Development of an Occlusion Protocol with Design Limits for Assessing Driver Visual Demand

by T Horberry, A Stevens, S Cotter, R Robbins and G Burnett

Development of an Occlusion Protocol with Design Limits for Assessing Driver Visual Demand

by T Horberry, A Stevens, S Cotter, R Robbins and G Burnett

PPR 256
PPAD 9/161/020

PUBLISHED PROJECT REPORT

TRL Limited

PUBLISHED PROJECT REPORT PPR 256

DEVELOPMENT OF AN OCCLUSION PROTOCOL WITH DESIGN LIMITS FOR ASSESSING DRIVER VISUAL DEMAND

Version 1

By   T Horberry, A Stevens, S Cotter, R Robbins (TRL Limited) and

G Burnett (University of Nottingham)

Prepared for: Project Record: PPAD 9/161/020

Client: TTS1, Department for Transport
(Contract Manager: Adrian Burrows)

	Approvals
Project Manager	**Sally Cotter**
Quality Reviewed	**Alan Stevens**

CONTENTS

Executive summary

The key objective of this project was to address the issues required to establish an effective Occlusion methodology for the assessment of In-Vehicle Information Systems (IVIS). It was the intention that the research would help develop a robust visual demand measurement protocol based on Occlusion, and to propose criteria for the acceptability of different tasks. This would help inform DfT and others regarding which IVIS tasks should not be undertaken while driving.

This work was undertaken in five work packages:

Work Package 1 - Age

The impact of age on performance of IVIS tasks whilst applying the Occlusion technique was investigated in order to understand how age should be controlled for when producing a protocol for Occlusion. A small number of previous studies have compared the age of participants within experiments, but no published studies have been found that looked specifically at the effects of the sample age on visual demand and the effectiveness of the Occlusion technique as defined by a recent ISO standard.

For this work, 60 participants in various age ranges (from 17-76) were tested on four different IVIS tasks using a draft Occlusion protocol based on the ISO standard. These tasks were:

1. *POI* – Entering a Point of Interest (POI) using a PDA version of a Tom Tom navigation system.
2. *Address* – Entering a street address using a PDA version of a Tom Tom navigation system.
3. *Short scrolling* – Finding a 3 letter stock code from a scrolling list (one column) and reading out the accompanying share price.
4. *Long scrolling* – Finding a 3 letter stock code from a scrolling list (three columns) and reading out the accompanying share price.

The results showed some differences between the age categories for the two Occlusion performance measures (Total Shutter Open Time (TSOT), and Resumability (R)). In particular the older participants showed a greater spread of scores (especially for TSOT).

Regarding data sampling, there was no overall difference in either TSOT or R due to the precise sampling strategy used (either 10 random participants or stratifying a sample of 10 with two participants aged over 50 years). However, our results found using subjects aged over 67 increased the variability of the sample, so we recommend that they should be excluded in order to obtain reliable results from a limited sample.

Finally, the results found no significant performance improvements due to learning beyond the ISO Standard-specified practice. From the findings obtained it does not seem that learning is occurring through the five trials on any task, and that the practice sessions required by the standard are enough for a participant to reach the top of the learning curve. Likewise, no significant differences were found between those participants who undertook the Occlusion trials first, and those who did them second (i.e. after non-Occluded trials). As such, our results imply that the practice sessions and experimental order outlined in the ISO standard are sufficient (at least for the range of tasks considered), and therefore they were included in the Occlusion protocol developed in Work Package 4.

Overall, these results imply that to obtain minimal inter-subject variability an experiment should ideally use younger/middle aged participants. Based on these results, work Packages 3 and 4 used these age ranges (and especially avoided the 67-76 age group). Gender imbalance may be maintained for specific reasons (for example, to reflect a population- such as truck drivers being mainly male), but this does not appear to be a significant factor in the Occlusion results. The Occlusion protocol written in WP4 used the results obtained here, in particular regarding both the use of older participants and using the experimental instructions that were developed in this study.

Work Package 2 - Consideration of IVIS tasks and expert evaluation

Work Package 2 considered the results of the four different IVIS tasks from WP1 with reference to other data sources. The purpose was to help understand the characteristics of IVIS tasks that are overly distracting. Such an approach assisted in later stages of this project to help establish external benchmarks for IVIS tasks using the Occlusion protocol.

Three additional sources of information were used:

1. Expert opinion/heuristic evaluation: a review of the interface characteristics of the tasks. By using a panel of subject matter experts, we examined the likely usability issues and demands (perceptual, cognitive and physical) from these tasks while driving, and uncovered the different types of distractions associated with each of the tasks.

2. A keystroke level task analysis (a fundamental decomposition of the four IVIS tasks to quantify the time theoretically needed to complete them).

3. Social acceptability questionnaires (to gauge whether drivers thought that undertaking any of these four tasks whilst driving was acceptable). By conducting two surveys (one specifically for the four IVIS tasks, and one covering a wider area), quantifiable data were produced about which tasks, and which components of the tasks, are considered to be acceptable to UK drivers.

These three sources produced a rich array of results and ultimately helped the overall objective of developing a Demand Reference Level criterion using the Occlusion protocol. In particular:

- The expert review and social acceptability results showed from first principles the characteristics of the four tasks that are overly distracting and/or may cause usability difficulties - these results both helped understand and predict obtained Occlusion scores, and helped establish what particular features of the task make it acceptable/unacceptable to the UK driving population (as evidenced in the social acceptability scores). When setting limits (in WP5), the list of positive and negative task characteristics can help explain why a new IVIS task has obtained its specific TSOT and R scores and how socially acceptable it may be. Such an approach may be useful for policy makers and researchers to understand obtained Occlusion scores, and possibly it could also be used to assist the design process.

- Likewise, the Keystroke Level task analysis showed that the theoretical task completion times (both Static and Occluded) were closely linked to the results obtained from WP1 for two of the four IVIS tasks. This does not, of course, directly imply that only a Keystroke Level task analysis is sufficient to demonstrate which IVIS tasks may be acceptable, as it was not a good predictor for the other two IVIS tasks. However it does help support, augment, and explain the obtained results. In particular, the Keystroke Level analysis highlighted the fundamental characteristics of tasks that contributed to differences in TSOT and R values.

Overall, the findings of this Work Package show what aspects of tasks may be distracting/lack usability, how the results relate to theoretical task performance time and what a large sample of the UK driving public would consider as "acceptable" tasks to be performed while driving.

