INFORMATION SOCIETY TECHNOLOGIES (IST)
PROGRAMME

AIDE
IST-1-507674-IP

Specification of a secondary task to be used in Safety Assessment of IVIS

Deliverable No. (use the number indicated on technical annex)		**D2.2.3**	
SubProject No.	**SP2**	SubProject Title	**Evaluation and Assessment Methodology**
Workpackage No.	**WP2.2**	Workpackage Title	**Driver workload and distraction assessment methods and tools**
Activity No.	**2.2.3**	Activity Title	**Potential of advanced secondary task methodology for driver assessment**
Authors		**N. Merat, E. Johansson, J. Engström, E. Chin; F. Nathan; and T. Victor**	
Status (F: final; D: draft; RD: revised draft):		**RD**	
File Name:		D:\nmerat\AIDE\Tasks2.2.3\2.2.3 reports\AIDE_Wp2.2_task223_V5.doc	
Project start date and duration		**01 March 2004, 48 Months**	

Table of contents

List of Figures

List of Tables

List of abbreviations

Term	Definition	References	Alternative definition	References alt. definition	Notes
ADT	Auditory detection task – an auditory version of the PDT where visual stimuli are replaced by auditory stimuli such as ‘beeps’ or burst of white noise	A new task implemented for AIDE			
DT	Detection task –Different from the *peripheral* detection task in that stimuli are no longer presented in the visual periphery				
IVIS	In-vehicle information system – an in-car system which provides warning/information to drivers which may or may not be driving-related. Examples include mobile telephones and navigation systems.				
PDT	Peripheral Detection Task -	Maartens & van Winsum (1999)			
TDT	Tactile Detection Task – a tactile version of the PDT, where visual stimuli are replaced by vibrotactile stimulators.	Enstrom et al. (2005)			
VDT	Visual Detection Task – similar to the PDT but visual stimuli are not necessarily presented in the peripheral field of view and can appear centrally or in the visual scene of a driving simulator.				

Executive Summary

This report presents the results of three experiments conducted at Volvo Technology, PSA and Leeds University, which were designed to examine the suitability of a series of detection tasks for the safety assessment of IVIS and ADAS. The basis for this work is the Peripheral Detection Task method. However, the visual stimuli used traditionally in this task were altered in some experiments, prompting the use of the term Detection task for all experiments.

The Volvo Technology experiment investigated the effect of visual stimulus eccentricity on detection, as well as assessing the effect of modality by using a tactile stimulus. The PSA and Leeds experiments also studied the effect of presentation modality on detection by using tactile, visual and auditory stimuli. The effect of IVIS demand on detection of these stimuli was measured by employing a variety of IVIS and surrogate IVIS, which differed in terms of their requirements for different processing resources.

Results from the Volvo experiment showed that neither visual stimulus eccentricity nor changes in stimulus modality had a major impact on the evaluation of results. All visual stimulus positions and the tactile detection task were sensitive to differences between baseline and the different IVIS tasks. Rather, performance was more influenced by other factors, such as the effect of sunlight on the visibility of the different visual stimulus positions.

The PSA experiments found an overall deterioration of performance in both detection tasks by the IVIS, although considerable differences in performance were observed between the tactile and auditory detection tasks. While some of these results were thought to be due to the difference in information processing resource used for these two tasks, some caveats were also highlighted, including fundamental differences in the physical characteristic of the two stimuli.

The Leeds experiments showed clear differences in performance between the visual, auditory and tactile detection tasks, and results concurred with multi-component models of human information processing which suggest that dual or multi-task performance is best achieved when there is least conflict between response and processing resources.

Based on earlier findings, described in the AIDE State of Art report 2.2.1 (Johansson, et al, 2005) and the results from the current three experiments, it is recommended that the tactile detection task can be used for assessing the safety of an IVIS whilst driving. The recommendation of this task is partly based on the fact that, currently, IVIS and ADAS rely least on the tactile/haptic modality. It can be assumed therefore that this is the least ‘overloaded’ modality and that any impairment in performance in the detection task will be due to a competition for general (rather than modality specific) resources. Following further assessment by the AIDE task 2.2.7, the intention is then to use this detection task for the evaluation of prototypes built in SP3.

1 Introduction

This report summarises the results of three experiments, conducted at Volvo Technology, PSA and The University of Leeds, all of which examined drivers' reaction time to a variety of detection tasks (DT) during interaction with a series of in-vehicle tasks and driving. Based on earlier findings described in the State of the Art report 2.2.1 in AIDE (Johansson, et al, 2005) and the results from the current three experiments, a specification for a detection task is provided at the end of this report. This recommendation will be used by partners in task 2.2.7 "Empirical comparison of methods for off-line workload measurement". The intention is then to use this specification for the evaluation of the prototypes built in SP3.

The basis for this work is the Peripheral Detection Task method (e.g. Martens & Van Winsum, 1999). However, the traditional layout used for visual stimuli in PDT experiments has been altered in the experiments described here. Therefore, the general term describing the tasks in the experiments will be the Detection Task (DT). In order to maximise resources, a variety of IVIS and detection tasks were distributed across the three partners, following consideration of task priorities and each partners' resources and capabilities.

The Volvo Technology experiment investigated the effect of different horizontal positions for the visual stimuli on performance, as well as comparing this with performance on a tactile detection task. The PSA and Leeds experiments looked at the effect of presenting the signals for the detection task in various modalities, by presenting visual, tactile and auditory stimuli.

Following a description of experiments from each site and a discussion of the results, comparative analysis has been made for all results and a specification on how the detection task should be used in future work is provided.

2 Input from previous AIDE work

A review of existing techniques and metrics for IVIS and ADAS assessment was carried out in AIDE, task 2.2.1 (Johansson et al., 2005). One chapter in this review focused specifically on secondary task methods and the experiments in this task are partly based on what was discussed in the earlier review.

3 Contribution to overall AIDE objectives

The results from the work in this task will be given to task 2.2.7 where the Detection Task data will be compared to the other metrics and tools developed within Work package 2.2. The results from the work in 2.2.7 will then be incorporated into the generic test regime developed in Work package 2.1 and will be used to assess the safety implications of AIDE demonstrators as well future ADAS and IVIS.

4 Common theoretical framework

The main aim of the experiments described in this deliverable was to determine how drivers' interaction with in-vehicle systems such as IVIS and ADAS influences their detection of other (perhaps unexpected) events in the driving scene. The following sections provide a brief description of the theoretical framework used to design these experiments.

4.1 Tunnel vision or general interference in the visual modality?

In previous studies, the PDT has been developed to test detection performance in the visual periphery; and an underlying assumption attributed to this test is that high demands create tunnel vision. As one aim of the experiments in this deliverable was to test this assumption, a more detailed analysis of tunnel vision in the visual modality is provided here.

The eyes are not cameras that deliver a uniformly detailed picture image. Vision becomes increasingly limited towards the periphery because of a decreasing density of receptors in the retina, but also because central visual regions receive an increasingly higher proportion of cortical processing in higher cortical regions (Findlay & Gilchrist, 2003). The fovea is not simply an area of high acuity, but rather the location at which visual processing is centred. This sensory limitation is easily mistaken as a central limitation.

Many visual functions show a progressive and dramatic reduction in performance as stimuli are placed more eccentrically from the fovea, decreasing until certain discriminations become impossible outside a certain central contour (Findlay and Gilchrist, 2003). Performance in the visual field is mainly dependent on stimulus features such as size and colour, contrast, movement, and luminance, but is also profoundly affected by lateral masking from the presence of background stimuli (Bouma, 1978). Detection performance of a stimulus will thus depend on *stimulus saliency*.

The visual field is “the sum of all directions from which the eye may perceive visual stimulation at a defined moment in time and the performance of the perception of this stimulation” (ISO, 1999). Perimetry is the systematic measurement of visual field function. A series of contours, or isopters, around the fixation point represent probabilities of target detection (e.g. Bellamy, 1984). Visual field *size* is the contour of the furthest visual angle from the fovea in which a particular stimulus can be detected, within a certain probability, without moving the eye or head (see Sanders and Donk, 1996). In addition to simply defining the outer contour of a visual field, detection probability rates and reaction times (RT) can be used to map out *sensitivity* to stimuli within that visual field.

Various similar concepts - useful field of view (UFOV), functional field of view, and visual lobe - define an area in which some additional *qualitative* information in the stimulus can be extracted (e.g. Rantanen and Goldberg, 1999). Ball (1988) defines UFOV as “the total visual field area in which useful information can be acquired without eye or head movements”. In addition to simply being detected, as in the visual field definition, a *useful* stimulus must be recognized, categorized, or identified to be part of the UFOV. The UFOV is thus simply equivalent to a visual field size for a specific qualitative aspect of a stimulus. Instead of using separate terms for ‘simple detection’ and ‘detection of qualitative information’, the term *visual field performance* can be used to refer to both simple detection and qualitative detections which both affect visual field size and sensitivity. Visual field performance (size and sensitivity) is intimately tied to the specific stimulus feature being studied.

4.1.1 Tunnel vision

Unfortunately, there is a good deal of misunderstanding in the literature with regard to the phenomenon called tunnel vision. Generally speaking, a wide variety of results show deterioration in visual field *size* and/or visual field *sensitivity* across retinal eccentricities with increasing processing demand. Visual field size reduction is not always reported because the stimuli used are not always tested at the edges of the visual field, eccentricities greater than 30 degrees are fairly uncommon. However, if the eccentrically declining sensitivity curves are extrapolated, then outer limits for detection can be estimated. In all cases, the ‘high demand’ results are poorer than baselines and therefore would result in smaller visual field sizes if they were extrapolated to visual field outer limits. Thus, visual field size reduction is a direct consequence of sensitivity reduction. Although this characteristic is not always evident in various interpretations, there is consensus in all results indicating visual field size reduction from increased demand.

Visual field *size* reduction is sometimes interpreted as *tunnel vision* simply because higher demands give smaller visual field sizes. But, some researchers reserve the term tunnel vision for a “funnelling” of sensitivity deterioration, as we shall see. Thus, one misunderstanding regarding tunnel vision is that visual field size only shrinks because increased demand produces ‘increasingly stronger peripheral deterioration’. For example, Raantanen and Goldberg’s (1999) or Jannelle’s (1999) findings of size shrinkage can depend either on general interference or an increasingly stronger peripheral deterioration.

Visual field sensitivity effects are disputed, partly because of confusion in terms. The basic disagreement is on whether sensitivity deteriorates more strongly in peripheral regions or not. It can be conceptualized as whether the curve describing sensitivity over retinal eccentricities is lowered or becomes narrower, see Figure 1. Although the term ‘funnelling’ (e.g. used by Sanders and Donk, 1996) may be more appropriate to describe increasingly stronger peripheral deterioration with increased processing demands, the term ‘tunnel vision’, or *tunnel vision hypothesis* has become the established term, see Figure 1B. This, of course, adds to the confusion, leading some researchers to treat any results showing reductions in visual field *size* to be supportive of the tunnel vision position (increasingly stronger peripheral deterioration), which simply is not a correct deduction. Size can equally well be reduced by a general interference, as shown in Figure 1A. The *general interference hypothesis* states that sensitivity is equally affected throughout the entire visual field from (the bell curve is lowered as in Figure 1A). Note that hybrid hypotheses are possible, in which sensitivity is affected disproportionately in different eccentricities.

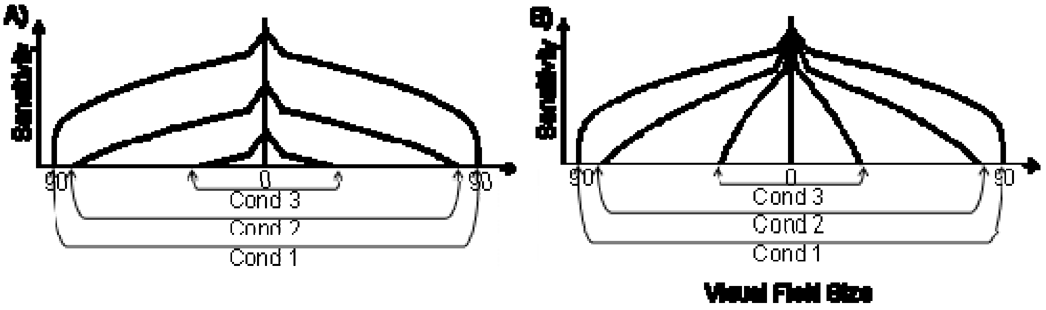


Figure 1 – Binocular visual field size and sensitivity deterioration as a function of experimental conditions (conceptual).

Figure 1 shows how visual fields are affected by factors such as stimulus size, contrast, or mental workload (conditions 1 to 3) such that both field size (visual angles from fovea on the x-axis) and the sensitivity profile within each field are affected. A) illustrates the general interference hypothesis stating that the field is equally affected across the retina for each condition, and B) illustrates the tunnel vision hypothesis stating that sensitivity is more affected in the periphery (from Victor and Åberg, 2005).

The general interference hypothesis has strong, clear support (Ball et al, 1993; Crundall et al 1999, 2002; Holmes et al, 1977; Sanders and Donk, 1996; Recarte and Nunes, 2003; van de Weigert, 1993). The general finding is a clear reduction in sensitivity and reaction time performance across the visual field. Indeed, Ball et al (1993), Crundall et al (1999; 2002), Holmes et al (1977), Recarte and Nunes (2003), and van de Weigert (1993) were unable to find support for eccentricity and workload interaction (tunnel vision) in specific statistical tests. Although foveal targets are generally not used as stimuli, one exception is Holmes et al (1977) who do include foveal targets at 1° and find support for general interference.

Some of the results commonly taken to support the tunnel vision hypothesis actually only show size reduction (e.g. Chan and Courtney, 1993; Ikeda and Takeuchi, 1975; Jannelle, 1999), or claim support for the funnelling position without performing statistical tests of eccentricity interaction (e.g. Plainis et al., in press).

Once this misinterpretation of field *size* reduction as being directly supportive of the tunnel vision hypothesis is removed from conclusions in the literature, it seems that little support for the tunnel vision position (introduced by Mackworth, 1965) is found in some, but not all, of William's work (1982, 1985, 1988, 1995). Notably, Williams only used a small sample of eccentricities (three visual angles), in a small region (between 1.5 and 9°), under speed stress, with specific instructions to prioritize the foveal task (Williams, 1988). At best, evidence in support of the tunnel vision hypothesis is weak and is produced under specific and stringent conditions. Moreover, it should be noted that Williams's studies were limited to the experimental situation with a primary foveal task and a secondary detection task. It could be argued that this situation is somewhat analogous to driving a vehicle while performing the peripheral detection task. However, this is very different from an IVIS evaluation experiment, where the detection task is used as a *tertiary* task, with the main purpose being the measurement of workload induced by a *secondary* (IVIS) task. To the knowledge of the present authors, there is no empirical evidence of tunnel vision occurring as a result of secondary task load.

Based on the above arguments, it is feasible to assume that an increase in secondary task demand, (e.g. as a result of an in-vehicle task) will lead to equally reduced detection of visual stimuli across the visual field. As described later, exactly this hypothesis was tested by Volvo, by comparing detection of LEDs which were presented at three visual eccentricities. If the general interference hypothesis is correct, the *degradation* in detection performance during a secondary task should be equal across the visual field, i.e. there should be no interaction between secondary task load and eccentricity (however, according to both the tunnel vision and general interference hypotheses, the absolute detection task performance is expected to be reduced as a function of eccentricity, as illustrated in Figure 1).

4.1.2 Gaze concentration and 'tunnel vision'

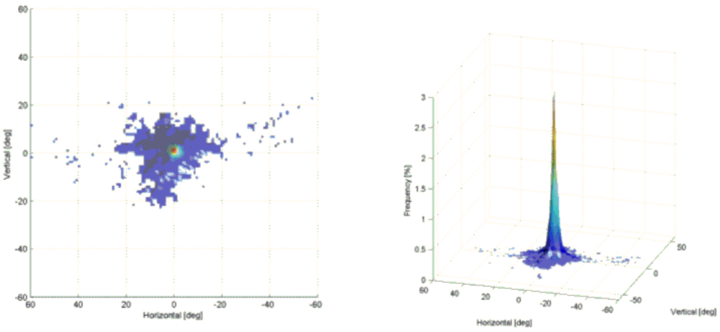
Another motivation for the experiments described in this report comes partly from the results of the European FP5 project HASTE, which found that the effect of a *visually demanding* IVIS on driving behaviour is very different to that of a *non-visual* (sound-based) IVIS, which nevertheless imposed a high demand on cognitive resources[1] (see Östlund et al., 2004; Jamson & Merat, 2005 and Engström, Johansson and Östlund, 2005). In particular, these two IVIS had very different effects on the distribution of drivers' eye-movements (Victor, Harbluk and Engström, 2005). Therefore, one conclusion from these HASTE studies was

---

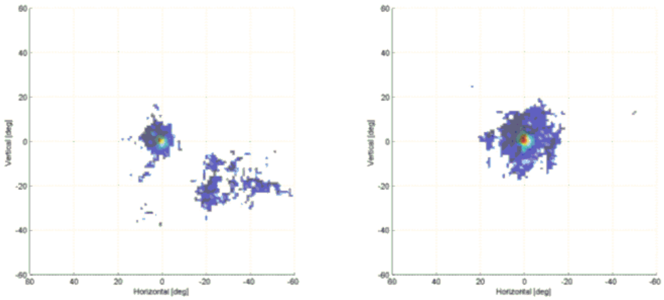
[1] This task was an auditory version of the visual continuous memory task, and named the auditory continuous memory task or aCMT.

that when performing a highly demanding non-visual IVIS, gaze concentration towards the road centre, and an associated reduction in visual field performance clearly have detrimental effects on safety. Therefore, it was important to establish whether this reduction in glances towards peripheral visual events would be accompanied by a reduction in detection of visual events, presented in the vehicle or driving scene. Before describing the experiments further however, it is perhaps important to provide a few words on the effect of task difficulty on glance behaviour and the difference between gaze concentration and tunnel vision.

Gaze concentration refers to the reduced spread of fixations over time, whereas tunnel vision refers to reduced detection performance across the visual field during a fixation (independent of eye movements). Gaze is shown to spatially concentrate to the road ahead as a function of increased demands on the driver. Gaze concentration has been shown to occur as an effect of curve negotiation (Land & Lee, 1994), cognitive load (Harbluk et al., 2002; Nunes & Recarte, 2002), the visual load imposed by in-vehicle glances (Victor & Johansson, 2005), in dangerous driving situations (Crundall & Underwood, 1998), and as a result of internal factors such as alcohol (Belt, 1969), anxiety (Janelle et al., 1999) and fatigue (Kaluger & Smith, 1970). Thus, as demands increase, the need to monitor the road ahead with the highest acuity is greater and fixations to this region are increased, see Figure 2.



a) Visual task baseline Mwy Field b) Auditory task baseline Mwy Field



c) Visual task SLv3 Mwy Field d) Auditory task SLv3 Mwy Field

Key:
SLv3 - high load IVIS task
Mwy - Motorway

Figure 2 – Example of fixation density plots of eye movements in different condition as seen from the drivers' perspective (Figure from Victor et al. 2005)

One direct consequence of spatial gaze concentration to the road centre region is that there is less acuity on peripheral sections of the visual scene over time, since the high acuity hovers around one place only. Increased spatial gaze concentration results in even less acuity than normal being spread out. Gaze concentration and reduced sensitivity to stimuli in the visual field are not identical, but they do seem to occur simultaneously (e.g. Recarte and Nunes, 2003) and are frequently confused. Both gaze concentration and reduced visual field performance occur simultaneously as an effect of higher processing demands, such as those induced by cognitive tasks, driving environment demands, anxiety, and vigilance.

4.2 Attention

Is the reaction time deterioration associated with increased demands in various driving and secondary task situations tied to the visual modality, thus reflecting visual detection performance or can the same reaction time deteriorations be found in other modalities (haptic and auditory)?

4.2.1 Modern attention theory

A commonly agreeable basic statement describing attention is “Key to the survival of many biological organisms is their ability to selectively focus neural processing resources onto the most relevant subsets of all available sensory inputs.” (Tsotsos, Itti, & Rees, 2005, p. xxi). Attention can thus be seen as a selection process where some inputs are processed faster, better, or deeper than others, so that they have a better chance of producing or influencing a behavioural response, for example by speeding up reaction time (Lamme, 2005).

Corbetta and Shulman (2002), and Koch (2004) show quite conclusively that visual attention is controlled by two functionally and neurologically distinct neural systems. At one extreme, attention can be dominated by external events, being drawn to salient objects or distracting events (stimulus-driven). On the other extreme, attention is controlled by cognitive factors such as knowledge, expectation, and current goals (goal-directed).

