tr>
<tr><td>Too late</td><td></td><td></td><td></td></tr>
<tr><td>Not at all</td><td></td><td></td><td></td></tr>
<tr><td rowspan="6">Lane usage</td><td>Choice / change</td><td>OK</td><td></td><td></td><td></td></tr>
<tr><td rowspan="3">Choice of lane before junction / obstacle</td><td>Too early</td><td></td><td></td><td></td></tr>
<tr><td>Too late</td><td></td><td></td><td></td></tr>
<tr><td>Wrong lane</td><td></td><td></td><td></td></tr>
<tr><td rowspan="2">Change of lane</td><td>Dangerous</td><td></td><td></td><td></td></tr>
<tr><td>Hesitant</td><td></td><td></td><td></td></tr>
<tr><td colspan="2" rowspan="3">Speed adaptation in front of junction / obstacle</td><td>OK</td><td></td><td></td><td></td></tr>
<tr><td>Late / hard braking</td><td></td><td></td><td></td></tr>
<tr><td>Bad</td><td></td><td></td><td></td></tr>
<tr><td colspan="2" rowspan="2">Speed</td><td>Too fast for the situation</td><td></td><td></td><td></td></tr>
<tr><td>Too slow for the situation</td><td></td><td></td><td></td></tr>
<tr><td colspan="2" rowspan="2">Overtaking</td><td>Overtake</td><td></td><td></td><td></td></tr>
<tr><td>Is overtaken</td><td></td><td></td><td></td></tr>
<tr><td rowspan="2">Priority</td><td colspan="2">Insist on his own priority and causes danger</td><td></td><td></td><td></td></tr>
<tr><td colspan="2">Renounce his own priority</td><td></td><td></td><td></td></tr>
</table>

Appendix 3: TRIP

Test Ride for Investigating Practical Fitness to Drive (TRIP)

Brouwer et al., University of Groningen.

Belgian version *(English translation by De Raedt, Free University Brussels)*

General instructions:

- Every question is followed by some space for explanation. Always fill in something when "insufficient" is given. For example, describe how a dangerous situation developed because of deviations in the concerning category.
- I = insufficient (score 1); D = doubtful (score 2); S = sufficient (score 3); G = good (score 4)
- Total score: add the scores of scales 1-11

1. Lateral position on the road

How would you describe the average lateral positioning on the driving lane (on a regular two-lane road) ?

[1] Too much to the left of the middle

[2] To the left of the middle

[3] In the middle

[2] To the right of the middle

[1] Too much to the right of the middle

How is the steadiness of steering (swaying and drifting away) ?

On straight roads					
- speed <= 50 km/u	a. without distraction	I	D	S	G
	b. with distraction	I	D	S	G
- speed > 50 km/u	a. without distraction	I	D	S	G
	b. with distraction	I	D	S	G
In curves					
- speed <= 50 km/u	a. without distraction	I	D	S	G
	b. with distraction	I	D	S	G
- speed > 50 km/u	a. without distraction	I	D	S	G
	b. with distraction	I	D	S	G

Explanation:

2. Lane position change

How is the position choice in the following situations ?

Lane choice to drive straight on	I	D	S	G
Lane choice to turn right	I	D	S	G
Lane choice to turn left	I	D	S	G
Lane choice at roundabouts	I	D	S	G

Explanation:

3. Distance from car in front

How would you classify the car following style of the driver ?

[1] Follows other cars at a too short distance

[2] Follows other cars at a short distance

[3] Follows other cars at an average distance

[2] Follows other cars at a long distance

[1] Follows other cars at a too long distance

How well is the following distance adapted to variations of speed of the cars ahead ?

- In town areas:	I	D	S	G
- Outside town areas:	I	D	S	G

Explanation:

4. Speed

How would you classify the driver in terms of his style of speed choice ?

[1] Has a driving style which is too fast

[2] Has a fast driving style

[3] Is average in terms of speed choice

[2] Has a slow driving style

[1] Has a driving style which is too slow

How good is the driver's adaptation of speed to the circumstances?

- In town areas:	I	D	S	G
- Outside town areas:	I	D	S	G

Explanation:

5. Visual behaviour & communication

(Head and eye movements and eye contact with other road users)

How do you judge the head and eye movements ?

When driving straight ahead	I	D	S	G
At crossings when crossing a main road	I	D	S	G
At crossings and junctions when turning right	I	D	S	G
At crossings and junctions when turning left	I	D	S	G
In curves	I	D	S	G
Use of the inside mirror	I	D	S	G
Use of the left outside mirror	I	D	S	G
Observation in the blind angle	I	D	S	G
Communication with other road users	I	D	S	G

Explanation:

6. Traffic signals

(Traffic lights and traffic signs)

Are traffic signals well perceived and responded to ?