Work Package 3 - Benchmarking IVIS tasks to another impairment inducing factor (alcohol) and performance measures (Lane Change Task)

The third Work Package (WP3) continued the Occlusion project by examining IVIS tasks using an alternative evaluation method (the Lane Change Task - LCT) and comparing results with known driving impairments (alcohol).

In particular, WP3 investigated the variability between drivers using the four IVIS tasks from the previous Work Packages. It used the LCT to compare impairment due to IVIS and impairment due to alcohol intoxication at the UK legal driving limit (80mg of alcohol per 100ml of arterial blood).

The LCT results found that driving whilst engaged in all four of the IVIS tasks was significantly worse than driving at the UK legal alcohol limit. Additionally, compared with single task performance (i.e. purely an IVIS task), there was an increase in the time taken to complete the IVIS tasks when performed in conjunction with the driving task. When compared with Occlusion task performance (from WP1), the LCT was not able to discriminate between the four IVIS tasks in relation to their visual demand and potential impact on driving.

By comparing the results obtained from these aspects to those found from Occlusion they helped to further clarify what would be an appropriate Occlusion Demand Reference Level, and showed the sensitivity of the Occlusion technique compared with the Lane Change Task.

Work Package 4 - Production of a detailed protocol for assessment of IVIS tasks, performing a reliability study of Occlusion procedure and testing the influence of a primary loading task

Work Package 4(a) – Protocol Development

The ISO standard for Occlusion had just been published, but we took an independent view of all of its requirements. Some aspects (e.g. precise experimental instructions) needed to be extended and others (e.g. the age groups suggested) needed to be modified for an effective protocol.

The results of WPs1-3 were used to make the protocol evidence-based. It was developed based on the latest Occlusion standard and supplemented with results from the first three work packages and our practical experience of applying the Occlusion method. The development process was iterative, whereby comments received by TRL from other organisations (including University of Nottingham, UK and BASt, Germany) were used to further refine the protocol.

The Occlusion protocol (Horberry et al., 2007: PPR 259) takes the form of a detailed method that contains all the necessary information for replication.

Work Package 4(b) – Reliability Study

A reliability analysis study was conducted to evaluate the protocol being developed. The protocol was used at an independent research centre (University of Nottingham), and the results were compared to TRL data.

Comparing the results achieved at University of Nottingham and TRL provided a clear indication as to the reliability of the protocol. Twenty participants were tested at the University of Nottingham using the most recent version of the Occlusion protocol; these data were then compared to earlier results from TRL. The same four IVIS tasks were performed at both Nottingham and TRL.

The results showed that there was considerable agreement between the Nottingham and TRL results for Tasks 1 and 2 (both using a navigation system) for both TSOT and R. This was encouraging as the primary use of the protocol is likely to be in the assessment of fully developed systems. There was less agreement for the two scrolling text tasks (prototype tasks). It was established that differences in scrolling rates contributed to the findings (both across the studies and across the two tasks). This indicated that the protocol should require

collection of information regarding the technical setting in which a task is implemented. This will better enable meaningful cross-study comparisons and the requirement (operationally defined in terms of performance speed) was added to the revised version of the protocol.

The key result was that the protocol was successfully evaluated and was shown to be reliable. For tasks that are fully developed it produced comparable results across laboratories. For prototype tasks the protocol was extended to ensure that the technical environment in which the testing takes place is better specified and more precisely controlled.

Work Package 4(c) – Primary Loading Tasks and Occlusion Performance

This work examined the effects on Occlusion parameters of participants undertaking a primary loading task i.e. an additional task during the occluded (blank) periods. It has sometimes been argued that Occlusion has limited validity and does not necessarily replicate the differing demands of performing an IVIS task whilst driving a real vehicle since the participant must attend to the driving situation in the periods between doing the IVIS task. This raises the question of whether a simulated driving task or "primary loading task" is necessary during the Occlusion interval. The aim of this work package was to determine whether or not the Occlusion measure should be accompanied by a simulated driving task or "primary loading task".

The study investigated the effects of two different primary loading tasks on Occlusion parameters. Ten participants were involved; they completed the four IVIS tasks under the following conditions:

- With no loading task.

- With a visual-spatial task (a visual search task on a separate screen).

- With an auditory task (holding auditory information in working memory and comparing it with new auditory information).

The study showed that the presence of a primary loading task whilst performing an IVIS task does modify the Occlusion results. However, the findings paint an unclear picture; in general the presence of the loading task produced an unwelcome variance in participant performance. For some tasks (the two navigation system tasks) performance was reduced (as expected) based on the Occlusion metrics. However, for the two scrolling tasks, participant performance presented a very indistinct pattern.

It was concluded that introducing a loading task has few beneficial effects in terms of Occlusion performance (and generally creates more undesirable inconsistencies). Consequently, using a loading task during an IVIS evaluation with the Occlusion technique is not recommended, and therefore the protocol was not modified to include the use of such a task.

Work Package 5 - Setting Limits: To determine performance limits based on the Occlusion measures

The objective for this final WP was to propose a clear "benchmark" or Demand Reference Level (DRL) that can be used by DfT (and others) to identify in-vehicle devices, tasks or functions that involve an unacceptably high level of demand if used while driving. It used a wide range of data sources enabling many viewpoints to be integrated; these included scientific, social acceptability and commercial views. These included:

- A meta-analysis of the literature review data especially focusing on the 'limits' used in other countries.

- Personal contact with Occlusion experts worldwide to try to reach an international consensus and to supplement the review of literature.

- Experimental Occlusion work conducted by TRL and Nottingham concerning age.

- Social acceptability/keystroke level analysis/likely usability issue findings.

- A comparison of impairment due to visual demand from using an IVIS with impairment caused by alcohol intoxication at the socially agreed UK limit.

- Results of reliability analysis and primary loading task investigations.

- Knowledge of the way industry uses design guidelines.

The developed DRL is based on the Occlusion parameter TSOT. For an IVIS task to 'meet' the DRL, the TSOT value (mean + 'spread' combined, termed DL_{Occl}) needs to be below 8 seconds.

The DRL was designed, and described in this report, in a manner so that it can be easily applied and understood by other researchers, practitioners, designers and policy makers.

Although meeting the DRL does not imply an IVIS task is 'safe' to perform whilst driving, it does suggest the visual demand required is within a benchmark limit. Conversely, generally tasks that do not meet the DRL involve an excessive level of visual demand; however, there might be situations where such tasks are still acceptable for other reasons (such as long tasks that are easily interruptible and impose only a low intensity of visual demand). In such cases, the justifications about why such tasks are acceptable should ideally also be evidence-based.