Stimulus-driven, bottom-up attention is driven by intrinsic stimulus qualities, acts rapidly, acts automatically, mediates pop-out, and acts across the entire visual field (Corbetta & Shulman, 2002). Stimulus-driven attention is specialized for the detection of behaviourally relevant stimuli and acts as a ‘circuit breaker’ for goal-directed attention, directing attention to salient events (Corbetta & Shulman, 2002). Selection is believed to be controlled via one or more explicit saliency maps, wherein neurons encode saliency or conspicuity of objects in the visual environment, not stimulus attributes (Corbetta & Shulman, 2002; Itti, 2005; Itti & Koch, 2000; Koch & Ullman, 1985). Competition among neurons in a saliency map gives rise to a single winning location that corresponds to the next attended target.

Goal-directed, top-down attention is involved in the cognitive selection of sensory information and responses, takes longer to deploy and can be directed to either a proscribed region in space, to individual objects, or to specific attributes throughout the visual field (Corbetta & Shulman, 2002; Koch, 2004). Determining which stimulus in a visual display is the most important to an observer requires the integration of both top-down and bottom-up processes. Thus, stimulus-driven aspects interact with ongoing cognitive goals (Corbetta & Shulman, 2002). Attentional selection is performed based on top-down weighting of the bottom-up feature maps that are relevant to a target of interest (Itti, 2005). This dynamic interaction is central to current theories of attention such as biased competition accounts (Desimone & Duncan, 1995; Itti, 2005; Reynolds & Desimone, 1999) and similar elaborations (Bundesen, Habekost, & Kyllingsbæk, 2005; Logan, 2002). The dynamic interaction explains how certain stimuli are enhanced, while other unattended stimuli are suppressed (see Koch, 2004; McAdams & Maunsell, 1999). In short, attention is the biased sum of bottom-up and top-down processes when it serves visual identification and selection of objects.

4.2.2 Implications on stimulus detection in the current experimental paradigm

The effect of increased demands upon both gaze concentration and reduced detection performance can be explained within the framework of this theory on attention. Increased attention to a subset of visual, auditory, or tactile stimuli facilitates the stimulus that is attended to, but it also simultaneously inhibits other stimuli. In the visual modality, the inhibition of unattended stimuli means poorer stimulation from stimulus-driven attention, leading to less stimulus-driven eye-movements. Likewise, the engagement of goal-directed attention to in-vehicle tasks while performing the dual task of driving reduces the capability to attend to other goal-directed attention stimuli. Thus, the interplay of stimulus-driven and goal-directed attention in biased competition for attention can explain both eye-movement and detection performance.

Stimulus-driven and goal-directed attention in biased competition is generally similar in other modalities (Spence and McDonald; 2004). However Spence and McDonald (2004, p 21) state: "while attention may well be controlled by a supramodal attentional mechanism, it will nevertheless exhibit some modality-specific features, given the differences in the spatiotemporal processing of stimuli in each of our senses." This difference in spatiotemporal processing of different modality stimuli might also suggest that ignorance or inhibition of some stimulus modalities is easier (or more difficult) than others, a factor which may also be ruled by the saliency of that particular stimulus, or even its importance/relevance to the subject.

4.3 Dual task studies

These basic observations on visual perception are taken a step further in driving, where drivers' attention is primarily directed towards the driving scene, but is also required by other events in and out of the car. Several studies (including those done in the HASTE project) have shown the effect of distraction on driving by using dual or secondary task methodologies. Conventionally, these experiments study the effect of distraction from a 'secondary task' (e.g. an IVIS), on performance in the 'primary task' (e.g. driving), as well as providing some insight into the limitations of human operators. Such dual task investigations have 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 two subsystems dealing with auditory/verbal and visual/spatial material respectively and a 'Central Executive', which acts very much like a supervisory attentional system (see Norman & Shallice, 1986), controlling and manipulating the two subsystems.

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.

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), a sound-based secondary task is shown to produce less disruption than a visual secondary task (see also Parkes & Coleman, 1990).

However, 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.

Interference can however be much more specific than would be predicted on the basis of a limited number of broad resources, but on the other hand there are cases of unspecific interference that seem not to depend on any specific resources (Heuer, 1996). For example, there is more interference between hand-hand than hand-foot responses, tracking-tracking or keypressing-keypressing than tracking-keypressing tasks, and listening to two words or tones is more interfering than listening to a word and a tone (Heuer, 1996). Therefore, in some cases, stimuli from the same sensory modality interfere more strongly with each other than stimuli from different modalities.

Interactions between multiple processes can be seen as impeding or facilitating performance depending on whether they are between different or similar processes (Heuer, 1996). Some abilities are enhanced and some are inhibited by ongoing processes. A term that is used fairly often to explain interactions is ‘crosstalk’, which refers to the spread of some aspects of the signals related to performance of the one task into the signals or data related to performance of the other task.

To summarise, 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:

i. The priority assigned to each task by the experimenter.
ii. The priority assigned to each task by participants (this may be regardless of experimenter instructions).
iii. Participants' idea of perfect performance.
iv. Participants' idea of acceptable performance.

In order to examine the safety effects of an IVIS, the design of the experiments for this AIDE task utilised the results of research on attention and multi-component models of information processing. In particular, it was thought that these frameworks would provide a suitable background for testing the limitations of our drivers in performing a variety of IVIS and detection tasks during driving.

4.4 Assessing the safety of an IVIS

In order to test the safety implications of using an IVIS or ADAS in driving, an important consideration is clearly to assess their effect on driving performance. Using two ‘surrogate’ IVIS tasks, results from the HASTE experiments showed a number of potentially unsafe driving-related manoeuvres by a visually distracting IVIS, including an increase in lateral deviation and a rise in the number of steering reversals. The effect of a sound-based cognitively demanding task on driving performance was more varied, but a common finding was the marked reduction in lateral deviation, which, as mentioned above, was accompanied by gaze concentration towards the road centre, and an associated reduction in visual field performance. Clearly, if performance in the driving task alone was taken into account, it can be argued that by reducing a certain amount of ‘weaving’ in the road, the demanding sound-based task effectively *improved* driving performance. However, this conclusion is problematic since the pattern of eye-movements clearly suggests that there should be an associated reduction in visual field performance and drivers may actually fail to see an object in their peripheral view during performance of a demanding sound-based task. Therefore, in this case, performance on the driving task alone would not provide adequate information on the distracting effects of an IVIS.

Consequently, one aim of the experiments described here was to examine how engagement in an IVIS affects drivers’ need to react to sudden unexpected events in the car or driving scene. Becoming aware of such events in the driving scene is a task which, in the first instance, almost exclusively requires visual detection, and the ability to direct the eyes. The simplest and most common technique used to assess drivers’ ability to react to such events in the laboratory is the peripheral detection task (PDT) method. This involves the random presentation of simple visual stimuli at 11-23^{o} to the left of the drivers’ visual line of sight (e.g. Martens & van Winsum, 1999). Results have shown a reduction in the number of hit rates and an increase in reaction time to the PDT by both visually and cognitively demanding secondary tasks (e.g. Verwey, 1993; Olsson, 2000). However, as outlined by Harms and Patten (2003), the PDT hit rate was lower during performance of the *visual* navigation system, compared to the *verbal* navigation system; leading these authors to argue that this may have been due a competition for visual resources by driving; the visual navigation system; and the visual PDT. This conclusion is certainly supported by the limited capacity models described above. Nevertheless, other studies have found very similar effects on visual and tactile detection performance, for both visual and auditory/cognitive secondary tasks (Engstrom et al. 2005), a result which is inconsistent with multiple resource theory, since for instance this theory would predict smaller effects of the visual secondary task on the tactile detection task (due to less inter modal interference).

To investigate this concept further, the experiments designed for task 2.2.3 of AIDE employed a variety of IVIS and ‘surrogate’[2] IVIS tasks, which varied in terms of their imposed input and response mode. The rationale for using such a variety of tasks was to examine the effect of conflicts in modality on performance in the visual detection task. In other words, to re-examine Harms and Patten’s suggestion that a visual peripheral detection task is disrupted more by a visual than a verbal IVIS. Also, to obtain a more comprehensive design, we incorporated the use of a variety of detection tasks, which varied both in terms of stimulus position and modality. For instance, as well as presenting the visual stimuli in the periphery, partners at Volvo also examined the effect of visual stimulus position on performance by presenting LEDs at different eccentricities (see Figure 4). This was done to investigate whether the detection of peripheral stimuli is affected by eccentricity (the ‘tunnel vision’ theory, e.g. Mackworth, 1965) or whether detection is impaired by workload regardless of eccentricity (the ‘general interference’ model; e.g. Acosta & Dickman, 1984; see also Plainis et al. 2001).

Finally, it was assumed that if Harms and Patten’s conclusions are correct, a visual IVIS would be less disruptive to an auditory or tactile detection task than to a visual detection task. Table 1 shows the selection of tasks used across the three sites for this purpose. Also outlined in this table is the processing stage, modality and response code required for each IVIS, as well as the predicted source of conflict with each detection task. So, for instance, if performance is ruled by Wickens’ MRT model (Wickens, 1984; 1992), the Backward Count task would be most disruptive to the auditory detection task (compared to the visual or tactile versions), because performance of both requires the same auditory/verbal modality, so that drivers would not hear the auditory beeps because whilst counting aloud.

Table 1 - The list of IVIS and surrogate IVIS tasks used in AIDE task 2.2.3

Task	Site	Processing demand	IVIS Modality	Response code	Source of conflict with DT
Answer Questions	Volvo	Medium	Auditory	Verbal	Auditory (modality)
Backward Count	Volvo, PSA, Leeds	High	Auditory	Verbal	Auditory (modality)
Read email	Volvo	Medium	Visual	Verbal	Visual (modality)
Dial ‘phone numbers	Volvo, Leeds	Medium	Visual	Manual	Visual (modality) Auditory (modality) All (response code)
IVIS (with/without) errors of recognition	PSA	Medium/High	Auditory	Verbal	Auditory (modality)

---

[2] It was assumed that the surrogate IVIS would be similar to a real IVIS in terms of the demand it required from the driver. For instance, answering biographic questions would be similar to a simple telephone conversation.

5 Experiment 1 (VTEC)

The VTEC experiment was conducted as a conjoint experiment between AIDE and the national project SAFE-TE funded by the Swedish Road Administration. The result from this experiment is presented here as well as in a separate report within SAFE-TE.

Stimulus presentation (in terms of position and modality) is probably one of the most important factors influencing detection performance. The objective of the present study was to investigate empirically a number of modifications to the “standard” PDT with respect to how the stimuli are presented. Specifically, the following modifications were investigated:

i. Use a single LED rather than the PDT array

It is unclear whether the spatial randomisation of the visual stimuli really is necessary. The area of 11-23 degrees horizontal and 2-4 degrees vertical angle specified by van Winsum et al. (1999), and employed in most PDT studies, is still quite small, so the driver is aware of the direction of the stimuli. If the current LED array could be replaced by a single stimulus it would simplify the technical set-up substantially.

ii. Place the LEDs in a central position rather than in the periphery

The main idea behind this modification was to increase the visibility of the stimuli thus reducing the variance induced by different lighting conditions. As reviewed above, this was motivated by the general interference hypothesis (e.g. Recarte and Nunes, 2003), which states that the sensitivity of detection task to secondary task load is independent of stimulus eccentricity.

iii. Increase the intensity of the LED (compared to the “standard” PDT LEDs)

This was done in order to reduce sensitivity to lighting conditions.

iv. Replace the LEDs by a tactile stimulus

As shown by Chin et al. (2004) and Engström et al. (2005), detection task sensitivity to secondary task load appears to be independent of visual stimulus eccentricity, and stimulus modality. Thus, the tactile detection task employed in Engström et al. (2005) was included in this study in order to validate this proposal. One main advantage of the tactile detection task (TDT) is that it is not sensitive to lighting conditions and does not affect eye movement measures.

The modifications i-iii lead to an alternative detection where the visual stimulus is not necessarily placed in the periphery. Thus, the more general term visual detection task (VDT) will henceforth be used to refer to this method. In the present study, three different horizontal stimulus positions were implemented for the VDT. These positions were compared to each other and to the TDT with respect to their sensitivity to four different visual and/or cognitive secondary tasks.

5.1 Method

5.1.1 Participants

30 participants, 10 women and 20 men, were recruited for the experiment. The average age was 28 years old (range: 23-53). The participants had had their driving licence for an average of 9 years (range: 1-35). 14 of the participants estimated their annual mileage to be less than 1000 km, while the remaining 16 estimated it to be in the range of 1000-4000 km.

5.1.2 Apparatus

A Volvo V70 with automatic gearbox was used for data collection. For both the PDT and the TDT, the Signal Detection Task (SDT) Tool was used to generate the stimuli and collect responses. The SDT Tool, illustrated in Figure 3, is a generic tool for signal detection task studies, developed in-house at Volvo Technology which features visual, auditory as well as tactile stimulus generators. The basic analysis of the raw data was done in Matlab and then exported to SPSS for statistical analysis.

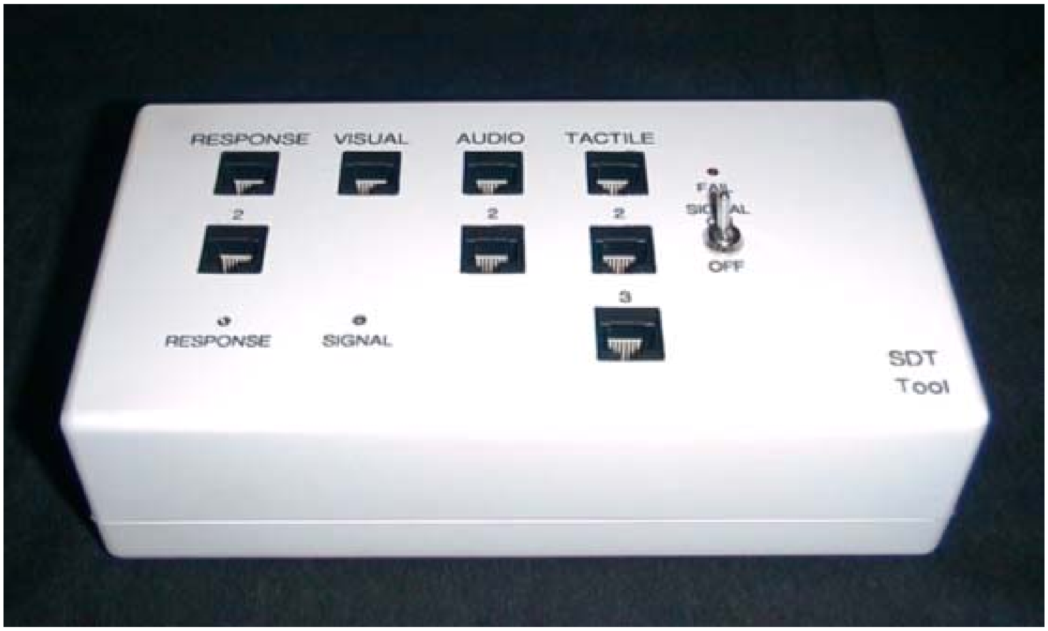


Figure 3 - The Signal Detection Task Tool

The subjects performed both the VDT and the TDT. The two tasks differed only with respect to the stimuli. For the VDT, the stimuli were given by three LEDs with a stronger luminance than the "standard" LEDs used in previous studies. The detailed specification for the LEDs is given in Table 2.

Table 2 - Specifications for the LED

Colour	Wavelength (nm)	Luminance at 20 mA	Beam angle
Red	660	2000	+- 12 deg.

The LEDs were positioned so that they were reflected either in the windshield or in the left window. The LED itself was masked and only the reflections could be perceived by the subject. Three different LED positions were included, as illustrated in Figure 4.

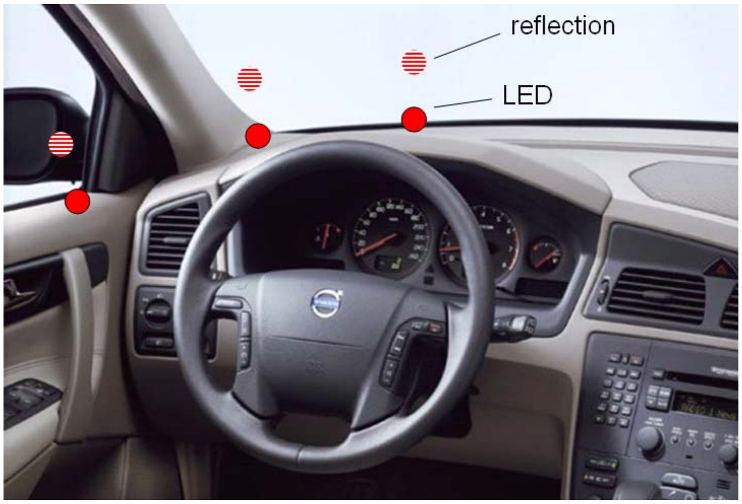


Figure 4 - Position of the LEDs in the Volvo study

The TDT was implemented in the same way as in Engström et al. (2005). The stimuli were given by two small electrical vibrators attached to the wrist of each hand (Figure 5). The stimulus activation was varied randomly between the two vibrators.

For both detection tasks, responses were given by means of a response button attached to the index finger (Figure 5). The stimulus duration was set to 1 second for both the VDT and TDT and the stimuli were presented at a rate of 3-5 seconds

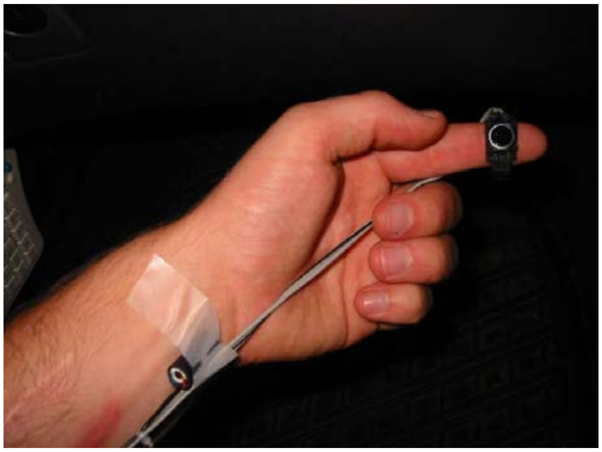


Figure 5 - Tactile vibrator and response button

5.1.3 Design and procedure

All data were collected on a motorway outside Gothenburg with divided lanes (two in each direction). The route was approximately 75 kilometres in total, and had two lanes in each direction. The mean width for each lane was 3.8 meters. The speed limit of the road was 110 km/h, although the weather and traffic conditions meant that it was difficult for drivers to adhere to this speed at all times.

A within-group design was employed with Secondary Task and Stimulus as independent factors. Secondary Task had five levels (Baseline, Question, Backward Count, Phone and Email). The levels for Stimulus were the three visual positions plus the tactile stimulus (Visual-leftmost, Visual-left, Visual-centre and Tactile). Every subject was exposed to each condition three times. The order of the Secondary tasks and Stimulus conditions were counterbalanced.

The dependent variables for the detection tasks were hit rate and response time. Hits were defined as stimuli responded to within 2000 ms from stimulus onset, with the exception of unrealistically fast responses (“cheats”, < 200 ms.). Hit rate was then defined as the number of hits divided by the total number of stimuli during a task or baseline condition. Response time was defined as the average response time of the valid responses during a secondary task or a baseline section. The rationale for the 2000 ms threshold is that responses slower than this should not be considered as responses to the stimulus, but rather as “arbitrary” responses given by mistake (e.g. to a falsely perceived stimulus).

Four secondary tasks were included in the study:

1. Answer biographical questions (Question)
2. Count backwards in steps of seven, for about 30 seconds, starting from a given three digit number (Backward Count)
3. Dial a 7-9 digit long telephone number using a hands-free kit (Phone)
4. Read an email located on the centre stack. The email consisted of approximately 10 lines of text written in Helvetica pt 20 (Email)

The first three tasks were similar, but not identical, to the tasks included in the Engström et al. paper (2005). In addition, a baseline condition (no secondary task) was included.

5.2 Results

The distributions of the hit rate and response time data are illustrated in Figure 6 and Figure 7. The figures represent data for all three visual DTs and the tactile DT. For hit rate, a strong ceiling effect can be observed, where the major portion of the data represent a 100% hit rate. For response time, the data is normally distributed, albeit a bit skewed to the left, similar to the distributions that have been found in other studies (e.g. Chin et al., 2004; Engström et al., 2005).