Perception	I	D	S	G
Reaction	I	D	S	G

Explanation:

7. Mechanical operations

(Fluency and timeliness of mechanical operation)

Operating the accelerator	I	D	S	G
Operating the brakes	I	D	S	G
Quality of steering	I	D	S	G

Explanation:

8. Anticipation

(Tactical anticipatory behaviour, e.g. slowing down when a pedestrian approaches a zebra crossing)

With regard to changing road situations	I	D	S	G
With regard to changing traffic situations	I	D	S	G

Explanation:

9. Understanding, perception and quality of traffic participation

How would you judge general insight, sense of context and practical implementation in traffic situations ?

Quality of traffic perception and traffic insight	I	D	S	G
Quality of traffic participation	I	D	S	G

Explanation:

10. Turning left

Merging onto a main road which has right of way
(no traffic lights).
How would you judge the following specific situations ?

When approaching the crossing or junction

- Adaptation of speed	I	D	S	G
- Mirror use and looking over the shoulder	I	D	S	G
- Operating the direction indicator	I	D	S	G
- Positioning on the roadway	I	D	S	G
- Looking (head movements)	I	D	S	G
- Effectivity of looking (having seen other traffic)	I	D	S	G

At the crossing or junction

- Positioning the car	I	D	S	G
- Looking (head movements)	I	D	S	G
- Effectivity of observation (having seen other traffic)	I	D	S	G
- Application right of way rules	I	D	S	G
- Swiftness and fluency of perception and action	I	D	S	G

Explanation:

11. Joining the traffic stream

On highway.

How would you judge the following specific situations ?

- Making speed on the acceleration lane	I	D	S	G
- Looking sideways	I	D	S	G
- Adaptation of speed to other traffic	I	D	S	G
- Operating the direction indicator	I	D	S	G
- Driving into the main lane	I	D	S	G

Explanation:

Conclusions

Circumstances test ride

Driver instructor:

Duration:

Particular circumstances:

Was the test ride broken off early because of traffic safety reasons ? Y N

Expert judgement

[ ] Unlimited fitness to drive without driving lessons

[ ] Unlimited fitness to drive after driving lessons (_____ lessons advised)

[ ] Fit to drive with restraints in time: _____

[ ] Fit to drive with restraints in place:_____

[ ] Fit to drive with technical adaptations:_____

[ ] Not yet fit to drive; propose driving lessons:_____

[ ] Unfit to drive / after driving lessons (number of lessons followed _____)

[ ] Unfit to drive / without driving lessons

Appendix 4: Wiener Fahrprobe

Free observer coding sheet

	Approaching a place of interaction
	checks the situation
	drives with anticipation
	does not drive with anticipation
	inappropriate speed
	inaccurate lane choice

Interaction			
	insists on right of way		does not insist on right of way
	does not allow to continue/merge		allows to continue/merge
	does not reduce speed		reduces speed
	presses other cars		
	obstructs others (e.g. at crossings, etc.)		
	others move into the safety distance of the subject		
	turns right near oncoming traffic		
	obstructs others when turning right		
	obstructs others when turning left		
	makes other road users decelerate		
	makes others accelerate		
	impedes cyclists/pedestrians		
	endangers cyclists/pedestrians		

	Overtakes or changes lane
	cuts up
	too small lateral distance
	Aborted

	Conflict
	subject provokes conflict
	subject does not provoke conflict

	Communication		comments
	Positive		positive
	Negative		negative

Description

Coding observer coding sheet

Standardised observation			
	Overtaking or		Speed
	lane change		Inappropriate
	Correctly		Inappropriate for road geometry
	not correct		too fast near VRUs
	in spite of oncoming traffic		in the platoon
	without sufficient vision		without platoon
	While forbidden		above the speed limit
	because of a stationary obstacle		at / below the limit
	lane change in time		considerably slower than the limit
	uses right lane mainly		brakes abruptly
	uses left lane mainly		unsteady speed
	Use of the indicator		Distance to the road user ahead
	indicates in time		correct
	Does not indicate		too short
	Does not indicate in time		Behaviour at traffic lights
	indicates ambiguously		drives against red
	Lane use		drives against amber
	inaccurate, weaving		does not start when it is green
	extremely on the right side of the lane		starts too early
	extremely on the left side of the lane		Checks the situation with respect to other road users
	cuts the curve		yes
	Lane choice for proceeding		no
	correct		Number of cars overtaking
	in time		
	at the last moment		
	Incorrect		
	Behaviour when merging		
	safe		
	Unsafe		
	with traffic		
	without traffic		
	inappropriate speed		

Appendix 5: Examples of SAGAT Questions

Level 1 Questions: Perception

- Did you receive a message in the last 30 s?
- What is your current speed?
- Which lane are you currently in?
- What colour was the light at the last intersection?
- On what side of the road were the last pedestrians?
- How many pedestrians at the last intersection
- Is there currently a vehicle ahead of you in the right lane?

Level 2 Questions: Comprehension

- What action did the most recent message call for?
- What is the speed of the car ahead relative to you?
- What was the last warning you received from the system?

Level 3 Questions: Predict Future Events

- Should you be driving less than 50kph over the next 30 s?
  Will you need to be in the left lane in the next 30s?