Overall outcomes

Taken as a whole it is argued that the project has successfully collected and analysed a great deal of valuable Occlusion-related data. It undertook the largest Occlusion study to date in the world (the age experiment reported in WP1). The work has received a great deal of international attention (for example, invitations to speak about the research at international events) and has helped promote use of the Occlusion method. A practical protocol has been developed and a Demand Reference Level has been defined by the research; it is believed that using both the protocol and DRL criteria will help identify IVIS tasks that require excessive visual demand, and if widely implemented will lead to improvements in road safety. The criteria developed here could also be incorporated into other, broader rating approaches (such as star ratings) for In-Vehicle Information Systems.

It is now recommended that steps are taken to disseminate the findings and protocol widely (this might be accomplished by the publication of an ISO technical report). Furthermore, once the DLR has been widely disseminated, provision should be made for periodical review. Finally, it is recommended that further data is collected to benchmark the visual demand for a wide range of IVIS tasks against the proposed DRL. Such research could assist DfT in understanding the range of interface types that could be considered to be unacceptable.

Abbreviations

ANOVA	Analysis of Variance (a statistical analysis method)
BAC	Blood Alcohol Concentration
df	Degrees of Freedom (used in the statistical analysis)
DRL	Demand Reference Level
DL_{Occl}	Demand Level Occlusion. This equates to the mean Total Shutter Open Time plus the 'Spread'. i.e: Mean $+ \frac{(85^{th}\ percentile-\ mean)^2}{mean}$
ESoP	European Statement of Principles on the Design of Human Machine Interaction
ISO	International Organisation for Standardisation
IVIS	In-Vehicle Information System
LCD	Liquid Crystal Display
LCT	Lane Change Test
MDev	Mean Deviation from a optimal model of driver lane keeping
MS	Mean Square (used in the statistical analysis)
Nomadic Device	A handheld wireless device such as a PDA or smartphone
Occlusion Interval	Time during which the driver interface is not visible when using an occlusion procedure
Outlier	Observation that lies outside the overall pattern of the sample data distribution
p	Probability
PDA	Personal Digital Assistant
POI	Point Of Interest
PLT	Primary Loading Task
PLT-VS	Visual Primary Loading Task (condition 1)
PLT2-AS	Auditory Primary Loading Task (condition 2)
POI	Point Of Interest
R	Resumability ratio (refer to page 10 for definition)
t	t-test statistic (a statistical analysis method)
Spread	$\frac{(85^{th}\ percentile-\ mean)^2}{mean}$
Task	Process of achieving a specific and measurable goal using a prescribed method
Trial (Test Trial)	Investigation of one participant undertaking one repetition of one task
TGT	Total Glance Time
TIVT	Total Interrupted Vision Time
TSOT	Total Shutter Open Time
$TSOT_{AS}$	Total Shutter Open Time under Auditory Spatial Task conditions
TTT	Total Task Time (refer to page 10 for definition)
TTT_{Occl}	Total Task Time under occluded conditions (refer to page 10 for definition)
TTT_{Unoccl}	Total Task Time under unoccluded conditions (refer to page 10 for definition)
Vision Interval	Discrete time during which the driver interface is visible when using an occlusion procedure
Visual Demand	Amount of visual activity required to extract information from an interface of an in-vehicle system to perform a specific task
Visual Occlusion	Measurement method involving periodic obstruction of the participant's vision or the obscuration of visual information under investigation
WP	Work Package

1   Introduction

Over the past few years there has been an increased uptake of in-vehicle information systems (IVIS) such as satellite navigation and communication devices. Although there are certainly many positive potential benefits from such systems, there are also some negative safety issues - these include the possibility that in-vehicle systems may distract drivers, increase their workload or encourage them to engage in non-driving related tasks during their actual driving (Horberry et al., 2006).

There is a variety of measurement methods available to assess the potential negative effects of such new in-vehicle technologies; one method that is becoming increasingly important internationally is 'Occlusion'. Occlusion assesses visual demand due to the use of visual or visual manual interfaces available to the driver while a vehicle is in motion. The Occlusion method determines visual demand and interruptability of a task by intermittent viewing of the in-vehicle system, generally by means of specially designed goggles. It should be noted that visual demand from an IVIS task does not necessarily imply driver distraction; the main intervening factor is the driver's motivation to engage in the task (Burnett et al, 2005).

Visual Occlusion helps identify designs which require long single glance durations by drivers to assimilate information necessary to complete a task using an in-vehicle system. Occlusion research generally examines two parameters: *Total Shutter Open Time* (TSOT: how long it took a participant to complete a task under conditions of Occlusion) and *R* (a measure of task resumability- assessing how 'chunkable' an in-vehicle task is). Definitions of these primary Occlusion metrics are given in Section 3.4.

The overall project goal was to develop a robust distraction measurement protocol based on Occlusion, and to propose criteria for the acceptability of different systems. The work is based on, and supports, the recent ISO standard (ISO 16673: 2007). It is anticipated that this report will ultimately provide evidence to help inform policy regarding acceptable and unacceptable In-Vehicle Information Systems.

This Occlusion research project was undertaken in five separate work packages; these are detailed in Sections 3-7 of this report. A literature review is provided in Section 2.

WP 1: Age (Section 3)

WP 2: Consideration of different IVIS tasks, and expert evaluation (Section 4)

WP 3: Benchmarking of different IVIS tasks to other impairment inducing factors (alcohol) and performance measures (Lane Change Task) (Section 5)

WP 4: (a) Protocol Development, (b) Reliability Study & (c) Primary Loading Task (Section 6)

WP 5: Setting Limits: To determine performance limits on the Occlusion measures for acceptable IVIS tasks (Section 7)

The final Occlusion protocol is available as a final report (Horberry et al, 2007).

2   Review of Recent Occlusion Literature

In 2004 a literature review of Occlusion research was completed for DfT, with particular emphasis on the Occlusion technique/methodology (Stevens et al., 2004). The review here updates that previous work looking particularly for age effects.

2.1   Literature Review Method

Data was sought and collected from a variety of information sources - published literature, research presented at the ISO 'Occlusion Standards Meetings' and through contacts with organisations who have undertaken recent Occlusion research (e.g. Transport Canada).

European, North American, Australian and Japanese transport researchers who have completed Occlusion studies, or have a known interest in, Occlusion, as well as several vehicle manufacturers were contacted by e-mail to inform them of our research study and to request any recent literature that may not yet have been published.