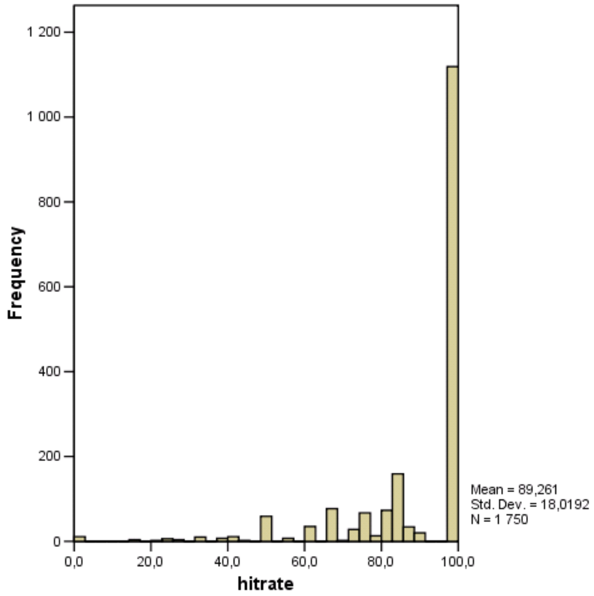


Figure 6 - Distribution of the hit rate data

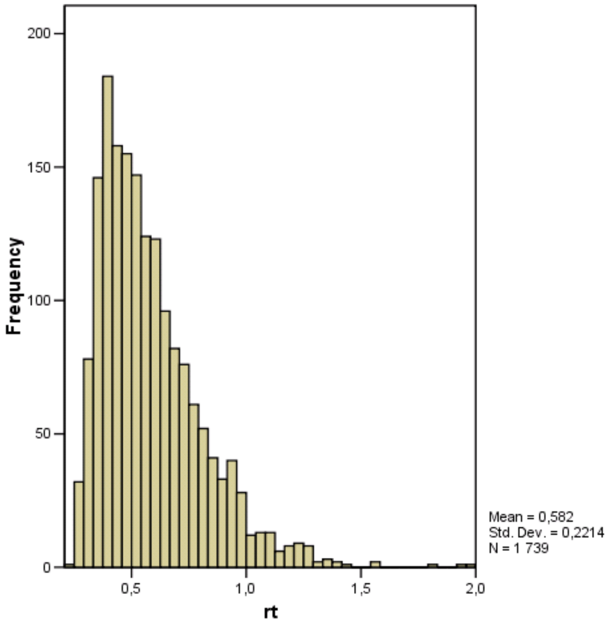


Figure 7 - Distribution of the response time data

Due to technical problems with the tactile vibrator, TDT data were missing for some subjects. In addition, a few data points were missing for both tasks due to other (unknown) reasons, e.g. erroneous annotations. In total, 51 out of 465 TDT data points (11.0%) and 59 out of 1395 (4.3%) VDT data points were missing.

Univariate ANOVAs with Secondary Task and Stimulus as main factors and Subject as a random factor were performed for hit rate and response time respectively. For hit rate, significant main effects were found for both Secondary Task (*F* (4, 123) = 11.9, *p*<0.001) and Stimulus (*F* (3, 84) = 4.48, *p*<0.01). The interaction between the two factors was not significant. The results from the hit rate analysis are illustrated in Figure 8. Sidak post hoc tests were performed to test differences between the factor levels. The results for the Secondary Task post hoc comparisons are given in Table 3. For Stimulus, the only significant differences were between the Tactile level and the others (p<0.05).

Table 3 - Results of post-hoc comparisons for Secondary Task, for hit rate

	Baseline	**Question**	**Backward Count**	**Phone**	**Email**
Baseline		**	**	**	**
Question			*	-	**
Backward Count				-	-
Phone					-
Email					

(**=significant at .001 level, *=significant at .05 level, -=not significant)

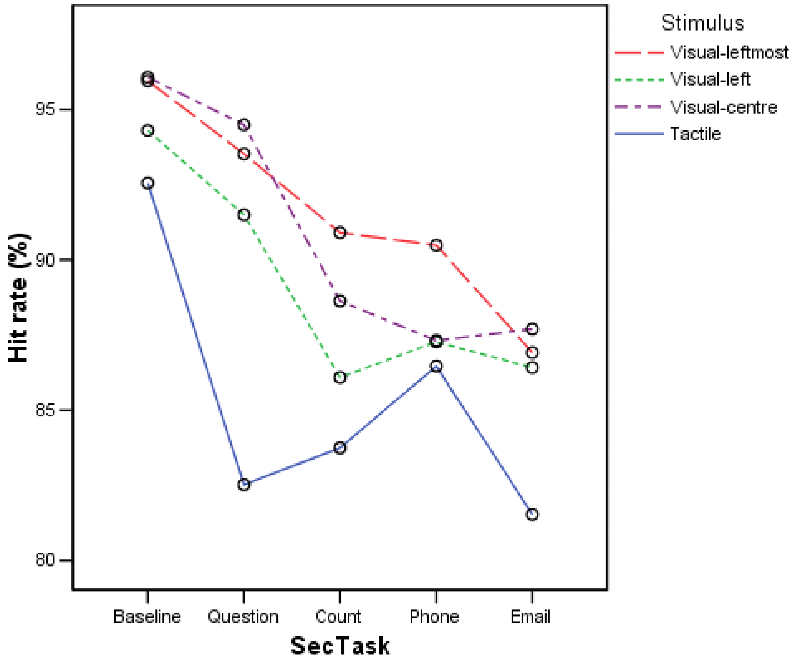


Figure 8 - Estimated marginal means for hit rate

For response time, main effects were found for both Secondary Task (*F* (4, 125) = 73.6, *p*<0.001) and Stimulus (*F* (3, 84) = 6.93, *p*<0.001). There was also a significant interaction between the two factors (*F* (12, 1485) = 2.05, p<0.05). Sidak post hoc tests were performed to test differences between the factor levels. For Secondary Task, all levels differed significantly from each other (p<0.05), except Phone and Email. For Stimulus, the Left-centre position differed significantly from the three others, but no other differences were found. The results are illustrated in Figure 9.

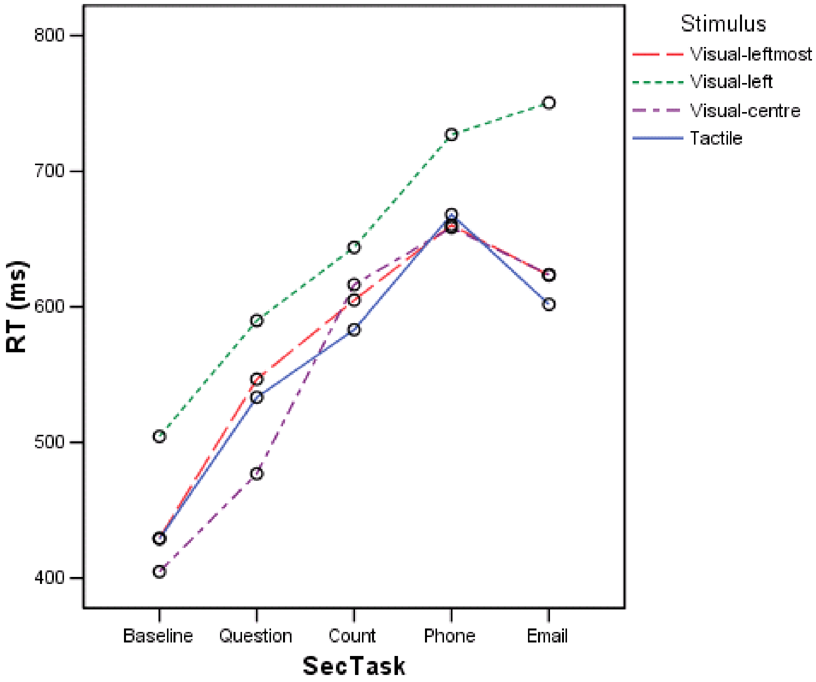


Figure 9 - Estimated marginal means for response time

In order to compare the sensitivity of the different detection task variants, some further analyses were performed. First, separate univariate ANOVAs were performed for each stimulus, and post-hoc tests were performed to test differences between Secondary Tasks. The results are given in Tables 4-7. It can be observed that the Visual-left and Tactile stimuli were slightly more sensitive to differences between tasks than Visual-left and Visual-centre.

Table 4 - Sidak post-hoc test for the Visual-leftmost stimulus

	Baseline	**Question**	**Backward Count**	**Phone**	**Email**
Baseline		**	**	**	**
Question			-	**	*
Backward Count				-	-
Phone					-
Email					

**=significant at .001 level, *=significant at .05 level, -=not significant

Table 5 - Sidak post-hoc test for the Visual-left stimulus

	Baseline	**Question**	**Backward Count**	**Phone**	**Email**
Baseline		**	**	**	**
Question			-	**	**
Backward Count				*	**
Phone					-
Email					

**=significant at .001 level, *=significant at .05 level, -=not significant

Table 6 - Sidak post-hoc test for the Visual-centre stimulus

	Baseline	**Question**	**Backward Count**	**Phone**	**Email**
Baseline		*	**	**	**
Question			**	**	**
Backward Count				-	-
Phone					-
Email					

**=significant at .001 level, *=significant at .05 level, -=not significant

Table 7 - Sidak post-hoc test for the Tactile stimulus

	Baseline	**Question**	**Backward Count**	**Phone**	**Email**
Baseline		**	**	**	**
Question			-	**	**
Backward Count				**	-
Phone					*
Email					

**=significant at .001 level, *=significant at .05 level, -=not significant

Moreover, effect sizes (i.e. the standardised differences between each task and baseline) were computed for the different stimulus positions. The results are shown in Figure 10. As can be observed in the Figure, the effect sizes are comparable, for Visual-leftmost, Visual-centre and Tactile, while Visual-left is generally slightly lower. On the other hand, the Visual-left appeared slightly more sensitive to differences between tasks, as shown by Tables 3-6 above.

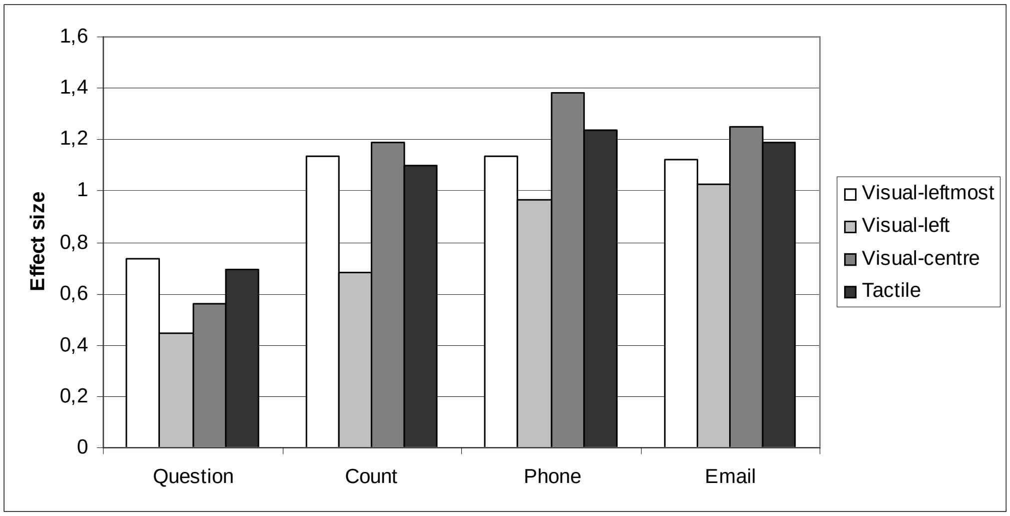


Figure 10 - Comparison of effect sizes for the different stimuli on the four secondary tasks

5.3 Discussion

The results show that hit rate is strongly sensitive to differences between the tasks and baseline, while the measure does not discriminate between the tasks. The reason for this is evident from the histogram in Figure 6, which reveals a strong ceiling effect with 100% hit rate for the main portion of the data. There is a discrete step down to a minor peak around 85%, a pattern which is clearly due to the discrete nature of the hit rate data. For a task with a duration of 30 seconds, there will be about 5-6 stimulus presentations. Thus, missing one stimulus means that the hit rate value is reduced by 17-20%. Thus, the measure is clearly too crude to distinguish between the secondary tasks. Moreover, since the data is not normally distributed, the assumptions of parametric testing are strongly violated. Based on this, it could be argued that hit rate is better used as a quality measure than a performance measure.

The use of two different performance measures (hit rate and response time) could be somewhat problematic. For example, in highly demanding situations drivers may shift strategy and respond fast and accurately only to a subset of the stimuli. This would lead to a low hit rate but unaffected response times, a pattern found for example in the study by Israelsson and Karlsson (2003). This makes it difficult to interpret and compare these data to “normal” data with unaffected hit rate and increased response times. Thus, it would be desirable to find a method which uses a single detection performance metric or considers hit rate to be a quality measure and reaction time the performance measure.

There were no differences between the LED positions with respect to hit rate. However, the hit rate for the tactile stimuli was significantly lower. It is also lower than the values reported in Engström et al. (2005). In the baseline condition for motorway driving in Engström et al., which should be directly comparable to the present baseline data, the hit rate was 99%

compared to 93% in the present study. Moreover, it is inconsistent with the response time results (further discussed below), where there were no differences between Tactile and Visual-leftmost and Visual-centre, and the tactile response times were faster than for Visual-left. Thus, the deviating result for Tactile hit rate is difficult to explain and it cannot be ruled out that this result is due to some technical error.

Response times were sensitive to differences between all task levels except between Phone and Email. Given that the data are approximately normally distributed, response time could be considered a more appropriate performance measure than hit rate.

There were no differences in response time between the Tactile, Visual-leftmost and Visual-centre stimuli. However, the middle LED (Visual-left) yielded significantly slower response times than the others, which is a rather unexpected result. A similar tendency can be observed in the hit rate data (lower hit rate for Visual-left), but in this case the difference was not significant. The deviation in response time for Visual-left is difficult to explain. The most likely explanation is that the visibility of the LED in this position was more affected by lighting conditions than the others.

Due to the deviating pattern of the Visual-left LED, it is difficult to give a definite verdict on the visual tunnelling vs. general interference issue based on the data from the visual stimuli. There was indeed a significant interaction between stimulus and task load, but this was mainly due to the deviation of the Visual-left (middle) LED (if the Visual-left data were removed, the interaction effect vanished). However, the similarity in detection performance for the visual and the tactile stimuli constitutes strong evidence in favour of the general interference hypothesis, e.g. that the reduced detection performance is due to cognitive rather than perceptual interference. This is in line with previous results (e.g. van Winsum et al., 1999; Recarte and Nunes. 2003).

The separate analyses for the Stimulus factor showed that the Visual-left position and Tactile stimuli were slightly better at discriminating between tasks than Visual-leftmost and Visual-centre. However, the Visual-left yielded slightly lower effects compared to baseline, as shown by the effect size analysis. In general, however, the effect sizes for the tactile and visual stimuli were quite similar, which contrasts to the results found in Engström et al. (2005), where the sensitivity was substantially higher for the tactile stimuli. This could be partly explained by the stronger intensity, and hence, visibility, of the LEDs used in the present study, which reduced the variance in the data. However, it should also be noted that the previous study only involved 12 subjects and, thus some of the differences found in effect size could be due to random variation. The current effect estimates, based on data from 30 subjects, should thus be more reliable.

A general conclusion from these data is that the stimulus eccentricity or even modality appears to have little influence on response time. However, the LED position is clearly of great importance for other reasons, in particular since it determines the LED visibility, which, in turn, seems to strongly affect response times (but apparently not hit rate). This seems to be the case even for the high-luminance LEDs used in the present study. Thus, it is of great importance to perform substantial pilot testing to identify a suitable LED position where the influence from lighting conditions is minimised.

It can also be concluded that both the VDT and the TDT are strongly sensitive to secondary task demands and, in general, yield similar results. However, additional criteria need to be considered when deciding which method should be included in a final recommendation. The main advantage of a visual detection task (VDT) is that it, at least to some extent, has a functional correspondence to real objects appearing in the visual scene while driving. Thus, it has a certain face validity that the TDT lacks. On the other hand, the VDT could possibly influence eye-movements, and hence visual demand measures. If this turns out to be the case, these measures cannot be collected simultaneously which greatly complicates the experimental design.

6 Experiment 2 (PSA)

The aim of the on-road experiment conducted at PSA was to provide a better definition of the secondary task method for field-test evaluation of adaptive integrated driver interface. For this purpose, particular care was taken to ensure that the perceptual modality of the secondary task and the IVIS system did not interact with each other.

In the RoadSense project, PSA conducted an on-road experiment, using a secondary detection task method, in order to assess the potential of this method for workload evaluation. The detection task performance decreased showing, a cognitive tunnelling effect, when primary task demands were high, suggesting a concentration of attentional resources on the primary task. The RoadSense experiment used auditory signals (beeps) as stimuli and a simple motor response (finger pressure) for the detection task. Driving task demands were natural (no specific instructions were given) and depended on traffic situations. While driving, drivers had to use a navigation system and a hands-free phone with a vocal input/output device (no screen, only voice command and voice messages).

Briefly, the results showed that:

- When demand increased (whether driving task demands, or IVIS task demands) detection task performance decreased (response time, number of late detections and percentage of misses increased).
- A single threshold (2s) for categorising ‘misses’ was not appropriate to all driving demands situations. This 2-second period for responding, is not equivalent in terms of time pressure depending on the situation.
- Furthermore, the binary criterion for ‘responses’ versus ‘misses’ is questionable: late responses (response time higher than 2s) has to be differentiated from omissions (no response). Therefore, at present, misses cover two different behaviours and refer to different types of cognitive processing.
- The detection task itself generates demands (mainly temporal pressure) that have to be estimated, in order to use and interpret the data correctly.

The aim of the experiments conducted in AIDE was to provide complementary results to the RoadSense findings, as well as providing new elements for the method evaluation. Therefore, the same IVIS tasks and several elements of experimental design are common in two experiments.

The objectives were to evaluate:

- Sensitivity and validity of the PDT method in two modalities: auditory and tactile signals.
- Impact on AIDE multimodal system: If two simultaneous information sources are processed (concurrent processing), which channel would be preferable in order to optimize the processing rapidity?

6.1 Method

A summary of the IVIS and detection tasks used in the PSA experiment is provided in

Table 8 - The range of tools used in the PSA experiments

IVIS	IVIS modality	ADAS	Cognitive tasks	DT Stimulus	Response	Road type
Telematic system (navigation system and hands free phone)	Vocal (output) Vocal (input)	No	Backward counting	tactile	motor	motorway
				auditory	motor	motorway

6.1.1 Participants

24 drivers (23 males and 1female) participated in the road tests (from 24 to 41 years old). 22 drivers were right-handed, 2 drivers were left-handled. All were experienced drivers, with 3 to 5 years of driving experience and travelled at least 10 000 km per year.

None of the drivers had auditory or sensory-motor impairment, and all drivers were able to drive and achieve the IVIS tasks and detection tasks.

The drivers were placed into two groups, with 50% of drivers (N=12) performing the auditory detection task and the remaining 50% (N=12) performing the tactile detection task.

6.1.2 Apparatus

The test vehicle was a Citroën C5 with automatic gearbox, equipped with a telematic multifunction system; consisting of navigation system and hands free ‘phone (see Figure 11). A list of the devices installed in the vehicle is shown in Table 9.

Figure 11 - Driver’s position (left) and experimenter’s position (right).

Table 9 - Devices installed in the PSA instrumented vehicle

Data	Devices
Video recording	4 cameras (2 for driver's eye, 1 for the scene ahead and one directed to the screens (synchronization between the different devices)
	Image mixer
	video tape recorder
Supply	Transformer 12V - 220V
	Transformer 220V – 12V
Vehicle data collection	Laptop
	Software
	Monitor for supervision
Auditory recording	2 Microphones
IVIS simulation	2 loudspeakers
	Laptop to generate wave file (voice synthesis simulated)

The detection task device consisted of:

- Main unit
- Electronic box
- 2 small loudspeakers used for the detection task auditory signals
- 1 screen and 1 keyboard
- 2 vibrators installed in the driver's seat

Figure 12 shows the response button used for this task, which was attached to the left finger of all drivers.

Figure 12 - The response button used in the PSA experiments

6.1.3 Design, procedure and tasks

6.1.3.1 Primary driving task

Participants were asked to drive as naturally as possible. In other words, no particular instructions were given with respect to the speed of travel and so on. This was done in order to avoid any additional artificial workload on the driver. During presentation of the IVIS and detection tasks, drivers were instructed to give priority to the driving task. Nevertheless, they were informed that they should try and perform all tasks as best they could (partners' common instructions).

Drivers drove on a stretch of motorway which mostly consisted of 2 (sometimes 3) lanes of traffic, although the third lane was dedicated specifically to slow vehicles travelling at a speed limit of 90 or 110km/h. This type of road was chosen to allow drivers to achieve the different tasks without too high a demand from the driving task. Traffic density and flow were expected to be low to medium considering the times selected.

6.1.3.2 Detection Tasks (DT)

The detection task signals were either auditory beeps or tactile vibration via the driver's seat. Signals were presented every 3 to 5 seconds and lasted 1 sec. Signal intensity was sufficient to be perceptible by drivers (checked with each driver). Drivers were asked to push the response button, as soon as they detected a signal. The detection task was a simple motor perceptual task, not thought to be cognitively demanding but imposed a high temporal demand considering that drivers were asked to push the response button as soon as possible.

6.1.3.3 Backward counting

This task was common to all partners, and lasted between 45 seconds and 1 minute. Subjects' task was to count aloud backwards (from 330 or 298) in steps of 7. This task is not thought to be perceptually demanding, but it is a highly cognitively demanding task involving mental arithmetic. This task was used as a 'surrogate' IVIS task, in the aim to evaluate the sensitivity of the detection task method.

6.1.3.4 IVIS tasks: dialling and destination selection

For these vocal tasks (presented without a visual display), the Wizard of Oz technique was used, although drivers were not aware of this. They were just aware that they were to use a simulation of IVIS functions with 2 voice recognition systems: one with high performance (no errors) and another one with low performance (some errors). The following IVIS tasks were chosen:

- Dialling : drivers dialled their own office phone number
- Navigation system use: the same well-known destination selection was used for all drivers

2 levels of difficulty were induced in the IVIS tasks:

- without any (system) errors of recognition: dialog was "fluent" and relatively easy
- with errors of recognition: system errors necessitated different types and levels of driver correction which varied in demand including: simple repetition, repetition after an error message, repetition after the false keyword was recognised by the system (e.g. digit correction).

a) Dialling task:

The driver pronounced (in French) their own work telephone number, to be dialled by the system. French telephone numbers are usually 10 digits long, and pronounced two digits at a time. During pronunciation of the number to the system, the driver had to pause between each 2-digit number, and listen to the result of the recognition, where the number was repeated by the system. If there was an error in recognition, the driver started the procedure of correction and had to repeat the correct 2-digit number. If the 2-digit number was correctly recognised, the driver validated it explicitly by saying 'yes' or just by waiting, and then he/she gave the next 2-digit number. When the fifth and last 2-digit number was correctly recognised, drivers validated the call explicitly by saying the keyword “call” and then had to hang up.

b) Destination selection task:

The task began with the experimenter’s instructions. The destination was imposed by the experimenter (drivers were not allowed to choose any destination) and was well-known (no memorisation was needed). The first keyword “navigation” was given by the experimenter, then the driver was prompted to give the name of the town e.g. “Paris”. If the town was correctly recognised i.e. repeated by the voice synthesis system, the driver had to validate it by saying “yes”. If not, he began the procedure of correction until the name of the town was recognised. He then gave the name of the street: e.g. “avenue of Champs Elysées”, followed by the same order of steps. When the name of the street was correctly recognised and repeated, the driver validated it by saying “yes” and the route was automatically announced. However, since the route function was not simulated, the last step of this task was to cancel the guidance function.

Each of the above tasks was performed twice in each condition, which meant that each driver performed the IVIS tasks four times in total.

The experimenter initiated the beginning of the task at appropriate sections of the road, i.e. when there were no forks in the road or any merging traffic. The navigation IVIS tasks consisted of a voice-based dialog, where the driver initiated the dialog and “answered” the system’s request. A degree of time constraint existed for this task and it was comparable to a real conversation as the driver expected a response from the system and tried to give keywords just after system response.

Before starting the IVIS tasks, drivers were familiarised with the dialogs with the car in a stationary position, following an initial demonstration by the experimenter. The task demands were as follows:

- Auditory: drivers had to listen carefully to the recognised keyword in a noisy environment.
- Mental: they had to remember and give the appropriate keyword even if they were trained with the tasks in the static condition. If there was an error of recognition by the system, the driver had to correct this and at the same time he/she remember the information for the next step of the task.
- Temporal: a time constraint existed for this task as it was comparable to a “conversation” (expectation of system response and keywords given just after system response).

6.1.4 Design and Procedure:

In total, participants took part in 8 short sessions of driving which lasted less than 10 minutes, and there was a brief break after each phase to complete workload questionnaires (the results of which are not reported here). Order of sessions was not counterbalanced and the level of difficulty increased stage by stage. In total, the driving lasted around 30 minutes, and the completion of questionnaires lasted around 40 minutes. The whole test was therefore around 3 hours, including the driver briefing, training and familiarisation phase.

The drivers started with a baseline condition of just driving, followed by performing the DT without IVIS. They then performed the IVIS tasks on their own, followed by concurrent performance of the IVIS and detection task. Drivers finished with the counting task firstly without DT and finally with DT. See Table 10 for further details of the design.

Table 10 - Order of events in the driving task

Driving	Condition	Duration	DT Auditory or tactile	IVIS no errors	IVIS with errors	Backward counting	Questionnaire
1		*8 min*					*20 min*
2	1	3 min	X				
3		*4 min*		*X*			*5 min*
4		*5 min*			*X*		*5 min*
5	2	5 min	X	X			5 min
6	3	7 min	X		X		5 min
7		*45 – 60 sec*				*X*	
8	4	45 – 60 sec	X			X	

The level of demand imposed by these tasks is thought to be as follows:

- involved cognitive process: condition 1 (baseline driving, no secondary tasks) < condition 2 (IVIS no errors correction) < condition 3 (IVIS with errors correction)
- tasks duration: condition 1 (baseline driving) < condition 2 (IVIS no errors correction) < condition 3 (IVIS with errors correction)

The counting task (condition 4) was expected to put a high demand on drivers' cognitive processes. However, the exact degree of demand was assumed to be based on drivers' counting ability and therefore checked following performance.

6.2 Results and Discussion

6.2.1 Data analysis

Table 11 shows the list of dependent variables used for each of the tasks.

Table 11 - Dependent variables

Data	Dependant variables
Detection task	Hit rate (percentage of signals with RT between 200ms and 2 seconds)
	Mean RT (for correct hits)
	Omissions rate (no response)
	Late detection rate (response time > 2 sec)
	Cheat rate (<200 ms)
IVIS task	Total task time
	Number of driver's errors
Counting task	Rapidity of counting (frequency of responses per minute)
	Accuracy of counting (% of errors)

6.2.2 Counting task performance

No specific instructions were given to drivers on how to perform the counting task. Table 12 shows performance in the counting task for the two driver groups when done in conjunction with the two detection tasks.

To estimate counting performance, two parameters were computed:

- rapidity of counting : frequency of items per minute
- accuracy of counting : percentage of errors

Table 12 - Counting performance for both driver groups (auditory DT and tactile DT)

Counting performance→	**Accuracy of counting (Percentage of errors)**		**Frequency of counts (per min)**	
	DT auditory	DT tactile	DT auditory	DT tactile
Mean	7.8	11.1	12.9	12.3
SD	16.7	18.2	5.1	4.8
Minimum	0	0	6.6	6.8
Maximum	55.6	50	23.3	19.6
median	0	0	11.5	10.1

Results showed a high degree of variability between drivers in accuracy of counting (standard deviation was higher than mean). The percentage of errors varied from 0 to 55%, so it seems that some drivers found it quite difficult to count and perform the DT while driving, although these results were similar for the two driver groups. Less variability was observed for the rapidity of counting, despite clear differences between drivers: the minimum and maximum were just over 6 to just over 20 items per minute respectively, for both detection tasks.

Considering the important variability between drivers in counting performance and in counting demands, the drivers were divided into four sub-groups as shown below:

Group A: counting was quick and accurate required low demand from drivers
Group B: counting was demanding, time was needed by the driver to be accurate, so accuracy was favoured to rapidity

Group C: counting was quick and inaccurate. Drivers favoured rapidity, although it is difficult to establish the exact level of demand.
Group D: counting was highly demanding, drivers needed time to perform the task and made many errors.

This categorisation is also summarised in Table 13. The criteria used to categorise drivers is shown in Table 14.

Table 13 - Categorisation of drivers by counting performance

Backward counting Performance		Percentage of counting errors	
		Low (accurate)	High (inaccurate)
Frequency of number given	Low (slow)	**B**	**D**
	High (quick)	**A**	**C**

Table 14 - Criteria used to define sub-groups

Sub-groups determined by counting ability		Percentage of errors	Frequency of items (per min)
A	Quick accurate	<SD	> SD
B	Slow accurate	< SD	< SD
C	Quick inaccurate	> SD	> SD
D	Slow inaccurate	> SD	< SD

The distribution of drivers for each group is shown in Table 15, which shows that none of the drivers were found to fit into sub-groups B and D.

Table 15 - Distribution of drivers into each sub-group

N° drivers for each sub-group		Auditory DT	Tactile DT
A	Quick accurate	1, 3, 7, 8, 9, 10, 13, 14, 15, 17	4, 6, 11, 18, 20, 24
B	Slow accurate	/	/
C	Quick inaccurate	12	2, 5,22
D	Slow inaccurate	/	/

6.2.3 IVIS task performance

With respect to performance in the IVIS no errors of recognition, and IVIS with errors of recognition tasks, the data for only 20/24 drivers has been taken into account, since sound was not successfully recorded for four subjects.

In the ‘no errors of recognition condition’, 14/20 drivers did not make any keyword errors. The other drivers made one or two substitution errors: for instance, they produced the wrong keyword (e.g. instead of “pick up”, “call” was given). Drivers also produced omission errors (e.g. forgot to say “no” in the correction procedure), misunderstood the voice synthesis response, or made an anticipation error (keyword was given too soon).

In the condition where errors of recognition were voluntarily induced, 7/20 drivers did not make any ‘keyword’ errors. Most drivers made between 1 and 5 errors. Therefore, the level of performance decreased in the condition which required system errors to be corrected. This suggests that workload was higher in this condition, compared to the IVIS no errors condition.

However, some caution must be taken for this assumption since the IVIS no errors condition always preceded the IVIS with errors condition. On the other hand, since taking time to become familiar with the IVIS no errors condition also required some extra workload, we can perhaps assume that the two conditions may well have placed the same degree of workload on drivers.

6.2.4 Distribution of hit rate

Hit rate corresponds to the percentage of responses between 200ms and 2 seconds. As shown in Figure 13 and Figure 14, a large variability was seen in hit rates to the auditory DT, compared to the tactile DT.

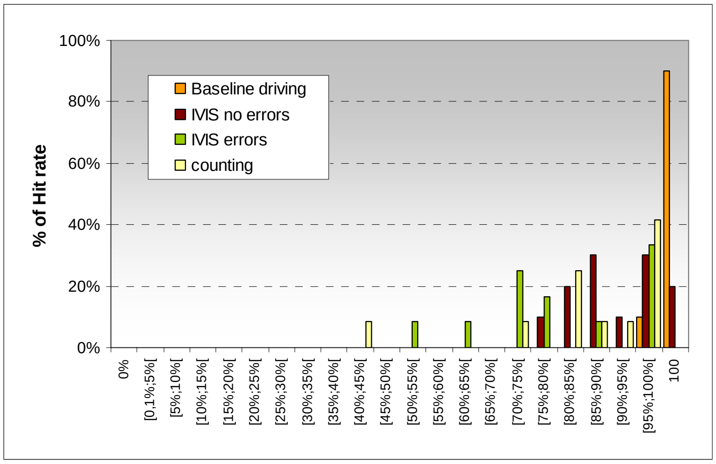


Figure 13 - Distribution of hit rates for the auditory detection task

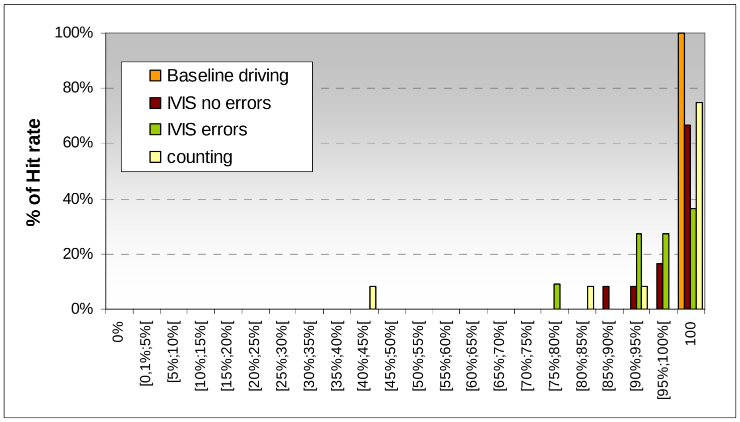


Figure 14 - Distribution of hit rates for the tactile detection task

In baseline driving, hit rate was very high in both DTs: all drivers (100%) detected all tactile signals, and almost all drivers (90%) detected all auditory signals. When performing the ‘IVIS with no recognition errors correction’ task (condition 2), hit rate decreased in both DTs compared to baseline driving (condition 1). This was particularly true for the auditory DT as variability between drivers was larger (hit rate varied from 75% to almost 100%) compared to the tactile DT: 65% of drivers detected all tactile signals, the others detect at least 85% of the signals.

When using the ‘IVIS with imposed correction of recognition errors’ task (condition 3), the pattern described in condition 2 was emphasized: hit rate decreased and variability between drivers increased in the auditory and tactile detection tasks. Also, auditory DT performance was more variable and lower than tactile DT performance (range from 50% to less than 100% versus variation from 75% to 100% respectively).

During the counting task (condition 4), hit rate was greatly variable between drivers, and was clearly degraded during performance of the auditory DT. However, a high hit rate was maintained in tactile DT, with 75% of drivers detecting all tactile signals.

To summarise, the distribution of hit rates showed the following main results:

- Global DT performance was at a high level for the auditory and tactile signals, as most drivers detected from 70% to 100 % of the auditory signals and 90% to 100% of the tactile signals.

- For both detection tasks, variability between drivers increased with secondary task demands (condition 3 and 4 > condition 2 > condition 1).

- For both DT, hit rate decreased with IVIS tasks and IVIS task difficulty (condition 1 > conditions 2 and 3 ; condition 2 > condition 3)

- Variability between drivers was higher in auditory DT compared to the tactile DT, for all conditions. Moreover, variability between drivers had a tendency to increase with secondary task demands (condition 3 and 4 > condition 2 > condition 1).

- Tactile DT was not found to be sensitive to the counting task.

As shown in Figure 15, for the auditory detection task, hit rate decreased significantly from baseline with the addition of all three secondary tasks. However, the ANOVA only showed a difference in hit rate between the IVIS errors and IVIS no errors conditions (Table 16). As shown in Table 17, for the tactile DT, only hit rates for the IVIS errors task were significantly lower than the baseline condition. Also, hit rates for the IVIS errors and IVIS no errors conditions continued to be reliably different from each other. The ANOVA also showed that hit rates were significantly higher for the tactile detection task compared to the auditory detection task, but only in the IVIS errors and IVIS no errors conditions ($p < .05$).

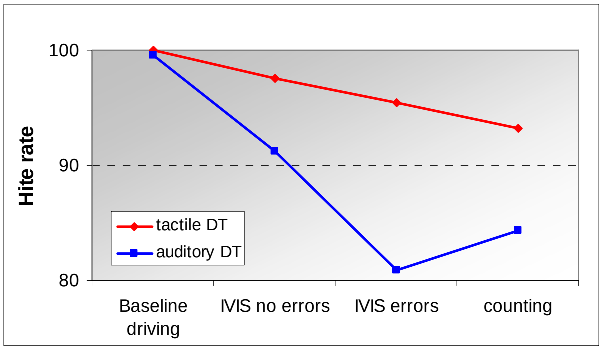


Figure 15 - Hit rate for each detection task and secondary task condition

Table 16 - Results of the ANOVA for the auditory DT

auditory DT	baseline driving	IVIS no errors	IVIS errors	counting
baseline driving				
IVIS no errors	p<.01			
IVIS errors	p<.005	p<.005		
counting	p<.05			

Table 17 - Results of the ANOVA for the tactile DT

Tactile DT	baseline driving	IVIS no errors	IVIS errors	counting
baseline driving				
IVIS no errors	-			
IVIS errors	p<.05	p<.05		
counting	-			

Hit rate was also plotted for each driver group according to their performance on the counting task (see section 6.2.2 for a description of how this categorisation was done). As shown in Figure 16, hit rate is high for drivers who made few or no errors on the counting task, for both tactile and auditory detection. On the other hand, drivers who made many errors in counting were also less successful at detecting the auditory signals, while they detected almost all tactile signals.

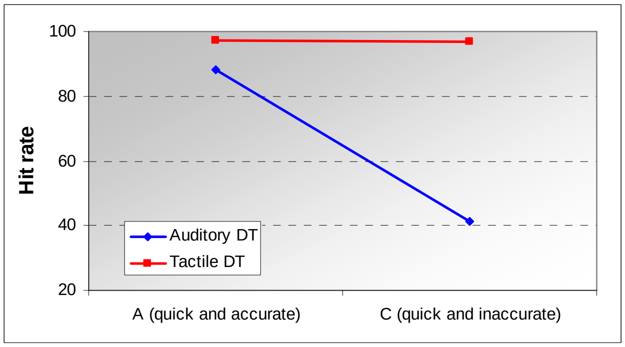


Figure 16 - Hit rate plotted by drivers' performance on the counting task

6.2.5 Distribution of RT

The RT distribution for each detection task is shown in

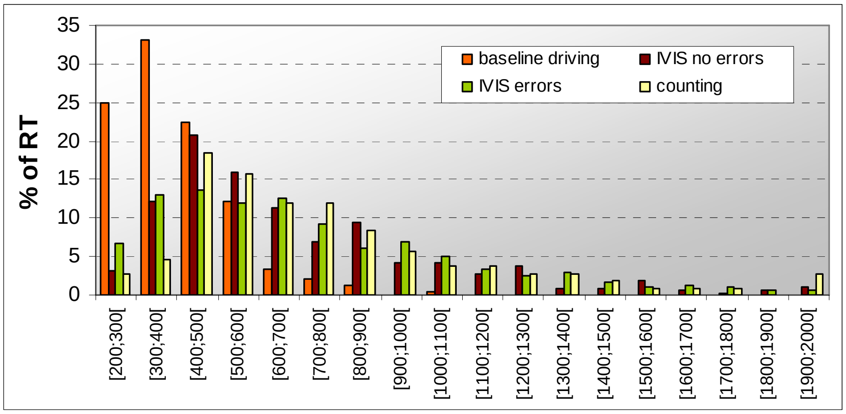


Figure 17 - Distribution of RT for the auditory DT

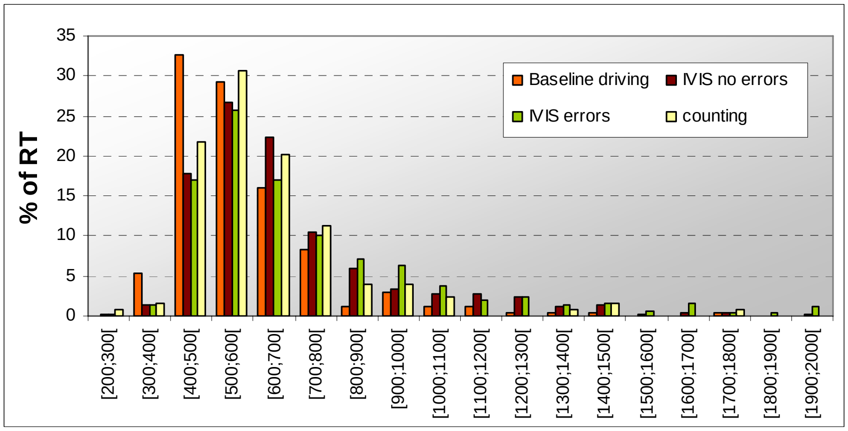


Figure 18 - Distribution of RT for the tactile DT

The distribution of response times between 200ms and 2seconds showed more variability for auditory DT compared to the tactile DT, for both baseline and secondary task conditions. Also, range of response times was larger for the auditory DT (starting from more than 200ms versus 400ms). Response times are concentrated in a number of time periods for the tactile DT (from 400ms to 800ms), while response times for the auditory DT were more widely spread. A major difference between the distributions of response times to the two detection tasks was seen in baseline driving, as response to auditory signals was much shorter (see Figure 19 and Figure 20).

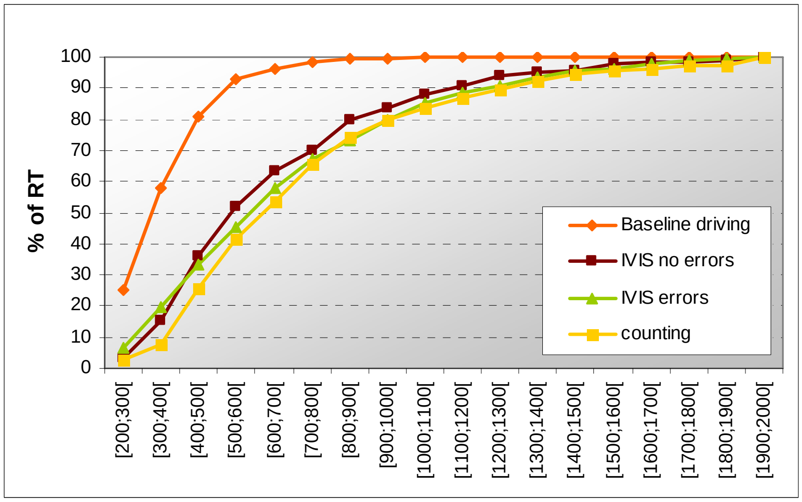


Figure 19 - Cumulative frequency of RT for the auditory DT

For auditory DT and tactile DT, significant differences are found when drivers were required to drive and perform the IVIS with recognition condition (high IVIS difficulty), compared to baseline driving and driving in the IVIS without errors condition: distribution of RT showed longer RT for the most difficult condition (IVIS with errors correction). Also, for the tactile DT, when drivers were required to drive and achieve the IVIS no errors condition, distribution of RT showed significantly longer RT compared to baseline driving.