An internet search of Occlusion terms was also carried out to obtain additional details of recent research.

A number of research reports and conference presentations were received in response to our queries that update the previous review. The results of this search are summarised below; they are presented in terms of the themes they addressed.

2.2   Literature Review Findings

2.2.1   Age of Participants

The age of participants is likely to be an important (perhaps the most important) individual difference affecting the validity and reliability of results from Occlusion measurements. The number of participants required in an Occlusion study has now been defined by ISO 16673; however, no requirement is stated concerning the age ranges of the participants. The standard simply makes a recommendation that 20% of subjects for Occlusion studies should be over the age of 50.

There have been several studies examining the effects of age on tasks performance under conditions of Occlusion; however, these have produced a somewhat mixed picture of results. Research by Weir et al. (2003) tested a group of younger (20-28yrs old) vs. older (65-71yrs old) drivers. Their research reported that older drivers were likely to take longer glances at the display and took longer to complete the required task.

Around the same time, Bengler (2003) tested 30 subjects, 20 of which were aged 19-30 and 10 of which were aged 48-63. Subjects completed 10 in-vehicle tasks, the results of which showed that there was a significant difference in Total Task Time (TTT) between the younger and older groups of drivers. The main study effect was participant age with the older drivers taking significantly longer to complete the tasks. Bengler also calculated an R value for his participants, and the age of the participants was again found to explain significant differences in R values.

More recently, Asoh et al. (2005) compared a group of older male drivers (62-71, average age of 65.2) to a group of younger male drivers (26-52, average age of 34) to investigate differences in total glance time (TGT) and total shutter open time (TSOT). The results showed that the TGT was lower for the older drivers.

2.2.2   Means of Occlusion

There are two widely used means of visual Occlusion for driving task demand research studies. The first, and more common method, is to use Translucent Technologies Plato Occlusion Goggles (see Figure 1).

Figure 1: Participant wearing the Plato goggles

Plato goggles use LCD lenses that can switch between an 'on' state (translucent) to 'off' (opaque) and are regarded as one of the simplest mechanisms for undertaking occlusion experiments.

The second method is to use screen/device blanking using software to occlude a display for set periods of time. However, unless the task is being carried out wholly on a touch screen, this method means that any device controls remain on view during the task.

Two recent studies have investigated the use of two different means of occlusion: Weir et al (2003) and Niiya (2000).

The Weir et al study found no significant differences in task time or driving performance between the two methods. The earlier study by Niiya did, however, report slower driver response times during an object recognition task when using goggles. It was reported that this difference was due to accommodation time of the display by the driver after the Occluded period.

Given that most Occlusion studies around the world have used the Plato goggles, this research has also made use of them. This allows a direct comparison to be made between the results of the new work reported here and those obtained elsewhere in previous studies.

2.2.3   The 'R' Value

The R value has recently been discussed by representatives of Transport Canada (Burns, 2006). Also, R results previously obtained have been compared to those obtained by the Lane Change Task (LCT) (Mattes, 2003).

Transport Canada were concerned that R does not discriminate between tasks thought to be acceptable and tasks that are thought to be unsafe whereas the measure of TSOT does. They conclude that TSOT findings are more consistent with LCT results.

They reported that it is unclear what R measures:

  • R > 1 suggests problems with resumability… but what does that mean for safety?

  • R < 1 does not mean a task is safer.

  • R did not discriminate by task complexity.

  • Measuring TTT to calculate R increases trial demands.

  • R alone is not sufficient and may be unnecessary or redundant to TSOT.

  • R ignores task duration and could be used to justify exceptionally long tasks.

  • R does not capture the other features of distraction (i.e., intensity, timing or frequency).

Therefore, Transport Canada believes that R can be used to support TSOT findings but never as a stand-alone measure (Burns, 2006).

However, some researchers take a different view, and have data to support the diagnostic value of R (Pettitt et al, 2006); so they argue that R values should be considered when evaluating IVIS systems using Occlusion.

2.2.4   Setting Limits

Asoh (2005) investigated the allowable upper limits of the JAMA guidelines for Total Glance Time towards a navigation system.

Asoh compared a range of younger and older drivers; he measured the relationship between TGT and anxiety of using a navigation system whilst driving. The correlation between the two was high for all conditions and anxiety increased and TGT increased. It was estimated that a TGT of less than 8 seconds did not cause anxiety or affect vehicle behaviour regardless of the type of road or navigation system.

Baumann (2004) reported and carried out experiments which built on a study by Tijerina (2000) which compared the Occlusion technique with the Green (1999) 15-second rule for Total Task Time for assessing whether a task is executable whilst driving. However, Tijerina argued that the 15 second total task time rule does not take into account how well the task can be completed in small chunks of time. The results showed that even though a task may be completed within the 15 second rule period, because of a task's chunkability it may still not be possible to carry it out whilst driving with short (< 2 second) glances. The Occlusion technique, however, acts as a method of assessing whether this is the case and so, it is maintained here, can identify unsuitable tasks.

In terms of cut-off limits, the Driver Metrics report (2007) described the values currently used around the world:

TSOT:

20 s (85th Percentile). *Society of Automobile Engineers* (International Organization) J-2364.

15 s (Mean). The *Alliance of Automobile Manufacturers*. This is a US based coalition of 9 car and light truck manufacturers, including BMW Group, DaimlerChrysler, Ford Motor Company, General Motors, Mazda, Mitsubishi Motors, Porsche, Toyota and Volkswagen.

7.5 s (Mean). *The Japan Automobile Manufacturers Association* (JAMA). This is a Tokyo-based trade association representing 14 Japanese car, truck, bus and motorcycle manufacturers.

R:

To be decided, but proposed < 1

2.2.5   Driver Metrics Report

A Driver Metrics Workshop was held October 2-3, 2006, in Ottawa, Canada to bring together a key group of global experts in driver performance research. The central issue addressed in the meeting, was which driver performance metrics can be employed early in product design to assess visual demand before more complex driving simulations and evaluations occur. Such early off-road/off-simulator measures can help driver interface designers understand how a new display/control concept might work before they engage in more complex, costly testing. Visual demand is the basis for many advanced telematics, IVIS and other infotainment devices - and thus is the focus of most attention demand research (Driver Metrics, 2007).