No significant difference was observed between driving with the counting task and baseline driving.

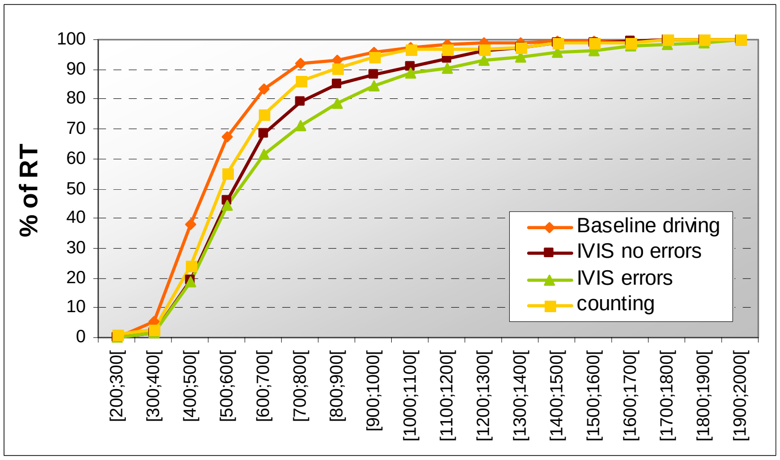


Figure 20 - Cumulative frequency of RT for the auditory DT

A major response time to auditory signals was observed between baseline driving and other conditions which required secondary task performance (whether low or high level of difficulty). When drivers had no additional task, most of them responded very quickly to most signals as 95% of RT was less than 700ms. However, in other conditions, whatever the secondary task performed (IVIS or counting), RT increased notably compared to baseline driving. However, the difference between the various levels of secondary task demand was relatively small.

To summarise, the distribution of response times showed the following main findings:

- For all conditions, response times to the auditory signals showed larger variability than response times to the tactile signals.
- A major difference between the two types of detection task was found in baseline driving, with shorter response times for the auditory task.
- With secondary task demands, response times to auditory signals and tactile signals had a tendency to increase, and this tendency was more visible for response times to the auditory signals.
- Counting was not found to affect the tactile DT (significant effect in auditory DT).

An analysis of variance showed a difference in mean RT for the auditory DT between baseline and the three secondary tasks (p<.01, Figure 21). However, there was no difference in RT to the auditory DT across the three secondary task conditions. For reaction time to the tactile DT, a difference from baseline was only seen for the IVIS errors and IVIS no errors conditions (p<.01), which were also found to be reliably different from each other (p<.05). Finally, in the baseline condition only, mean RT to auditory signals was found to be significantly shorter than mean RT to tactile signals.

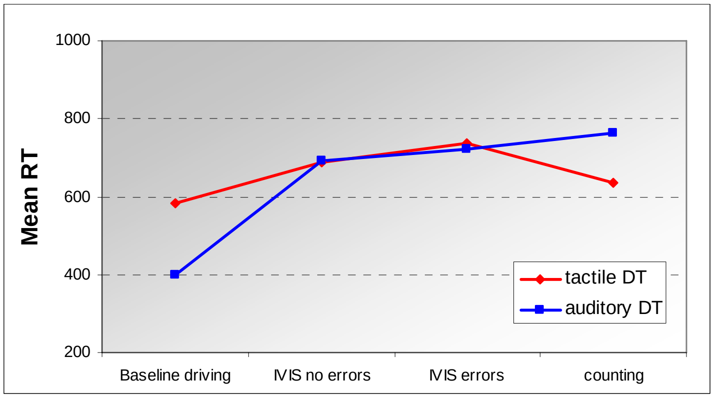


Figure 21 - Mean reaction time for each detection task and secondary task conditions

Reaction time data for each detection task was also plotted according to drivers’ counting ability (Figure 22). An analysis of variance showed that the worse the counting performance, the higher the subject’s RT to the two detection tasks (auditory: *p*<.001, tactile *p*<.01).

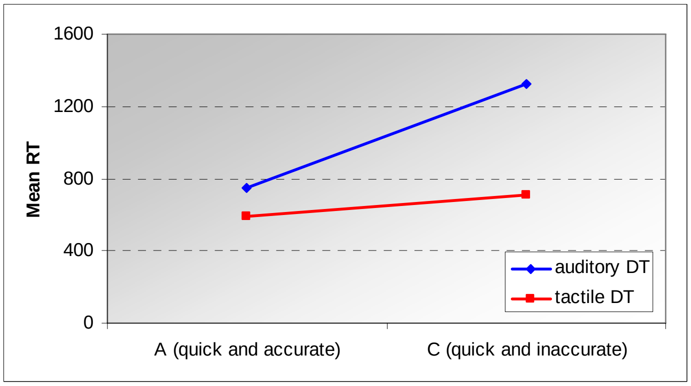


Figure 22 - Mean reaction time to the two DTs plotted by drivers’ counting performance

6.2.6 Percentage of late responses

The percentage of late responses (RT > 2 sec) in the two detection tasks was found to be quite low (<10%). A comparison of the two types of detection task showed a larger proportion of late responses for the auditory DT, compared to the tactile DT (Figure 23). However, this difference was only significant for the 'IVIS no errors' condition. Analyses of variance also showed a rise in late responses for the auditory DT with the addition of secondary tasks, although this was only significantly different for the 'IVIS with errors' and 'counting' conditions. The percentage of late responses in the tactile detection task did not show a reliable rise with secondary task performance.

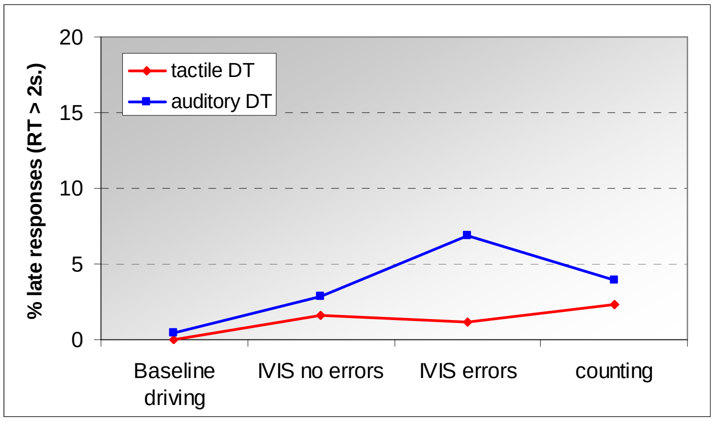


Figure 23 - Percentage of late responses for each detection task

6.2.7 Omissions rate

The rate of omissions (signals not detected) was quite low (<20%), and once again, there were generally more omissions in the auditory DT compared to the tactile detection task, with this difference being reliable for the 'IVIS with errors' and 'IVIS no errors' conditions (Figure 24). Addition of the secondary tasks caused a reliable rise in the number of omissions in the auditory DT, which was significantly different from baseline for the 'IVIS with errors' and 'IVIS no errors' conditions. However, performance of the secondary tasks did cause a reliable increase in the number of omissions for the tactile DT

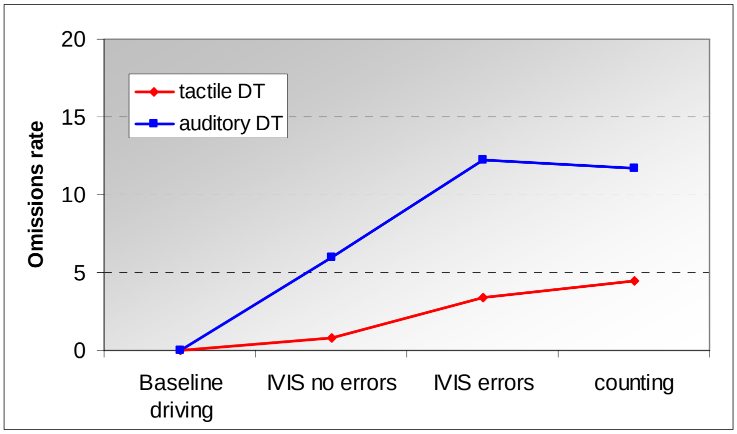


Figure 24 - The rate of ommisions for each detection task

6.3 Summary of results

Drivers were found to miss less tactile signals but also respond later to tactile signals, compared to auditory signals.

In general, the tactile DT was found to be less sensitive to disruptions by the secondary tasks, compared to the auditory DT (see Table 18). For example, Hit rate and RT to the tactile DT were not found to be affected by concurrent counting (except for a few drivers who had serious difficulties in the counting task).

Table 18 - Summary of performance in the two detection tasks

Comparison of conditions		**IVIS no errors - driving**	**IVIS errors - driving**	**Counting - driving**	**IVIS no errors – IVIS errors**
Auditory DT	Hit rate	p<.01	p<.005	p<.05	p<.005
	Mean RT	p<.001	p<.001	p<.001	-
	RT Distribution	-	p<.001	p<.001	-
	% late response	-	p <.01	-	p <.01
	% omissions	p <.05	p <.01	-	p <.05
Tactile DT	Hit rate	-	p<.05	-	p<.05
	Mean RT	p<.01	p<.001	-	p<.05
	RT Distribution	p<.01	p<.001	p<.001	-
	% late response	-	-	-	-
	% omissions	-	-	-	-

6.4 Discussion

A series of road tests were conducted to evaluate the sensitivity of the Detection Task (DT) method, to evaluate IVIS functions.

Drivers were required to perform two tasks: a counting backwards task and a multifunction system with vocal interactions (input/output) which required a destination selection or dialling task. At the same time, one group of drivers (50%) had to detect auditory stimuli whilst another group (50%) were required to detect tactile signals.

The results may be explained by a “narrowing of the attentional field”, since detection of signals in both modalities – auditory and tactile – is impacted by an increase in task demand (IVIS difficulty and counting task). Results also showed differences in performance between the two DT, with more disruption from the IVIS secondary task (with vocal dialog) on the auditory DT. This is thought to be due to a possible “confusion effect” (Wickens, 2002) between the vocal interactions with the IVIS and detection of the auditory signals from the DT.

Moreover, results showed different level of sensitivity of measures of DT performance (response time, hit rate, percentage of late response, omissions rate) : leading to question the driver’s strategy when performing the detection task.

For example, tactile DT was not sensitive to the counting task, unlike the auditory DT. Also, mean response time to tactile signals was sensitive to demand induced by IVIS task difficulty, while mean response time to auditory signals was not. Moreover, when the comparison between tactile DT and auditory DT was made for each condition of driving, differences were found between the two detection tasks for baseline driving and IVIS conditions.
Such differences in performance between tactile and auditory DT can be explained by differences in the physical characteristics of signals (felt intensity / frequency of signals, position of sensors). In other words, since the level of signal perceptibility and the potential confusion with ambient in-vehicle noise was not estimated, they could have been different for the two modalities. Also, difference in response time to auditory and tactile signals may be linked to the rapidity/length of information processing channel, and the position of sensors could also present differences (loudspeaker are placed in rear seat passenger, and vibrator placed in the driver's seat). Finally, any difference in auditory and tactile performance may also be due to physical interference of the auditory task by sounds related to the IVIS or drivers' own voice during counting.

The findings from this experiment were perhaps difficult to interpret due to one or a combination of the following factors:

- Response time in different conditions can be compared providing conditions are comparable in terms of duration, knowing that response times present more variability along time. Here, same temporal scale should be used for all compared conditions.
- The instructions given to drivers, which was to drive and use the IVIS (or count) and detect signals and the consequent status of the DT in cognitive processing (central or peripheral processing) is questionable. In other words, asking drivers to perform "*secondary tasks and DT at the same time as well as they can"* may be interpreted differently from one driver to another. Moreover, since the instructions given to drivers for the DT was to "*respond by pressing the response button, as soon as you detect a signal*" a bias may have been introduced, promoting the driver to give priority to DT because of high temporal demand. Especially if the driver is highly motivated to perform as well as possible.
- Therefore, perhaps drivers should have been instructed to treat the DT as a "peripheral task", if this task was to be used specifically to evaluate residual attentional resources.
- Despite the relevance of the counting task, interpretation of DT in the counting condition was difficult as many factors influenced performance. This included, the instructions given for the counting task, where no particular objectives for best performance were given. Therefore, drivers were free to assign their own objective to the task, perhaps leading to different levels of performance. In other words, some drivers may have favoured to count accurately while some drivers preferred to count quickly. Depending on drivers' strategy, attentional demands induced by the counting task were different. Moreover, ability to count (to estimate mental demands) should have been taken into account

7 Experiment 3 (Leeds)

The purpose of the Leeds simulator experiment was to study performance on a variety of Detection Tasks, and examine how this performance is affected by different driving environments and various secondary tasks.

Earlier findings from the European project HASTE suggest that the effect of a *visually demanding* IVIS on driving behaviour is very different to that of a *non-visual* IVIS which nevertheless imposed a high degree of workload (see Ostlund et al., 2004; Jamson & Merat, 2005). In particular, these two IVIS had very different effects on the distribution of drivers' eye-movements, with the demanding *non-visual* IVIS causing a reduction in peripheral saccades, and a concentration of gaze towards the centre of the road (Victor et. al, 2005). Therefore, one conclusion from these HASTE studies was that when performing a highly demanding non-visual IVIS, drivers' attention towards peripheral visual events in the driving scene may be reduced, and this would clearly have detrimental effects on safety.

Therefore, the first aim of the experiment described in this report was to investigate this idea further, by examining drivers' reaction time to a number of visual stimuli, which were presented in the drivers' peripheral view, within the driving scene. In many ways, the criteria used for these visual stimuli were akin to the conventional Peripheral Detection Task (PDT, van Winsum et al., 1999). However, the crucial difference was that the stimuli were presented in the driving scene itself, rather than within the car. Clearly then, this is a method which can only be used within a driving simulator study. Also, since attention to moving objects is known to be captured faster than stationary objects (e.g. Raymond, 2000), this 'visual detection task' was either stationary in the visual scene, or made a brief movement upon appearance on the visual scene. The reason for using two versions of this visual detection task was that if drivers' reaction time to the moving version was faster, this would prove to be a more sensitive version of the task, and could therefore be used for this purpose in subsequent studies. To examine drivers' reaction time to detection tasks presented in other modalities, a tactile and auditory version of the detection task was also presented to subjects.

Performance on these four detection tasks during driving was examined, and then compared to when drivers were also engaged in one of two in-vehicle information systems (IVIS). All drivers completed two IVIS tasks: a 'Phone task' which required dialling a series of telephone numbers into a touch-screen pad and a 'backward count' task, which involved counting backwards from a given number, in steps of 7. It was assumed that the first IVIS task involved relatively little demand from central resources but did require a reasonable degree of visual/manual interaction by the driver. On the other hand, the counting backwards task was thought to be quite demanding on central cognitive resources, but required no visual or manual input from the driver. The selection of these two tasks would therefore allow us to examine the effect of two very different IVIS on the detection of various other signals during driving.

7.1 Method

7.1.1 Participants

Twenty four drivers (8 male, 16 female) participated in this study. The average age of participants was 33.29 (range 19-58) and average driving experience was 13.25 years (range 1-37).

7.1.2 Apparatus and Tasks

Driver behaviour was studied using the Leeds Advanced Driving Simulator (LADS). The LADS is based on a complete Rover 216GTi with all of its basic controls and dashboard instrumentation still fully operational (Figure 25). A real-time, fully textured and anti-aliased 3-D graphical scene of the virtual world is projected on a 2.5m radius, cylindrical screen in front of the driver. Realistic sounds of engine and other noises are generated by a sound sampler and two speakers mounted close to each forward road wheel. Although the simulator is fixed-base, feedback is given by steering torques and speeds at the steering wheel. Data were collected at 60Hz and include information of the behaviour of the driver (i.e. driver controls), that of the car (position, speed, accelerations etc.) as well as information on other autonomous vehicles in the scene (e.g. identity, position and speed).

Figure 25 - The Leeds Advanced Driving Simulator

The Phone secondary task was presented on a 6.4” X 6.4” LCD touch screen, positioned to the right of the steering wheel (see Figure 26). The telephone numbers for this task were presented from a set of speakers positioned in the back of the car. These speakers were also used to present the numbers for the backward count secondary task.

Figure 26 - The LCD touch screen showing the phone secondary task

Presentation of the detection tasks was controlled by the simulator software. The visual and moving visual detection tasks consisted of a red circle (radius 29.3mm x 30.93mm) presented randomly to the left or right of the screen. For the moving visual DT, the red circle moved 0.1 second after its appearance, by a distance of 0.1 metre and in a randomly chosen direction. The auditory stimulus was a burst of broadband noise, presented through the left and right speakers of the Rover, which were located in the driver and passenger doors. The intensity of this stimulus was not formally measured, although we ensured that the broadband stimulus was clearly audible during the drive and secondary task presentation. Finally, the touch detection task was presented via a small 1 cm^2 vibrotactile pad (tactor), built for this experiment by Leeds University. This pad was placed on drivers' neck, just below their left ear (Figure 27). The presentation of all stimuli for the detection task was at a random rate of once every 2 to 5 seconds. The stimulus then lasted for 2 seconds, or until the subject pressed the response button. The response button used for all DTs was similar to that employed by Volvo (see Figure 28).

Figure 27 - The vibrotactile pad

Figure 28 - The response button used for all detection tasks in the Leeds experiment

7.1.3 Driving scenario

A rural environment with two levels of difficulty (two lane straight sections, two lane s-shaped curves) was used for this experiment.

To allow maximum data collection, a car following task was also introduced. To create a following scenario that was sensitive to the fact that different drivers adopt different driving speeds, the first section of the virtual drive included both a straight and curved section in which no lead vehicle was present. This was to allow the measurement of *"free speed"* choice for each individual driver. This value was measured separately on both straight and curved sections. Once this was achieved, the lead vehicle was introduced at an intersection, building up to a *driver-dependent speed*. The driver dependent speed was defined as either:

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

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

The data collection part of the road layout (test section) consisted of 5 straight and 5 curved sections which were laid end-to-end. Two sections (one curved and one straight) were also required at the beginning of the road to allow for *free-speed* measurement. At the end of the *free-speed* sections, there was a cross-road intersection to allow for the introduction of the lead vehicle and around 1km of road to allow the lead vehicle to reach its desired speed (free speed – 10%, maximum speed 96km/h speed limit). In addition, there were the 10 second *filler sections* and 10 second *task description sections* between each *test section*. The width of the road was 7.3 metres made up of two carriageways of 3.65 metres each.

7.1.4 Design and procedure

All participants were provided with written instructions for the experiment, and completed a consent form. They then had the opportunity to practice each IVIS and detection task. Each driver then drove a practice road, the layout of which was similar to the experimental road described above. Here, they practiced driving in isolation, followed by driving with the IVIS and detection tasks.

Each subject completed three runs of the experimental road: one run for each of the two IVIS and one baseline run. A within-subjects design was used, whereby each of the 4 detection tasks was presented twice in a particular run, once for straight and once for curved sections. The start of each detection task coincided with the start of an IVIS (apart from baseline conditions). Presentation of the detection tasks was in a random order and each run of 10 sections also consisted of a baseline drive where no DT was presented. The order of the three runs was counterbalanced across subjects.

For the *Phone* IVIS task, drivers heard a series of 7 digit telephone numbers, which were pre-recorded and announced at predetermined sections of the road, using speakers in the rear of the car. Drivers were asked to enter the digits on the touch screen monitor, followed by the 'enter' key when they were happy with their response. The next 7-digit telephone number was then presented after the enter key was pressed. This Phone task lasted for 60 seconds.

For the *Backward Count* IVIS task, subjects heard a pre-recorded three digit number, which was presented from the car's rear speakers. They were asked to count aloud backwards from this number, in units of 7, and were then told to stop counting after 60 seconds had lapsed. Subject response was recorded by the experimenter.

Finally, all drivers were asked to rate their driving performance after completion of each IVIS/DT combination. They heard an announcement from the rear speakers of the car which asked them to rate their driving performance, using a number between 1 (I drove very badly) to 10 (I drove very well).

7.1.5 Data analysis

The driving and DT data were analysed using Univariate analyses of variance (ANOVA). Sidak post hoc tests were used when appropriate. The factors and levels used are shown in Table 19.

Table 19 - Factors and levels used for the Leeds Univanova

Factor	Levels	Type
IVIS task type	3 (2 tasks + baseline)	fixed
DT task type	5 (4 tasks + baseline)	fixed
Road level	2 (straight, curve)	
Subject	24 (number of participants)	random

7.2 Results

7.2.1 Detection task performance

The distribution of reaction time responses for all detection tasks is shown in Figure 29. This pattern was found to be the same for all detection tasks, when plotted individually.

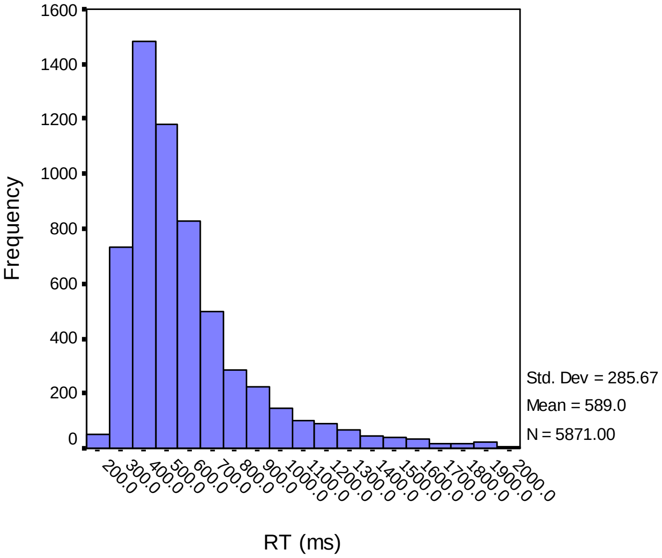


Figure 29 - Distribution of the reaction time data

Univariate ANOVA on reaction time showed a significant main effect of secondary task (*F* (2,46) = 176.657, *p*< .001) and detection task (*F*(3,69) = 25.147, *p*< .001). Post hoc tests showed that, for all detection tasks, reaction time increased from Baseline to Backward Count to the Phone task (see Figure 30). Such a high reaction time for the Phone task may have simply been because there was a competition in manual response for the Phone and detection tasks. Reaction time for the tactile detection task was found to be significantly lower than for all other tasks, suggesting that subjects were perhaps most sensitive to this task. Also, the results suggest that regardless of whether or not subjects' visual attention was taken away from the driving scene, their least rapid response was always to the visual detection tasks, with reaction time significantly higher than for the other two tasks (*p* < .01). Of course, in the absence of eye movements it is perhaps difficult to be certain of the direction of drivers' visual attention at these times. Finally, a significant interaction was found between the detection task and secondary task. One source of this interaction is seen between the visual and tactile task, with reaction time to the visual task being higher during the phone task.

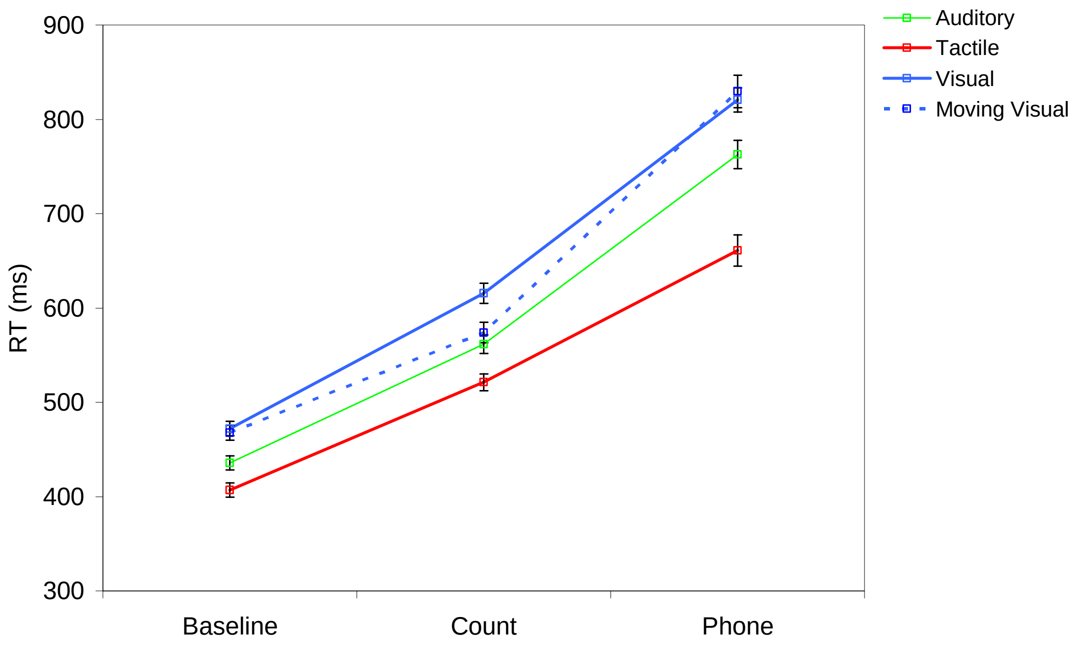


Figure 30 - Reaction time for each detection task[3]

The Unianova also showed an effect of secondary task on hit rates (F (2,46) = 4.631, $p< .05$), where the percentage of hits for the Phone task was significantly lower than both baseline and the backward count task (Figure 31). Therefore, the manual demand required by the phone task clearly reduced the capacity for responding to the detection tasks. Hit rate for the moving visual task was found to be generally higher than the visual detection task but this difference was not found to be reliably different according to the Unianova.

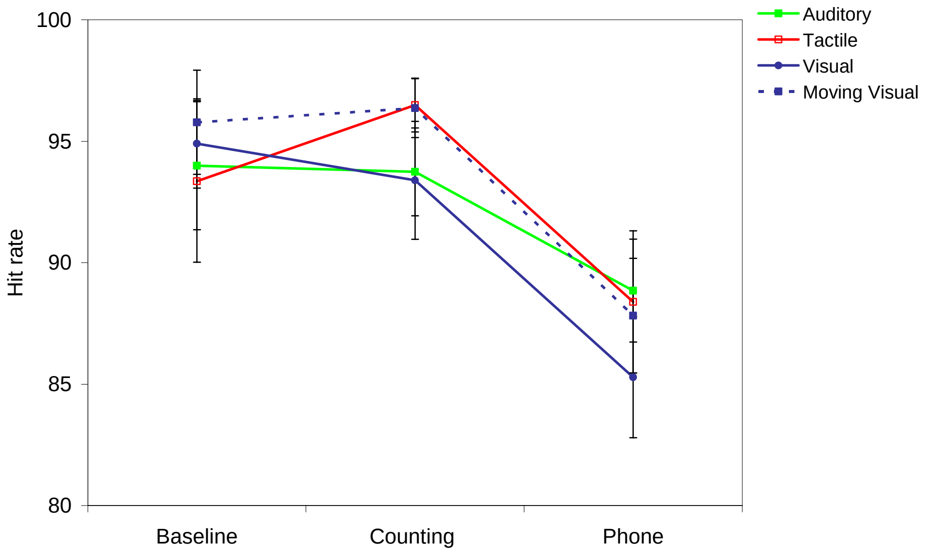


Figure 31 - Percentage of hit rates for each detection task

---

[3] Error bars for this and subsequent figures represent standard error of mean.

No formal analyses were done for the percentage of misses and cheats in the detection tasks, but the effect of driving and secondary tasks on these values is shown in Figure 32, which shows that both these measures were highest for the Phone task.

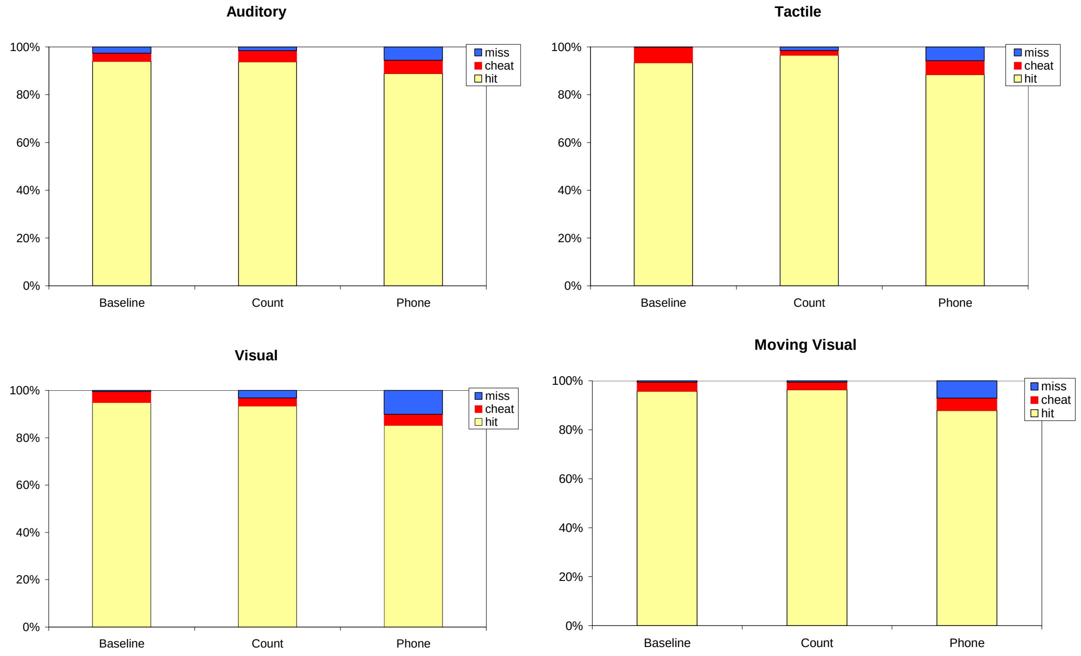


Figure 32 - Hits, misses and cheats for each detection task

7.2.2 Performance on the IVIS tasks

7.2.2.1 The Phone task

Performance in the Phone task was not differentially affected by the various detection tasks. On average, participants attempted to enter just over 5 seven-digit telephone numbers during each 1 minute Phone task. However, the average number of correct responses was only 39.6%, and a large variability was observed between subjects (SD = 27.6). The mean reaction time for a correct response was found to be 8.9 seconds (SD = 2.6 seconds).

7.2.2.2 The backward count task

Performance in the counting task was not found to change in the presence of the different detection tasks. However, there was great variance in performance between subjects (Table 20). In order to examine whether subjects' counting ability influenced the response to the detection task signals, subjects were categorised by their performance in the backward count task and detection task data were re-plotted accordingly.

Table 20 - Summary of performance in the backward count task

Counting performance→	Accuracy of counting (Percentage of errors)				Frequency of counts (per min)			
	Auditory DT	Tactile DT	Visual DT	Moving visual DT	Auditory DT	Tactile DT	Visual DT	Moving visual DT
Mean	5.73	5.72	4.67	5.12	17.60	17.25	17.96	18.60
SD	9.54	9.78	9.23	9.10	9.82	8.75	9.37	10.02
Minimum	0	0	0	0	2	2	2	4
Maximum	33.33	44.44	50	42.86	45	41	45	43

As with the method used by PSA (see section 6.2.2), subjects were divided into four sub-groups based on their counting ability (see Table 21).

Table 21 - Categorisation of subjects by counting ability in the Leeds experiments

Subject group	Counting ability	%Error	Frequency of counts
A	Quick and accurate	0	>mean
B	Slow and accurate	<mean	<mean
C	Quick and inaccurate	>mean	>mean
D	Slow and inaccurate	>mean	<mean

NB. Means were used instead of SD (as used by PSA) since too many subjects fell into more than one group with SD.

When plotted according to performance on the counting task, the percentage of hit rates was shown to be higher for the group who performed well on the counting task, compared to the group of drivers who had the most difficulty with the task (Figure 33). Similarly, reaction times to the detection tasks were found to be smaller for subjects who performed well on the counting task, compared to those whose response was slow and inaccurate (Figure 34).

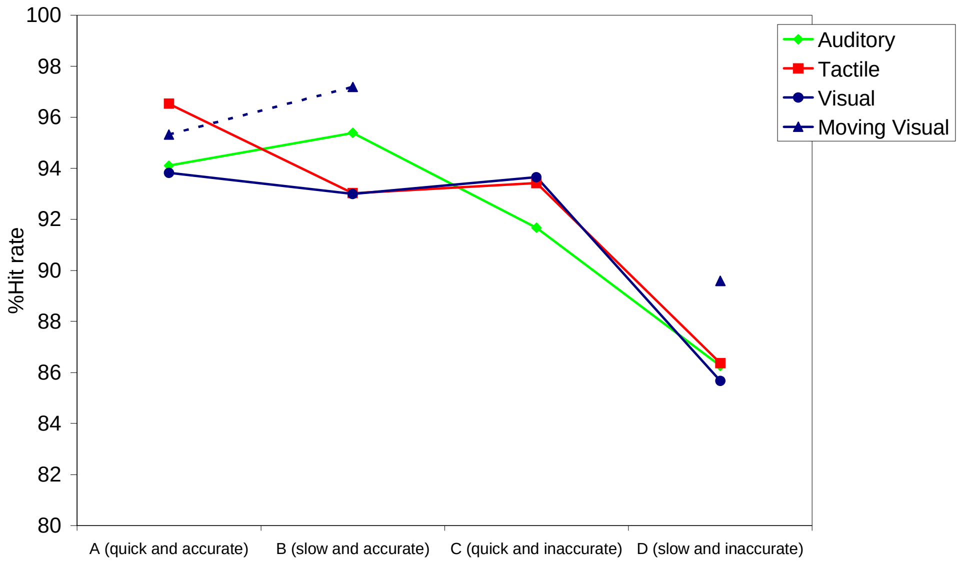


Figure 33 - Hit rate for each detection task plotted by drivers' ability on the counting task

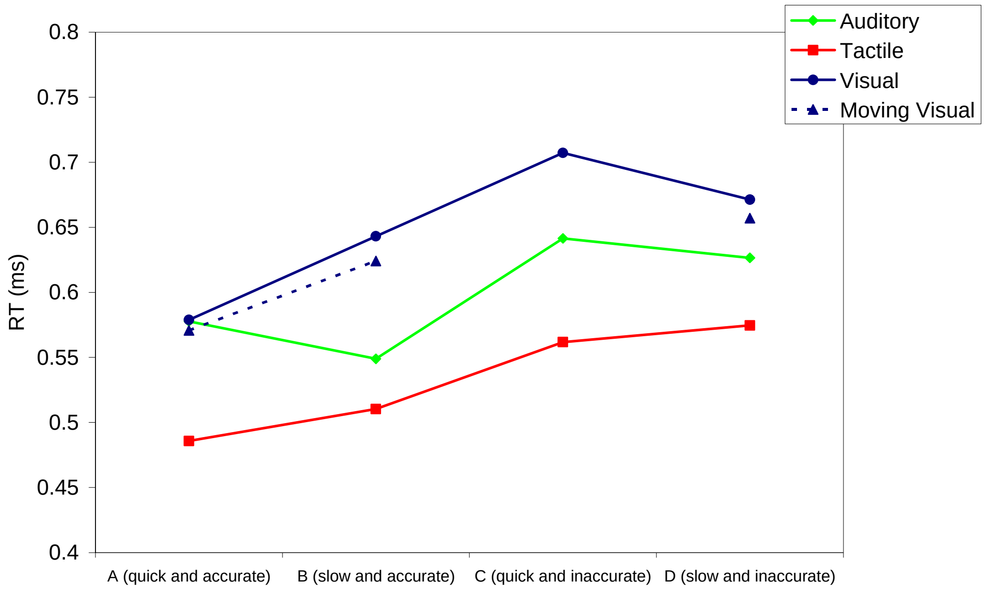


Figure 34 - RT to each detection task plotted by drivers' ability on the counting task

7.2.3 Driving performance

Driving performance was found to be the same across all detection tasks, and there was no difference between baseline driving performance and driving performance with the detection tasks. The effect of the two secondary tasks on driving is outlined below.

7.2.3.1 Longitudinal measures

Drivers reduced their mean speed when performing the Phone task (*F* (2,46) = 10.882, *p*< .001), but continued to drive at the same speed as baseline during the backward count task (see Figure 35). Mean speed was slower for the curve sections than the straights. The interaction between road type and secondary task was also found to be significant (*F* (2,505) = 7.112, *p*< .01), showing a much sharper reduction in mean speed by participants when they performed the Phone task in the curve sections. The effect of secondary tasks on minimum speed was similar.

Standard deviation of speed was also found to increase significantly with the addition of the Phone task (*F* (2,46) = 26.206, *p*< .01), but showed no difference from baseline when subjects conducted the backward count task.

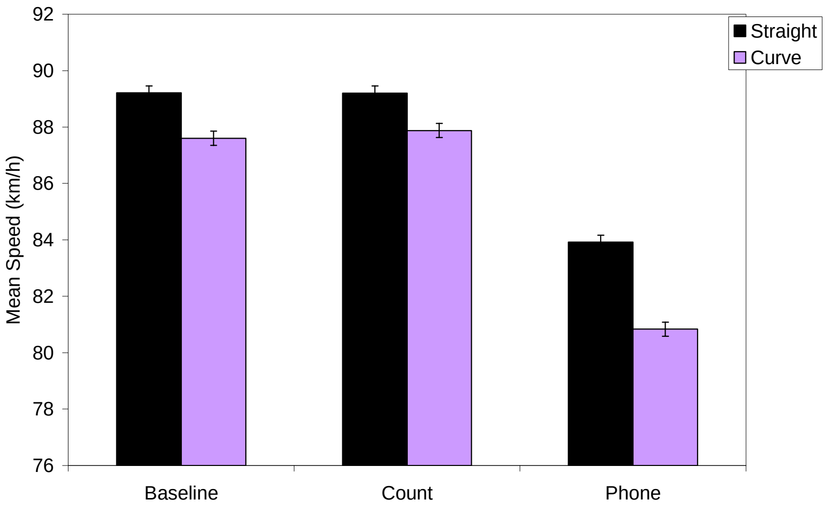


Figure 35 - The effect of secondary task performance on mean speed

Drivers increased their mean and minimum headway to the lead car when they were performing the Phone secondary task (*F* (2,46) = 23.881, *p*< .001 and *F* (2,46) = 12.876, *p*< .001 respectively). As shown in Figure 36, there was no main effect of road type, but there was an interaction between road type and secondary task, where headway was even longer when subjects were attempting the Phone task during their drive of the curved sections.

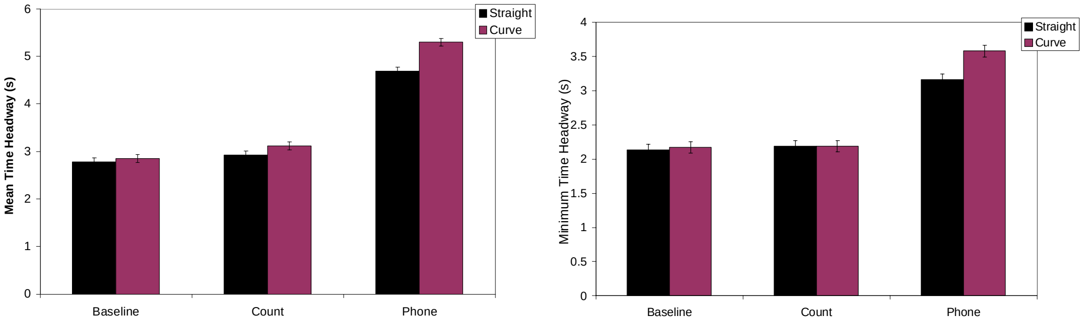


Figure 36 - Mean (left) and minimum (right) time headway

7.2.3.2 Lateral and steering measures

Standard deviation of lateral position was shown to rise significantly when subjects were driving with the Phone task (*F* (2,46) = 27.186, *p*< .001), but showed no difference from baseline (and indeed a minor fall) during the backward count task (Figure 37).

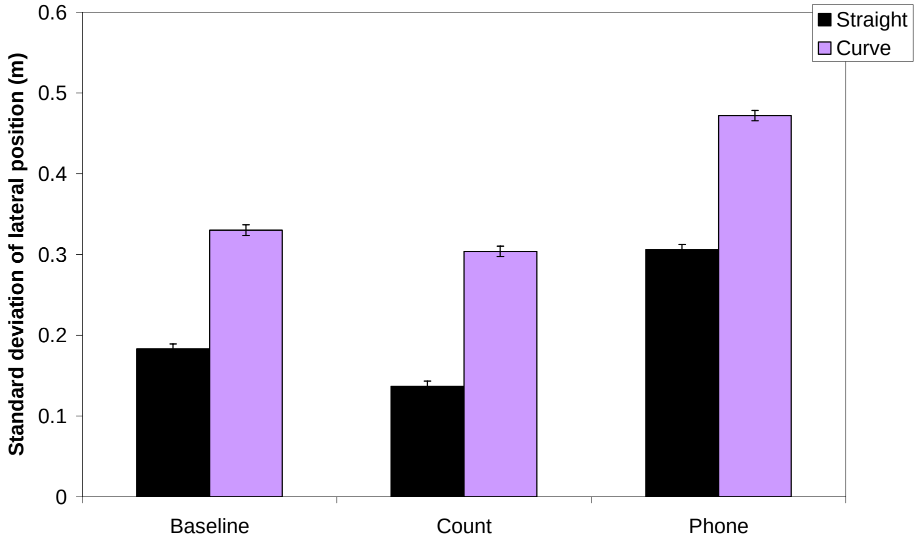


Figure 37 - Standard deviation of lateral position

This reduction in lateral stability with the Phone task was also observed in the steering data. For instance, the number of small (1^o) steering wheel reversals (rr_st1) was found to increase significantly from baseline (F (2,46) = 26.533, p< .001). As shown in Figure 38, the number of small reversal rates were also seen to rise with the backward count task, suggesting that both tasks resulted in some increase in workload, compared to baseline. However, the number of rapid steering wheel turns (RSWT), which are a measure of rapid reactions by the driver to avoid driving off the road or colliding with an object, was only seen to rise for the Phone task (this was true for thresholds of 20, 40 and 70 per minute, see Figure 39).