Four groups of metrics were considered:

1. Naturalistic driving and crash data.

2. Occlusion methods.

3. Lane change task.

4. Direct measurement.

In terms of Occlusion, the report stated:

"The Occlusion procedure simulates visual demand in order to measure "resumability" after a visual distraction. The longer someone is visually distracted (i.e., keeps their eyes off the road), the greater the crash risk. The Occlusion method is accomplished by blocking the participant's view of the relevant driving scene area or cockpit areas either physically (e.g., Occlusion goggles, partition) or by simulating a partition in a simulated environment for a relevant Occlusion interval. The resumability ratio is determined as the total time that vision is not occluded (TSOT) to the total static task time (TTT_{Unoccl}), i.e., $TSOT/TTT_{Unoccl}$ (ISO/DIS 16673.2).

One metric that received unanimous agreement in the Driver Metrics Workshop was Total Shutter Open Time (TSOT). Higher TSOT means a task requires longer visual attention. In the attached simplified matrix, TSOT has been shown to be particularly effective in detecting interface limitations related to diminished longitudinal control (0.9 R squared). On the other hand, the group was unable to come to a true consensus on criterion values for the resumability ratio ($R = TSOT / TTT_{Unoccl}$), except to agree that values below 1.0 suggest the task affords blind interaction while tasks above 1.0 exceed the typical time limits for blind interaction.

The Occlusion method can be used as a surrogate to implementing an actual secondary or primary driving task, but should always be acknowledged as an approximation of the actual distraction task load. Results can be used to determine the relative distraction of adding a secondary interface to the vehicular environment. The methodology allows for reliable and objective viewing times with low experimenter effort and cost in a lab or vehicle setting. It can measure visual control during participant- or experimenter-based presentation times. Finally R may have diagnostic value in determining the interruptability of a task.

That said, the Occlusion method is not sensitive in combination with short, auditory, or purely manual tasks. The true utility of R is still under examination, especially in terms of driving safety, as the values of R do not capture task intensity, timing, or frequency. Also, the definition of TSOT may vary with task complexity, depending on a particular implementation's duration or definition.

Workshop attendees identified a desire for new methodologies and options other than Occlusion goggles. They also would like to include a measure of cognitive demand control over the Occlusion interval metrics. Participants also reported a problem of system response delays in collecting Occlusion data that needs to be worked out, including determining an acceptable threshold for the response time of opening/closing the Occlusion device mechanism.

While there was strong agreement on TSOT as a metric, there was a considerable range of discussion put forth by participants. Further work is required to determine when a consensus can be reached on criterion values. The tentative agreement of the attendees was that issues with TSOT should be dealt with first, especially those dealing with the differing criterion values (e.g. 7, 15, and 20 seconds).

*As for resolving the R metric differences, there will be difficulties in achieving a consensus. The group agreed to disagree." (*Driver Metrics draft report, pp 5-6, 2007.)

2.3   Key Implications

Overall, not a great deal of Occlusion research has been published in the open literature in the past three years, despite the topic being extensively researched. Speaking to professional contacts overseas, quite a large amount of unpublished research has been performed in this general area[*]. Now that the Occlusion standard has been produced it is anticipated that more research will be published.

For the purpose of this project, the main points are:

- It is clear that several studies have found contradicting results when testing older and younger participants. Given that the ISO Occlusion standard makes only limited mention of age, it is clear that further work is needed to investigate the differences across an age range. This will be undertaken in WP1.

- Furthermore, little research has tried to compare results with existing guidelines and standards. Asoh seems to have been one of the few to have tried this in comparing his results to JAMA guidelines. Similarly, Transport Canada tried to compare the Occlusion results to the lane change task results, and to lateral lane deviation. In terms of cut-off criteria, the driver metrics report (2007) stated:

    "…finally, a limit needs to be stated. For TSOT we have 8 seconds, 15 seconds and 20 seconds, as 'lines in the sand."

- Finally, little work has been done to undertake Occlusion meta-analyses or to develop and agree benchmarks (for example, three different ones are mentioned in the quote above for TSOT). In part this is because of differences in experimental measurement procedures. Therefore one of the main objectives of the work in WPs 2-5 was to develop Occlusion criteria values.

---

[*] Many organisations who have undertaken Occlusion research are motor vehicle manufacturers; for commercial reasons, they are often not able to publish their work in the open literature.

3   Work Package 1- Age Experimental Study

3.1   Background

The age of participants is likely to be an important individual difference affecting the validity and reliability of results from Occlusion measurements. As seen in the literature review above, previous research has not considered age directly in relation to the utility of the ISO standard itself.

At present, it is quite possible that researchers or developers using the ISO protocol independently with varying samples based on age but the same systems/tasks would come to markedly different conclusions.

3.2   Study Aims

The study's *primary* aim was to establish the age range needed in the protocol for minimum variability (and therefore minimum participant numbers). As part of this, the study identified how performance values (TSOT and R) vary with age for a range of IVIS tasks. In this respect, we anticipated that a hypothetical graph of the results might be as follows:
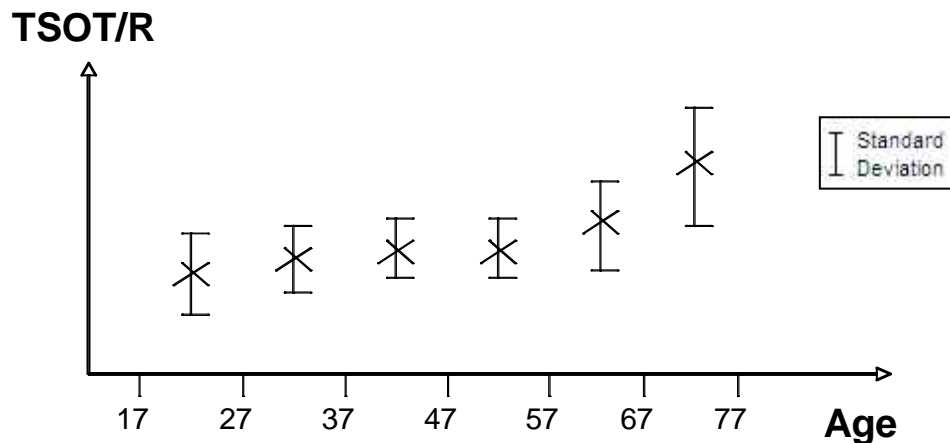

Figure 2: Hypothetical distribution of TSOT/R by age

It was important to compare both TSOT and R across tasks for different age groups to examine whether there are consistent significant differences. For example, if TSOT for Task 1 is significantly greater than TSOT for Task 3 with the 17-26 age group, is this also the case for all other age groups?

The final aim was to test the current ISO version of the Occlusion standard to ascertain the effect of incorporating 20 % of older drivers within the total participant group.

3.3   Method

3.3.1   Participants

Sixty participants were recruited in six age categories (17-26; 27-36; 37-46; 47-56; 57-66; 67-76) with ten in each group. Gender was balanced across the groups. All participants held a valid driving licence.

3.3.2   Tasks

Four different in-vehicle tasks of differing difficulty were evaluated. Such an approach allowed consideration of the impact of variability within varying age ranges, and how the effects of age differ across the different tasks.

Based on an emerging view of design targets presented in the literature review above (for TSOT less than 15 seconds and R less than 1), four tasks were chosen to ensure a range of values for TSOT and R. The approach gave a 2 by 2 framework for the four tasks to assess. The actual tasks chosen were based on earlier pilot studies that demonstrated that the TSOT and R measures were likely to be in the appropriate ranges (see Table 1).

Table 1: The four different in-vehicle tasks

	TSOT expected to be <15 secs	TSOT expected to be >15 secs
R expected to be < 1	**Task 1 – Acceptable?** Destination entry by Point of Interest using a commercial route navigation system.	**Task 2 – Acceptable/ Unacceptable?** Destination entry by address using a commercial route navigation system.
R expected to be > 1	**Task 3 –Acceptable/ Unacceptable?** Short reading task: Finding share prices from a scrolling list (1 column) presented on an in-vehicle display.	**Task 4 – Unacceptable?** Longer reading task: Finding share prices from a scrolling list (3 columns) presented on an in-vehicle display.

The tasks where R is expected to be <1 are relatively easy to establish, as there are many reported in the literature. Finding tasks where R is likely to be >1 is more difficult, largely because of the nature of the R ratio where potential problems with resumability may be offset by opportunities for "blind operation" i.e. continued operation without sight of the device. It was therefore felt that predominately visual tasks are required – from previous studies, the only tasks where R values >1 have been reported have required the reading of text (either static text or scrolling text). For such tasks, it is important to give participants a realistic goal associated, to identify key information from the text. The longer reading task (Task 4) involves the automatic scrolling through several screens to locate information.

3.3.3   Procedure

The procedure was based largely on the current version of the Occlusion standard. In addition, a pre-trial questionnaire was completed by participants in order to record information about: age, gender, number of years driving experience, average annual mileage, and about their views on technology. Such information was used to analyse inter-individual effects.

Participants were tested individually in a normal road vehicle (Vauxhall Astra). The participant sat in the driver's seat and the experimenter was seated in the passenger seat. The in-vehicle devices were positioned in the centre console of the vehicle. Figure 3 shows the experimental arrangement.

Figure 3: Occlusion experimental arrangement

Before the experiment, the instructions were standardized (written explicitly so delivered in the same way to each participant) and were presented orally. The display and controls of the interfaces were visible during instruction. An instruction was repeated at the request of the participant.

For each of the four tasks, the participants were shown how to use the device in the vehicle and then asked to complete five training sessions with the device. The test sessions then began and participants completed five Occluded and five 'Static' (non-occluded) attempts at each task. The order of Occluded/Static attempts and the order of the four tasks were randomised to minimize any order effects. During the Occlusion interval, neither the interface displays nor controls were visible, but 'blind' operation of the controls was permitted.

Each participant was instructed to attend to the task in a continuous manner. The specific data to be viewed or entered for each of the five test trials was unique but representative of the level of difficulty for each task. As such, each task was of equal difficulty and involved an identical number of steps.

During the Occlusion task the participants wore the goggles. Following the ISO standard, the vision interval was 1.5 s and the Occlusion interval was 1.5 s. Periods of vision and Occlusion occurred automatically without interruption until the task was completed or the trial terminated. Thus, the pacing of Occlusion intervals was controlled by the system, rather than the participant.

3.4   Calculating the Occlusion Measures (TTT, TSOT and R)

A full definition of terms and a detailed description of how the Occlusion measures were calculated are provided within the 'Occlusion protocol' (Horberry et al., 2007: PPR 259). Brief definitions of the main Occlusion metrics of Total Task Time (TTT), Total Shutter Open Time (TSOT) and R (Resumability ratio) are provided below:

TTT_{unoccl} Definition. The total task time unoccluded (TTT_{Unoccl}) shall be determined as follows:

> Start: Timing starts at the end of the task instruction.
> End: Timing ends when the instructed task has been completed and the participant says he or she is "done".
> Duration: Tasks are timed from start to end without interruption, including errors.

TTT_{occl} Definition. The total task time in Occluded conditions (TTT_{Occl}) is as follows:

> Start: Timing starts with the beginning of the first vision interval.
> End: Timing ends when the instructed task has been completed and the participant says he or she is "done".
> Duration: Tasks are timed from start to end without interruption, including errors.

TSOT Definition. The total open shutter time (*TSOT*) is as follows:

> Start: Timing starts with the beginning of the first vision interval.
> End: Timing ends when the instructed task has been completed and the participant says he or she is "done".
> Duration: Tasks are timed from start to end without interruption, including errors. The task time (TTT_{Occl}) is divided by the Occlusion cycle time (shutter open time plus shutter close time). The result is then multiplied by the shutter open time. This determines the time which the shutter was open during the task.

R value Definition

> The R value or resumability ratio, is the ratio of the duration of the total shutter open time (*TSOT*) to the total task time unoccluded (TTT_{Unoccl}), i.e., *TSOT/$TTT_{Unoccl.}$*

3.5   Results

3.5.1   TSOT and R Results

The most important analyses concerned how Occlusion performance varied within and between different age groups for each of the four different tasks. Table 2 below is a summary of the Occlusion results:

Table 2: Mean TSOT and R values for Tasks 1 to 4

Task	Mean TSOT	R (Mean of ratios)[†]
Task 1	7.63	0.85
Task 2	11.77	0.81
Task 3	10.94	0.78
Task 4	11.62	0.87

As seen above, R (and to a lesser extent TSOT) did not vary greatly between the four tasks. This is despite pilot testing which found a bigger variation. However, as will be detailed later, the relative differences between the tasks on these two Occlusion measures were greater for some age categories (in particular the older age groups).