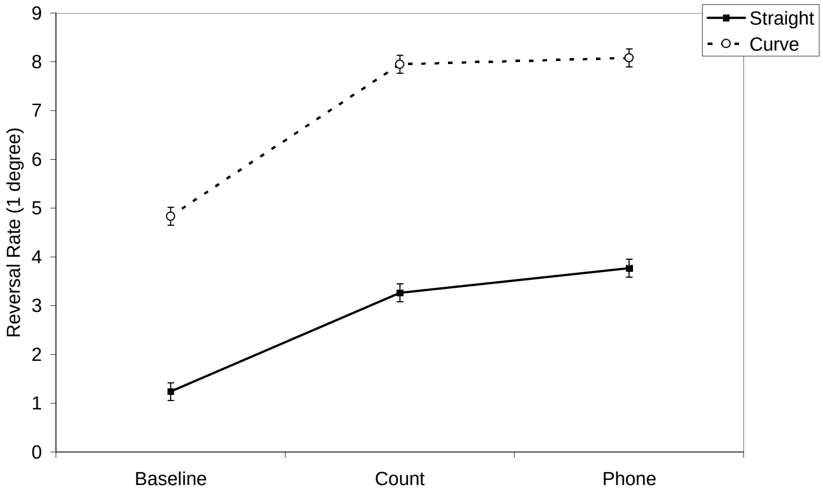


Figure 38 - Steering wheel reversal (1^o)

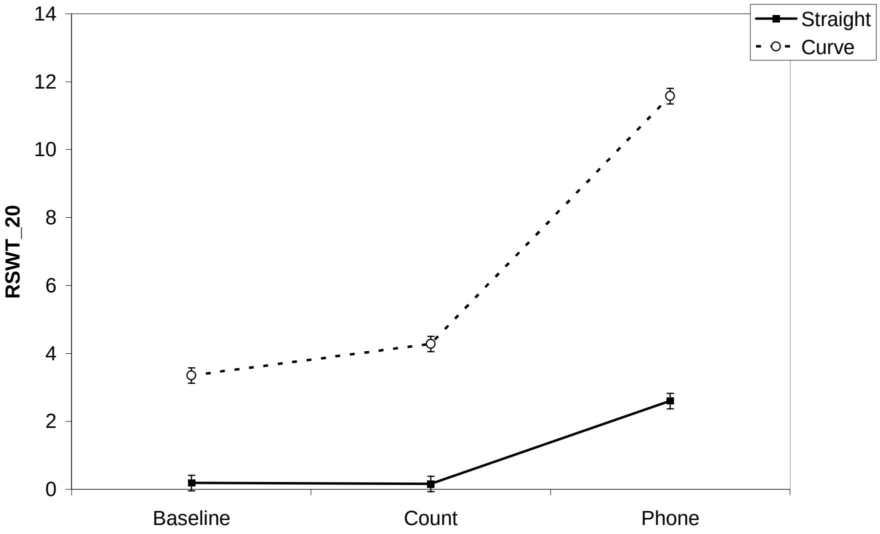


Figure 39 - Rapid steering wheel turns, using a threshold of 20°

7.3 Subject ratings

Results of the subjective rating showed no difference in drivers' evaluation of their own driving performance across the three secondary task conditions (baseline, backward count, phone; see Figure 40).

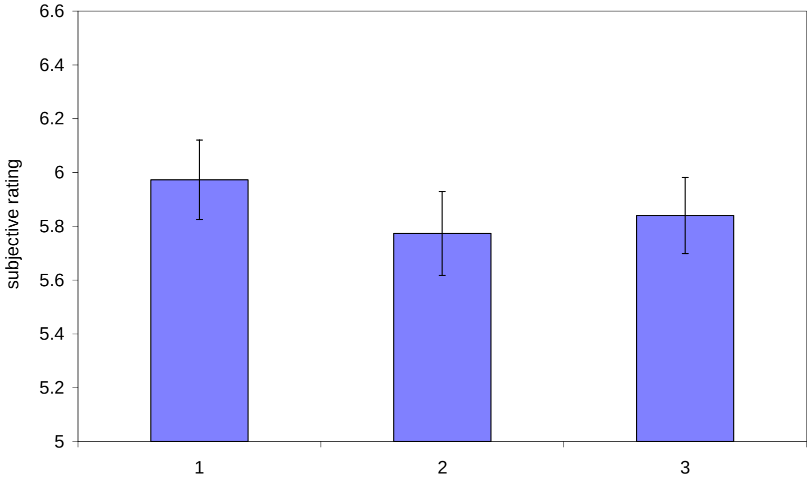


Figure 40 - Subjective rating

7.4 Discussion

Performance in the detection tasks was mainly assessed by examining the percentage of hit rates and the average reaction time. As with the results of PSA and Volvo, hit rate was quite high, with average performance between 85-95%. Compared to baseline, drivers' ability to react to the various detection tasks deteriorated with the addition of both IVIS tasks, but performance was worst during the Phone dial task, regardless of detection task. This longer reaction time during the Phone task is thought to be related to a conflict in manual response which was required both by the detection tasks and the also the Phone task. The results also showed a significant difference in response across the detection tasks, with best performance in the tactile detection task and the longest reaction times observed for the visual DT. It is difficult to conclude, however, whether this difference in reaction time across the detection tasks was related to a competition for shared modality or processing resources with the IVIS task.

For example, since the backward count task is likely to be quite demanding on central resources (and therefore quite similar to the HASTE aCMT task), we would expect a reduction in glances towards the peripheral visual field (as confirmed by the driving performance) and therefore a higher response rate to the visual DT, which was indeed observed. However, since both the backward count task and the auditory detection task were sound-based, they may have competed with each other in terms of input modality, which could explain the rise in reaction time (compared to baseline) seen in the auditory detection task. This would also explain why reaction time to the tactile detection was lowest, since this task did not share the same modality as the backward count task. Similarly, entering the numbers in the Phone task required subjects' visual attention, while the announcements of the telephone numbers itself may have interfered with the auditory detection task. Once again, this would account for the relatively lower mean reaction time seen for the tactile detection task.

In conclusion, such a difference in RT values across the three modalities of detection task provides support for the multi-component models of information processing, which suggest that performance of two or more tasks is best achieved when there is a minimum degree of conflict between processing resources.

Performance in the counting and phone task was not found to be affected by the detection tasks. However, a re-organisation of detection task performance by counting ability showed that the subjects that produced the lowest number of responses and the most number of errors in the counting task were also those with the least number of hit rates and longest reaction time to the detection tasks. Therefore, subjects' overall processing speed was clearly playing a big part in their ability to perform the two tasks whilst driving.

Driving performance was not found to be affected by the detection tasks. However, the two IVIS were found to have quite different effects on driving performance, with results quite similar to those found in the HASTE experiments. In other words, drivers reduced their average speed and increased their headway with the lead car during performance of the mostly visual Phone task. On the other hand, they continued to travel at the same speed as baseline driving and failed to increase their headway when performing the backward count task. In addition, while lateral deviation was found to increase with the Phone task, a slight reduction in this measure was seen during the backward count task. These findings suggest that drivers were perhaps aware of their limitations during performance of the Phone task, and adjusted their driving behaviour accordingly. However, they were clearly underestimating the effect of the backward count task on driving, which was nevertheless causing possible detrimental effects to their driving performance.

8 Comparison between different Methods/metrics

In order to compare results between the three sites, estimated effect sizes (i.e. the standardised differences between each task and baseline), were calculated for each detection and secondary task.

8.1 Description of procedure

For each site, estimated effect (d′) was calculated as the standardised difference of means:

$$d = \frac{\bar{x}_{Treatment} - \bar{x}_{BL}}{SD}$$, where

$$SD = \sqrt{\frac{(n_{Treatemetn} - 1)SD^2_{Treatement} + (n_{Treatment} - 1)SD^2_{BL}}{(n_{Treatment} + n_{BL} - 2)}}$$

Figure 41 shows the result of these calculations for all sites. For Volvo and Leeds, the various detection tasks show similar effect sizes for the Phone and Email secondary tasks. All detection tasks have a similar effect on the Question task used by Volvo, although, clearly, neither detection task was as sensitive to the Question task as the other secondary tasks. Similarly, while the effect of the backward count used by Leeds was the same for all detection tasks, neither detection task was as sensitive to backward count as to the Phone task.

The effect of each detection task on the backward count (which was the only secondary task common to all sites) shows quite different effects for each site. In particular, a very large effect of this task is seen on the PSA auditory task, although it is difficult to establish whether some of this effect might be due to confusions between hearing the auditory beeps during the backward count task. The tactile detection task used by PSA was not found to be very sensitive to any of the secondary task, compared to the auditory detection task, although this may have been because of difficulties felt by drivers in terms of detecting the vibration in the seat.

One reason for this attempt to standardise results between the three sites, was to establish whether, in terms of its sensitivity to secondary tasks, one type of detection task would be favoured against another. In particular, vastly different effect sizes on the various detection tasks would suggest that a particular detection was more/less sensitive to secondary task performance than another. However, only the results of the PSA studies showed a distinction in sensitivity of the detection tasks, and this may well have been related to the problems with hearing the beeps or feeling the tactile device. On the other hand, the Leeds and Volvo experiments suggest that, broadly speaking, performance was not affected by a difference in modality or position of detection task stimuli. However, it is difficult to make firm conclusions without further work and it is certainly not wise to base all conclusions on the results from estimated effect sizes. For example, reaction time data for the Leeds experiments did show more sensitivity of the tactile task compared to the other detection tasks, and this cannot be disregarded.

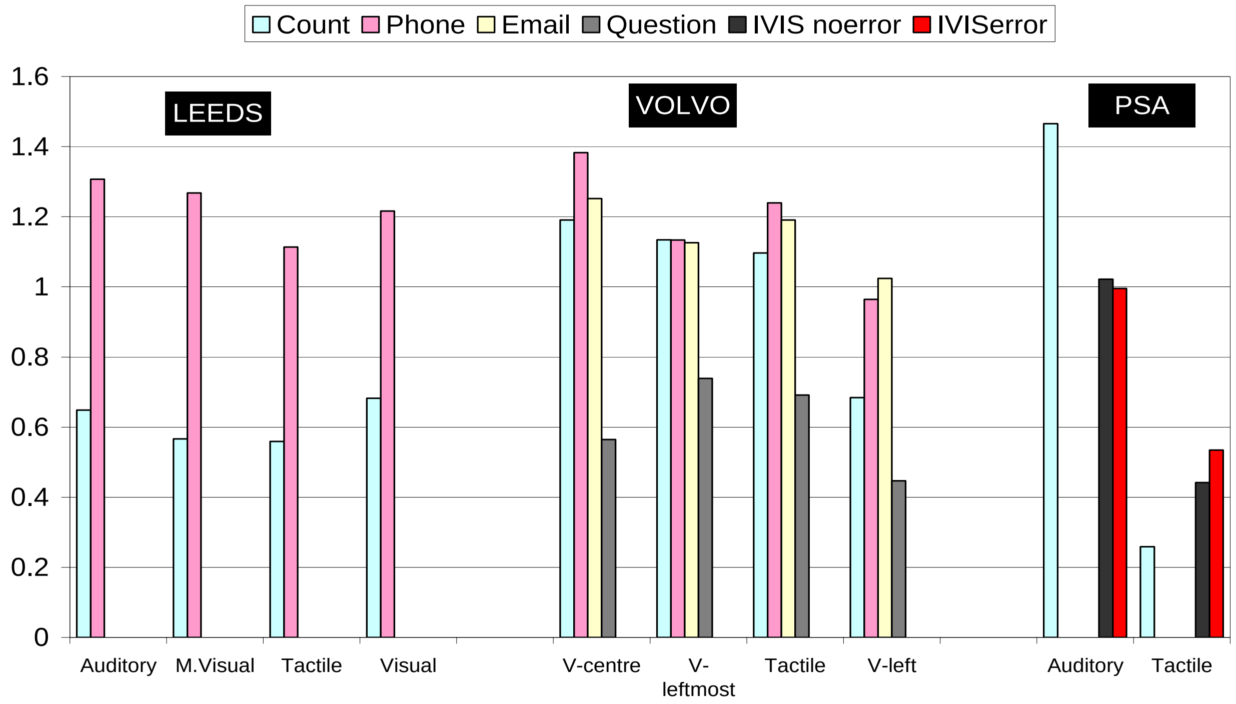


Figure 41 - Estimated effect size for each detection secondary task across the three sites

9 Specification of suggested method/metric to be used in the AIDE test regime

Based on the above findings and following extensive discussion between the three sites, the partners involved in task 2.2.3 were required to recommend the use of one particular detection task for further assessment and use by partners in task 2.2.7 “Empirical comparison of methods for off-line workload measurement”.

Overall, the effect size plots suggest that there was not much difference in sensitivity of the various detection tasks, while reaction time data for the detection tasks showed variable sensitivity, but only for the Leeds and PSA data.

Since, a recommendation was to be made on the most appropriate detection task for assessing the safety of an IVIS, discussions across partners highlighted two points which were pivotal to the reaching the decision on the final recommendation:

i. At present, most IVIS rely on visual and/or auditory output, while the tactile modality is (at present) not used as frequently.
ii. Detection of visual and auditory signals in the field will depend greatly on the surrounding environment. In particular, bright lighting or noise related to other events in the field environment can reduce detection rate of tasks based on the visual and auditory modality respectively.

Based on these observations, the tactile detection task has been recommended for further testing, since its detection is less likely to be hindered by unexpected or changing environmental factors, and it is also less likely to be competing for the same perceptual modality as the IVIS task under examination.

9.1 Description of Method

9.1.1 Modality, presentation rate, duration and response

As described above, after careful consideration of the three detection tasks, the tactile stimulator was considered the most appropriate to use for further testing by task 2.2.7. The most appropriate form of tactile stimulation is thought to be one produced by small a vibration motor, similar to that used by Leeds University. The device used to cause this vibration at Leeds was obtained from a mobile telephone and encased in a 1 cm^2 balsa wood holder. There are no special recommendations in terms of stimulus intensity, although it is important that the stimulus is tested under various conditions to ensure it is always perceptible by the subject. The rate of presentation is random between 2 and 5 seconds (each value has the same probability of being drawn). Each tactile stimulus should last for at least 1 second (can be up to 2 seconds long) and vibration should be suspended when a response is given by the subject. Response to the stimulus should be given by a button attached to the index finger of the hand opposite to the one used for changing gear. Also, the response button itself should produce a haptic feedback to the subject.

9.1.2 Placement of equipment

In order to avoid interference with changing gear or any other driving-related manual movements, the tactile stimulator should be placed on the driver's neck and secured with medical tape. The wire attached to the stimulator should be long enough to prevent interference with driving. For an example of the desirable setup, see Figure 42.

Figure 42 - The recommended position for the tactile stimulus

9.1.3 Scenario

A mixed environment including field, rural and urban roads is considered a suitable scenario. In the field, consideration should be given to some degree of control, using similar road sections, intersections etc.

9.1.4 Preferred design

A within-subject is suggested, and a minimum of 12-15 subjects will be required. In order to collect an adequate number of data points, the detection task should be at least 30 seconds long, and around 3 repetitions of the task are recommended.

9.1.5 Instructions

Drivers should be told that their primary task is driving. After that, equal priority should be given to the IVIS/ADAS and detection tasks. Also: *Drivers should be asked about how they prioritised between the tasks at the end of the drive.*

9.1.6 Metrics

Four main measures should be calculated for the detection task. These are:

- The percentage of hit rates, defined as the number of stimulus responses between 200 and 2000 ms, over the total number of stimuli.
- Mean reaction time of the hit rates.
- Percentage of cheats: any stimulus response lower than 200ms.
- Percentage of misses or omissions: a ‘no response’ or any response longer than 2000 ms, respectively.

While it is important to calculate all of the above measures, interpretation is largely based on the rate of hits and the average reaction time.

Analysis should ideally use Univariate ANOVA with Secondary Task (IVIS/AIDE system) as the main factor and subject as a random factor. Alternatively repeated measures ANOVA can be used if sample sizes are the same. The following syntax can be used in SPSS:

```
UNIANOVA
  rt BY task pos tp
  /RANDOM = tp
  /PLOT = PROFILE( task*pos )
  /EMMEANS = TABLES(task)
  /EMMEANS = TABLES(pos)
  /SAVE = RESID
  /PRINT = HOMOGENEITY
  /DESIGN = task pos tp task*pos task*tp pos*tp.
```

9.1.7 Interpretation

In order to interpret the results, a hit rate of at least 70% for the detection task is desired. Analysis of variance on reaction time can then provide information about how performance in the detection task is affected by performance on the IVIS/AIDE system.

10 Summary and Conclusions

The aim of the AIDE task 2.2.3 was to examine the suitability of a series of detection tasks for the safety assessment of IVIS and ADAS. Using previous findings from the HASTE project, the AIDE State of the Art report 2.2.1, and by utilising models such as Wickens's multiple resource theory, and modern theories of attention as a theoretical framework, a series of dual task studies were designed which examined the effect of IVIS performance on drivers' ability to detect and respond to random stimuli, presented in and out of the vehicle. In particular, the effect of stimulus position, stimulus modality and IVIS demand was investigated.

Results were somewhat in conflict across the three sites, in that only Leeds and PSA found a difference in reaction time to the various modes of detection task, which suggested that detection was faster when the IVIS and detection task shared the least number of resources of task processing or response. Nevertheless, since a recommendation was to be made to partners involved in task 2.2.7, following some discussion, the tactile detection task is proposed to be the most appropriate for use in further investigations.

The reason for choosing the tactile device is that, at present, not many IVIS tasks use this modality for the presentation of warnings or information. Therefore, this modality is the least 'overloaded' in drivers. The danger with using a visual or auditory modality is that, if performance in either of these modes is shown to drop in the presence of an IVIS, it is not as easy to establish whether this is because drivers failed to see (or hear) the detection task because their vision (or hearing) was engaged in the driving or IVIS task (modality conflict), or because they failed to respond to the detection task because they were responding to the driving/IVIS task (response conflict). If a failure to detect is due to shared modality, then the IVIS is clearly not safe to be used whilst driving, as it is taking drivers' visual/auditory attention away from the driving task. Of coursem this is particularly dangerous if it is drivers' visual attention that is taken away from the scene.

An alternative to this argument is that since the lowest response times were found for the tactile task (at least in the Leeds study), it would imply that use of this modality for the safety assessment of IVIS may suggest that many IVIS are safe to use with driving. It is at this point that the question, "how much interference is too much interference?" must be asked. Unfortunately, the answer to this question is still missing within the literature in this area. However, it is perhaps significant to end this report by referring to a line of reasoning used by Lamble, Kauranen, Laakso & Summala (1999) on this subject. In most cases, average reaction time to any of the detection tasks used in the above experiments rose by around 200-400ms between the 'baseline' and 'IVIS' conditions. Previous studies on the effect of alcohol in driving (e.g. DeWaard & Brookhuis, 1991) have shown a rise in brake reaction time of 168 ms, by blood alcohol concentrations around the legal limit. Therefore, it is important to consider whether the use of any device which increases drivers' reaction time more than that of legal BAC levels is acceptable to traffic safety.

11 Innovation

The visual version of the PDT has been used extensively. However, as far as we are aware, this is perhaps the first time that comparisons have been made between performance of the auditory, tactile and visual version of the same task in the same driving environment, using the same in-vehicle task in each case.

12 References

Acosta, E. & Dickman, J.D. (1984). Information processing in the visual periphery: influence of mental workload. *Proceedings of the Human Factors Society, 28th Annual meeting, 737-739.*

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. & 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.

Ball, K.K., Beard, B.L., Roenker, D.L., Miller, R.L., Griggs, D.S. (1988). Age an visual search: expanding the useful field of view. *Journal of the Optical Society of America*, 5, 2210-2219.

Belt, B. L. (1969). Driver eye movements as a function of low alcohol concentrations. *Technical Report, Driving Research Laboratory*, Columbus, OH:Ohio State University.