The mean values for both TSOT and R were calculated and then split by age category and task performed. Subsequently, the standard deviation of the data were also calculated. The results below are present first for TSOT and then for R.
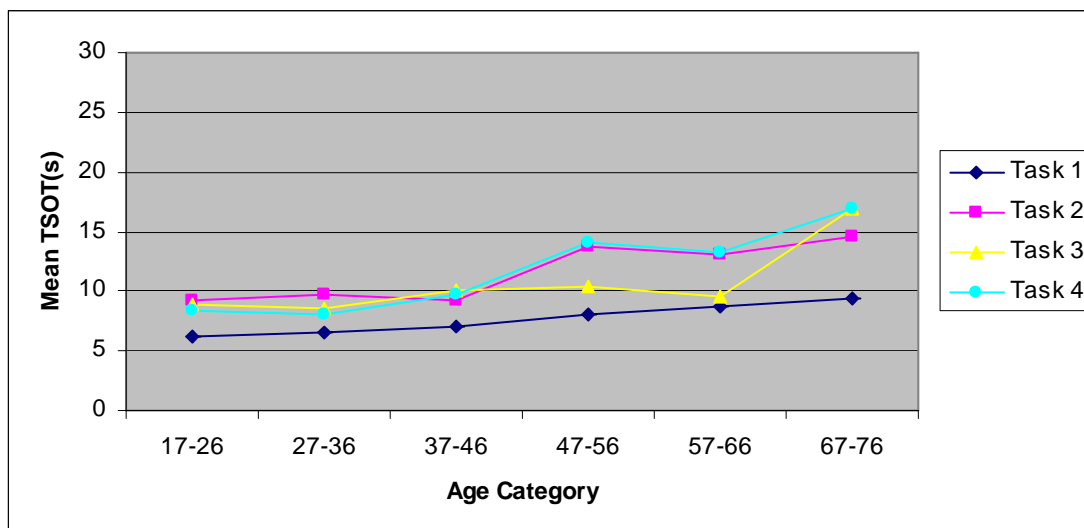
TSOT vs. Age Category

Figure 4: Tasks 1-4, TSOT against Age Category

[†] R is calculated by TSOT/TTT_{Unocc} When calculating 'R' for the whole sample of 60 participants a method of 'mean of ratios' was used. The R value for each participant (based on the mean of each of their 5 performances) was calculated and then the mean of all participants' R values was determined. For the purposes of this work this is considered a preferred alternative to calculating the 'ratio of means' i.e. mean TSOT/mean TTT_{Unocc}.

As Figure 4 above shows, Task 1 was the quickest to complete for all age groups, whereas performance on the other three tasks varied between different age categories. From first principles, Task 1 was hypothesised to be the least demanding; the results obtained here support this.

Also, in overall terms, task performance declines as a function of age. For example, compared with the other age groups, the age category 67-76 had the longest TSOT for all tasks.

The experimental hypotheses predicted the younger to middle aged categories would have quicker task performance than the older age groups (above 57 years). To some extent this pattern is evident in the 67-76 group, but less so for the 57-66 group. Overall, the hypothesised effect was partially found.

Another revealing metric is the spread of scores **within** each age group; Figure 5-Figure 8 below shows this for the four different tasks.
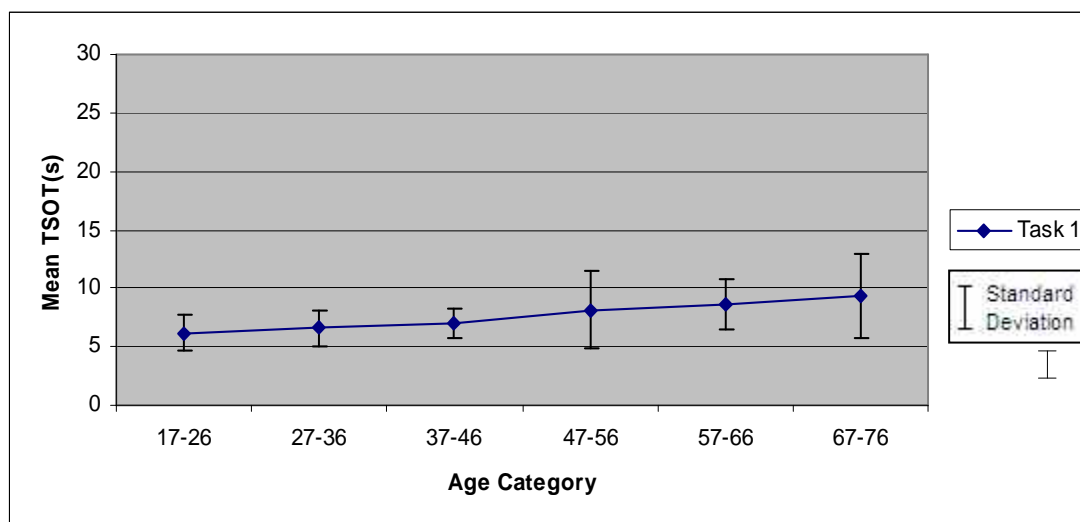

Figure 5: Task 1, TSOT against Age Category

The results for Task 1 (see Figure 5 above) were approximately as expected, where the older age groups 67-76 and 47-56 produced the largest amount of deviation compared with the other categories. Figure 5 also demonstrates an upward trend in TSOT scores as age increases: an ANOVA performed on these data revealed a significant difference ($MS = 15.78$, $df = 5$, $p = 0.031$) amongst the six age categories.
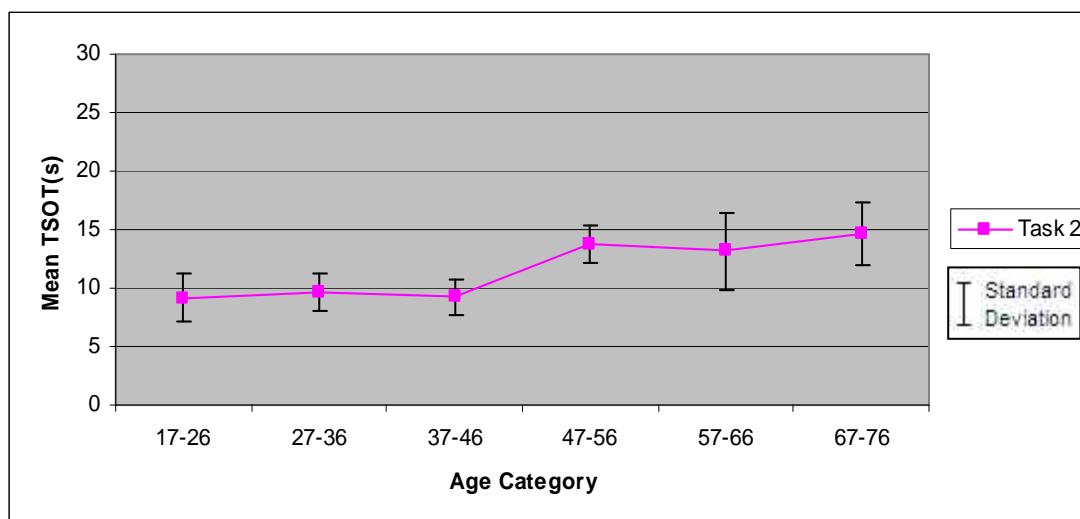
Figure 6: Task 2, TSOT against Age Category

Figure 6 above demonstrates the results of Task 2. These present a slightly mixed picture, where the highest spread of scores is in the age group 57-66, followed by the 67-76 group. An ANOVA was performed on the data for Task 2 and this produced a very significant result (MS = 56.05, df = 5, $p < 0.001$).