Bellamy, L. J. (1984). The application of visual lobe measurement to visual perception. In A. G. Gale and F. Johnson (Eds), *Theoretical and Applied Aspects of Eye Movement Research.* Amsterdam: Elsevier.

Bouma, H. (1970). Interaction Effects in Parafoveal Letter Recognition. *Nature*, Vol. 226, pp. 177-178.

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.

Bundesen, C., Habekost, T., & Kyllingsbæk, S. (2005). A neural theory of visual attention. Bridging cognition and neurophysiology. *Psychological Review, 112*, 291-328.

Chan, H.S. & Courtney, A.J. (1993). Effects of cognitive foveal load on a peripheral single-target detection task. *Perceptual and Motor Skills*, 77, 515-533.

Chin, E., Gautheret, L., & Nathan, F. (2004). Peripheral Detection Task (PDT): RoadSense results. Internal document within the AIDE project.

Corbetta, M. & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews Neuroscience*. 3 (3), 201-215.

Crundall, D. & Underwood, G. (1998). Effects of experience and processing demands on visual information acquisition in drivers. *Ergonomics*, 41, 4, 448-458.

Crundall, D., Underwood, G., & Chapman, P. (1999). Peripheral Detection Rates in Drivers. In A. G. Gale, I. D. Brown, C. M. Haslegrave, & S. R. Taylor (Eds.), *Vision in vehicles VII,* (pp. 261-270). Amsterdam: Elsevier. (North-Holland).

Crundall, D., Underwood, G., & Chapman, P. (2002). Attending to the peripheral world while driving. *Applied Cognitive Psychology*, 16, 459-475.
Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, 18, 193–222.

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

Engström, J., Åberg, N., Johansson, E., Hammarbäck, J. (2005) Comparison between visual and tactile signal detection tasks applied to the safety assessment of in-vehicle information systems. In PROCEEDINGS of the Third International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design.

Findlay, J. M., & Gilchrist, I. D. (2003). *Active vision: The psychology of looking and seeing*. Oxford: Oxford University Press.

Harbluk, J.L., Noy, Y.I., & Eizenman, M. (2002). The Impact of Cognitive Distraction on Driver Visual Behaviour and Vehicle Control. *Transport Canada Technical Report*. Available at: http://www.tc.gc.ca/roadsafety/tp/tp13889/en/menu.htm

Harms, L & Patten, C. (2003). Peripheral detection as a measure of driver distraction. A study of memory-based versus system-based navigation in a built-up area. *Transportation Research Part F: Traffic Psychology and Behaviour 6*, 23-36.

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.

Holmes, D.L., Cohen, K.M., Haith, M.W. & Morreson, F.J. (1977). Peripheral visual processing. *Perception and Psychophysics*, 22, 571-577.

Ikeda, M. & Takeuchi, T. (1975). Influence of foveal load on the functional visual field. *Perception & Psychophysics*, vol. 18 (4), 255-260.

Israelsson, E., & Karlsson, N. (2003). How does a speech user interface affect the driving task?. *In Proceedings of the Contemporary Ergonomics Conference*, Edinburgh 2003.

Itti, L. (2005). Models of Bottom-up Attention and Saliency. In L. Itti, G. Rees, and J. K. Tsotsos (Eds.), *Neurobiology of Attention* (pp. 576-582). Amsterdam: Elsevier.

Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, 40, 1489-1506.

Jamson, A.H & Merat, N. (2005). Surrogate In Vehicle Information Systems and Behaviour: Effects of Visual and Cognitive Load in Simulated Driving. *Transportation Research Part F: Traffic Psychology and Behaviour*, 8(2), 79-96.

Janelle, C. M., Singer, R. N., & Williams, A. M. (1999). External distraction and attention narrowing: visual search evidence. *Journal of Sport and Exercise Psychology*, 21, 70-91.

Johansson, E., Engström, J., Cherri, C., Nodari, E., Toffetti, A., Schindhelm, R., Gelau, C. (2005). Review of existing techniques and metrics for IVIS and ADAS assessment. EU project AIDE, project IST-1-507674-IP, Deliverable 2.2.1

Kaluger, N. A. & Smith, G. L., Jr. (1970). Driver eye movement patterns under conditions of prolonged driving and sleep deprivation. Night Visibility and Driver Behaviour. *Highway Research Record*, No. 336, Washington, D.C.

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

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.

Koch, C. (2004). *The quest for consciousness. A neurobiological approach*. Englewood: Roberts and Company Publishers.

Koch, C., & Ullman, S. (1985). Shifts in selective visual attention: towards the underlying neural circuitry. *Human Neurobiology*, 4, 219–227.

Mackworth, N.H. (1965). Visual noise causes tunnel vision. *Psychonomic Science*, 3, 67-68. marching to two different drummers: Perceptual aspects of the difficulties. *Journal of Experimental Psychology: Human Perception and Performance,* 11, 814-828.

Koch, C. (2004). *The quest for consciousness. A neurobiological approach*. Englewood: Roberts and Company Publishers.

Lamme, V. A. F. (2005). The difference between visual attention and awareness: A cognitive science perspective. In L. Itti, G. Rees, and J. K. Tsotsos (Eds.), *Neurobiology of Attention* (pp. 167-174). Amsterdam: Elsevier.

Lamble, D., Kauranen, T, Laakso, M. & Summala, H. (1999). Cognitive load and detection thresholds in car following situations: safety implications for using mobile (cellular) telephones while driving. *Accident Analysis and Prevention,* 31, 617-623.

Logan, G. D. (2002). An instance theory of attention and memory. *Psychological Review, 109*, 376-400.

Land, M.F., & Lee, D.N. (1994). Where we look when we steer. *Nature*, 369, 742-744.

Mackworth, N.H. (1965). Visual noise causes tunnel vision. *Psychonomic Science*, 3, 67-68.

Martens, M. & Van Winsum, W. (1999). Activation criteria and warning strategies: Driving simulator results. IN-ARTE Project (TR4014). I.D. 7.1 Unpublished technical report. TNO Soesterberg.

McAdams, C. J., & Maunsell, J. H. R. (1999). Effects of attention on orientation-tuning functions of single neurons in macaque cortical area V4. *Journal of Neuroscience*, 19, 431-441.

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.

Nunes, L., & Recarte, M. A. (2002). Cognitive demands of hands-free-phone conversation while driving. *Transportation Research Part F, 5,* 133-144.

Olsson, S. (2000). Measuring driver visual distraction with a peripheral detection task. Master's thesis. ISRN LIU KOGVET-D-0031-SE, Linkoping University, Department of Behvioural Sciences, Sweden.

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.

Plainis, S., Chauhan, K., Murray, I.J. and Charman, W.N. (2003). The effect of performing a cognitive task on detection thresholds; eccentricity and task-difficulty effects. *Vision in Vehicles III*

Plainis, S., Murray, I.J. and Chauhan, K. (2001). Raised visual detection thresholds depend on the level of complexity of cognitive foveal loading. *Perception*, 30, 1203-1212.

Recarte, M. A., & Nunes, L. M. (2003). Mental Workload While Driving: Effects on Visual Search, Discrimination and Decision Making. *Journal of Experimental Psychology: Applied, 9 (2)*, 119-137.

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

Raymond, J.E. (2000). Attention modulation of visual motion perception. *Trends in Cognitive Sciences*, 4(2), 42-50.

Recarte, M., & Nunes, L. (2003). Mental Workload While Driving: Effects on Visual Search, Discrimination, and Decision Making. *Journal of Experimental Psychology: Applied,* 9, No. 2, 119-137.

Reynolds, J. H., & Desimone, R. (1999). The role of neural mechanisms of attention in solving the binding problem. *Neuron, 24*, 19-29.

Sanders, A.F., & Donk, M. (1996). Visual search. In: A.F. Sanders & O. Neumann (Eds.), *Handbook of perception and action: Attention III (pp. 43-77)*. London: Academic Press.

Schindhelm, R., Gelau, C., Montanari, R., Moreale, D., Deregibus, E., Hoedemaeker, M., De Ridder, S., Piamonte, P. (2003). Human factor tests on car demonstrator. EU project COMUNICAR, project IST 11595, Deliverable 6.4.

Spence, C., & McDonald, J. (2004). The cross-modal consequences of the exogenous spatial orienting of attention. In G. Calvert, C. Spence, C., & B.E. Stein (Eds.), *The handbook of multisensory processes*, pp. 3-25.
Cambridge, MA: MIT Press.

Tsotsos, J. K., Itti, L., & Rees, G. (2005). A brief and selective history of attention. In L. Itti, G. Rees, and J. K. Tsotsos (Eds.), *Neurobiology of Attention* (pp. xxii-xxxii). Amsterdam: Elsevier.

van der Weijgert, E. C. M. (1993). Foveal load and peripheral task performance: tunnel vision or general interference? *Proceedings of the second international conference on visual search.* Taylor & Francis. Pp. 341-348.

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

Verwey, W.B. (1993). First test on man-machine interface adaption to driving situations. Report TNO-IZF 1993 C44. Soesterberg: TNO Human Factors Research Institute.

Victor, T. W., & Åberg, N. (2005). *The driving performance cost of gaze eccentricity in full attention and cognitive load conditions.* Manuscript. Department of Psychology, Uppsala University, Sweden.

Victor,T.W., Harbluk, J.L. & Engström, J.A. (2005) Sensitivity of eye-movement measures to in-vehicle task difficulty *Transportation Research Part F: Traffic Psychology and Behaviour.*, 8(2), 167-190.
Victor & Johansson, 2005

Wickens, C.D. (1984). Processing resources in attention. In R. Parasuraman and R. Davies (Eds.), *Varieties of Attention* (pp. 63-101). Academic Press: New York.

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

Wickens, C.D (2002). Multiple resources and performance prediction. *Theoretical issues in ergonomic science*. Vol. 3, p. 159-177.

Williams, L.J. (1985). Tunnel Vision Induced by a Foveal Load Manipulation. *Human Factors,* 27(2), 221-227.

Williams, L.J. (1982). Cognitive Load and the Functional Field of View. *Human Factors,* 24(6), 683-692.

Williams, L. J. (1988). Tunnel vision or general interference? Cognitive load and attentional bias are both important. *American Journal of Psychology,* Vol. 101, No. 2, pp. 171-191.

Williams, L. J. (1995). Visual field tunneling in aviators induced by memory demands. *Journal of General Psychology.*

13 Appendix A: Specification summary

General parameter	Specific parameter	Specification	Specification
Stimulus presentation	Modality	Tactile (default)	Visual
	Spatial and temporal randomisation	No spatial randomization. One vibrator used. Random time interval between 3 and 5 seconds	No spatial randomisation. Single LED used. Random presentation between 3 and5 seconds
	Position	The tactile stimulator should be placed on the driver's neck and secured with medical tape. The wire attached to the stimulator should be long enough to prevent interference with driving.	The LED should be placed so that the stimulus is presented in or near the driver's central field of view through reflection in the windshield. Care should be taken to find a position where the influence of lighting conditions on stimulus visibility is minimised.
	Intensity	Be sure the stimulus is clearly perceivable while driving. *Caution:* If tested in the field ensure road conditions do not obscure the sensation of the tactile stimulus.	The intensity of the LEDs should be 2 candela
	Duration	The stimuli should last for a maximum of 1 second (discontinued when a response is given).	The stimuli should last for a maximum of 1 second (discontinued when a response is given).
Responses	Response device	The response should be given by a button attached to the index finger of the hand opposite the one used for changing gear.	The response should be given by a button attached to the index finger of the hand opposite the one used for changing gear.
	Type of response feedback	The response button should give mechanical haptic feedback. The click sound should be very subtle. Stimuli should d discontinue when a response is given.	The response button should give mechanical haptic feedback. The click sound should be very subtle. Stimuli should discontinue when a response is given.
Performance measures	Threshold criteria for a 'hit'	200-2000 ms.	200-2000 ms.
	Response time	Only computed for 'hits'. Defined as the time from stimulus onset to response.	Only computed for 'hits'. Defined as the time from stimulus onset to response.
	Percentage of misses	Calculated as 'no response' or anything above 2000ms	Calculated as 'no response' or anything above 2000ms
	Quality criterion	Minimum 70% average hit rate per task required for the measure for the task to qualify for analysis.	Minimum 70% average hit rate per task required for the measure for the task to qualify for analysis

General parameter	**Specific parameter**	**Specification**	**Specification**
	Additional measures	If testing is done in the field, drivers should be asked to rate their driving performance (simple scale of 1 to 10 is fine) after each Detection Task episode. This is to ensure that difficult traffic manoeuvres are not the only cause of lapses in DT performance. It is also wise that observers are present in the car to note unusual traffic conditions.	
Scenario		Mixed environments (see Fastenmeiers's divisions for example.). Controlled field experiment (road sections same, same intersections etcetera)	Mixed environments (see Fastenmeiers's divisions for example.). Controlled field experiment (road sections same, same intersections etcetera)
Experimental design	Number of test participants	Minimum 15. Three repetitions per task.	Minimum 15. Three repetitions per task.
	Grouping of subjects	Within group	Within group
	Statistical analysis	- Univariate ANOVA with Secondary Task (IVIS, AIDE system) as the main factor and subject as random factor. Alternatively repeated measures ANOVA (if sample sizes are the same). - Post-hoc analysis.	- Univariate ANOVA with Secondary Task (IVIS, AIDE system) as the main factor and subject as random factor. Alternatively repeated measures ANOVA (if sample sizes are the same). - Post-hoc analysis.

```
UNIANOVA
  rt BY task pos tp
  /RANDOM = tp
  /PLOT = PROFILE( task*pos )
  /EMMEANS = TABLES(task)
  /EMMEANS = TABLES(pos)
  /SAVE = RESID
  /PRINT = HOMOGENEITY
  /DESIGN = task pos tp task*pos task*tp pos*tp.
```

```
UNIANOVA
  rt BY task pos tp
  /RANDOM = tp
  /PLOT = PROFILE( task*pos )
  /EMMEANS = TABLES(task)
  /EMMEANS = TABLES(pos)
  /SAVE = RESID
  /PRINT = HOMOGENEITY
  /DESIGN = task pos tp task*pos task*tp pos*tp.
```

Instructions		Drivers should be told that their primary task is driving. After that, equal priority should be given to the IVIS/ADAS and detection tasks. Also: *Drivers should be asked about how they prioritized between the tasks at the end of the drive.*	Drivers should be told that their primary task is driving. After that, equal priority should be given to the IVIS/ADAS and detection tasks. Also: *Drivers should be asked about how they prioritized between the tasks at the end of the drive.*

14 Appendix B: Suggested terms for the AIDE glossary

* Terms in italics: suggested to be inserted in the general AIDE glossary.

Term	Definition	References	Alternative definition	References alt. definition	Notes	Category
*ADT**	*Auditory detection task – an auditory version of the PDT where visual stimuli are replaced by auditory stimuli such as 'beeps' or burst of white noise*					
ADAS	Systems that interact with the driver with the main purpose of supporting the driving task on the tracking and regulating levels.	Original AIDE definition			Based on the ECOM model, adopted as the Conceptual Framework of AIDE (Hollnagel, E. & Woods, D. D., 2005), the driving task can be described in terms of, potentially simultaneous, layered control processes: (1) tracking, (2) regulating, (3) monitoring and (4) targeting. Hollnagel, E. & Woods, D. D. (2005). Joint cognitive systems: Foundations of cognitive systems engineering. Boca Raton, FL: Taylor & Francis/CRC Press.	General
AIDE system	The Adaptive Integrated Driver-vehicle Interface targeted by the AIDE IP,	Original AIDE definition			The AIDE system consists of a basic set of HMI management components, in particular the ICA and the DVE monitor. Thus, the AIDE system does not include a specific set of applications or HMI I/O devices. Rather, the AIDE system should support	General

	implementing the AIDE meta-functions				different applications, I/O devices and configurations in a modular way.	
*DT**	*Detection Task – Different from the peripheral detection task in that stimuli are no longer presented in the visual periphery*					
Driver distraction	Attention given to a non-driving related activity	ISO TC22/SC13 WG8 CD 16673 (Occlusion Committee draft)			This definition is somewhat unclear due to the fact that “driving-related activity” and “driving performance” are not further defined. According to the AIDE Conceptual Framework, the driving task should not be seen as a single activity, but rather as a set of multiple simultaneous, and layered, control tasks. Distraction with respect to a given control process (e.g. tracking) could thus be viewed in terms of interference by another (driving- or non-driving related control process, typically resulting in degraded performance on the given control task. In practice, however, “driver distraction” is normally used to refer to interference with the tracking and/or regulating control tasks	Behavioural effects

Driving demand	The demands of the driving task	de Waard, D. (1996). The Measurement of Drivers' Mental Workload. ISBN 90-6807-308-7. Traffic Research Centre. University of Groningen.			Demand is determined by the goal that has to be attained by means of task performance, and is, once the goal has been set, external and independent of the individual driver (c.f. mental workload)	
Driving performance	Performance on the driving task				According to the AIDE Conceptual Framework, the driving task should not be seen as a single activity, but rather as a set of multiple simultaneous, and layered, control tasks. Thus, driving performance could be defined with respect to any of these control tasks. For example, driving performance on the tracking level would be related to the ability to keep the vehicle within acceptable safety margins. Similarly, performance on the regulating level would be related to the ability to select these safety margins based on the general situation assessment at the monitoring level.	Behavioural effects

Driving task	All aspects involved in mastering a vehicle to achieve a certain goal (e.g. reach a destination), including tracking, regulating, monitoring and targeting.	Original AIDE definition			Based on the ECOM model, adopted as the Conceptual Framework of AIDE (Hollnagel, E. & Woods, D. D., 2005), the driving task can be described in terms of, potentially simultaneous, layared control processes: (1) tracking, (2) regulating, (3) monitoring and (4) targeting. Hollnagel, E. & Woods, D. D. (2005). Joint cognitive systems: Foundations of cognitive systems engineering. Boca Raton, FL: Taylor & Francis/CRC Press.	General
IVIS	Systems that interact with the driver with the main purpose to support tasks on the targeting and monitoring levels, or do not support driving at all.	Original AIDE definition			Based on the ECOM model, adopted as the Conceptual Framework of AIDE (Hollnagel, E. & Woods, D. D., 2005), the driving task can be described in terms of, potentially simultaneous, layared control processes: (1) tracking, (2) regulating, (3) monitoring and (4) targeting. Hollnagel, E. & Woods, D. D. (2005). Joint cognitive systems: Foundations of cognitive systems engineering. Boca Raton, FL: Taylor & Francis/CRC Press.	General
*PDT**	*Peripheral Detection Task*				Maartens & van Winsum (1999)	*General*
Primary task	The task with the highest priority in a multi-tasking situation.	Original AIDE definition			In the automotive domain, the primary task is normally the drivimg task	General
Secondary task	A task with lower priority than the primary task in a multi-tasking situation.				In the present context, a typical secondary task is interacting with an in-evhicle system	General

Specification	Precise (formal if possible) description of an object within the scope of the task.	EAST-EAA (Safety terms for automation systems reliability and safety of complex systems; VDI/VDE 2000.)				SW Development /Architecture
System	A collection of components organized to accomplish a specific function or set of functions.	EAST-EAA (IEEE Recommended Practice for Architectural Description of Software-Intensive Systems; IEEE Standard P1471, IEEE Architecture Working Group (AWG), 2000.)	Set of elements, which interact according to a design; an element of a system can be another system, called a subsystem, which may be controlling system or a controlled system and may include hardware, software	Functional safety: safety instrumented systems for the process industry section; Part 1: Framework, definitions, system, hardware and software requirements; IEC2002.	The different components of the system are organized to accomplish a specific function or a set of functions.	SW Development /Architecture

			and human interaction			
Task	Process of achieving a specific and measurable goal using a prescribed method	ISO TC22/SC13 WG8 CD 16673 (Occlusion Committee draft)			Example: Obtaining guidance by entering a street address using the scrolling list method, continuing until route guidance is initiated (visual-manual task)	General
*TDT**	*Tactile Detection Task – a tactile version of the PDT, where visual stimuli are replaced by vibrotactile stimulators.*	*Enström et al. (2005)*				*General*

Use case	An intended or desired flow of events or tasks that occur within the vehicle and are directed to or coming from the driver in order to accomplish a certain system-driver interaction.	Original AIDE definition	A model of the usage by the user of a system in order to realise a single functional feature of the system.	EAST-EAA		SW Development /Architecture
*VDT**	*Visual Detection Task – similar to the PDT but visual stimuli are not necessarily presented in the peripheral field of view and can appear centrally or in the visual scene of a driving simulator.*					*Behavioural effects*
Visual demand	Degree of visual activity required to extract information from an object to perform a specific task	ISO TC22/SC13 WG8 CD 16673 (Occlusion Committee draft)			In general, visual demand depends on: (a) the quantity of information to be extracted and (b) the ease with which information can be resumed following any interruption	Behavioural effects