Further analyses (using a Bonferonni post-hoc test) indicated many significant differences between the age categories. These results can be summarised as dividing the age categories into two groups: the lower three age categories (17-46) and the older three age categories (46-76), with these two groups being significantly different from each other (with the exception of the 37-46 and 57-66 categories where the difference was not significant (p = .12)).

Therefore, older groups were both slower and more variable as compared to the younger categories for this task.

Figure 7: Task 3, TSOT against Age Category
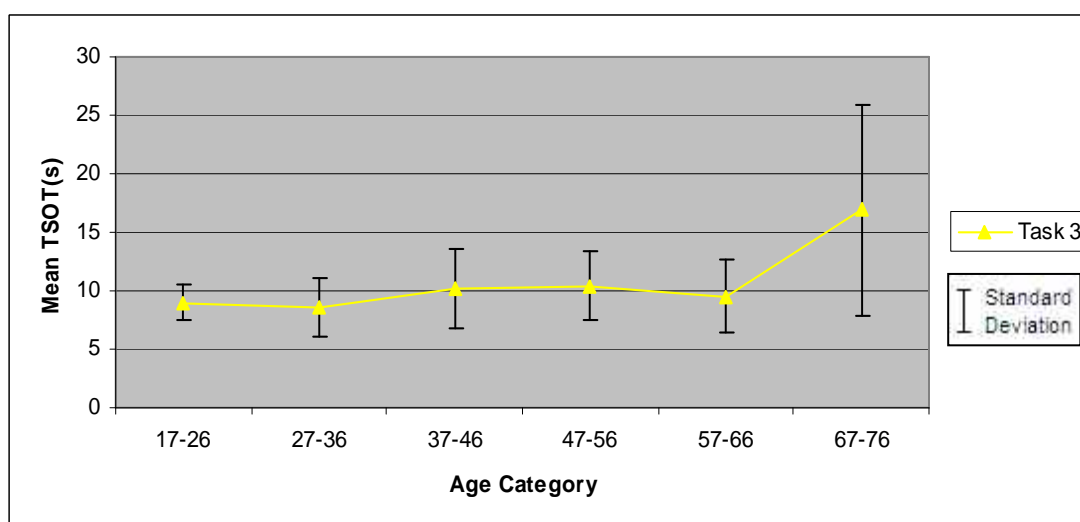
Task 3 (as reported in Figure 7 above) also shows scores as predicted, where the oldest group (67-76) had the highest spread of scores. The results of an ANOVA performed on these data demonstrated the presence of significant differences between the age categories (MS = 92.25, df = 5, p = 0.001). Again, a Bonferonni post-hoc was performed on these data which established that

the only age category which was significantly different from the others was the 67-77 age category.
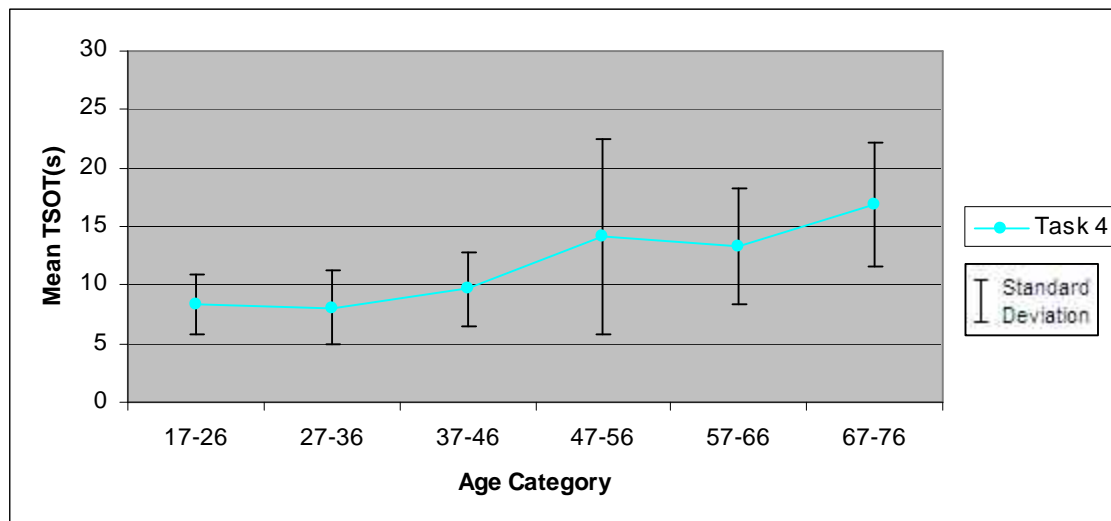

Figure 8: Task 4, TSOT against Age Category

Task 4 revealed a more mixed pattern, where category 47-56 had the highest spread, followed by the 67-76 group (see Figure 8 above). However, it must be noted that the 47-56 age category has larger error bars because two participants had unusually high TSOT scores; two to three times as high as most other participants.

To establish which of these differences were significant, an ANOVA was performed on the data, including Bonferonni post-hoc tests. The ANOVA reported the presence of a significant difference (MS = 121.12, df = 5, p = 0.001). The results of the post-hoc tests demonstrated that age category 67-76 has significantly higher TSOT scores than age categories 17-26, 27-36 and 37-46 (much the same as for Task 3). No other significant differences were found.

R value vs. Age Category

As with TSOT, the data were split per task. Overall, the pattern for R was reasonably similar to that obtained for TSOT, suggesting a slight general decline in ability to resume tasks both as a function of increased task difficulty, and higher participant age.
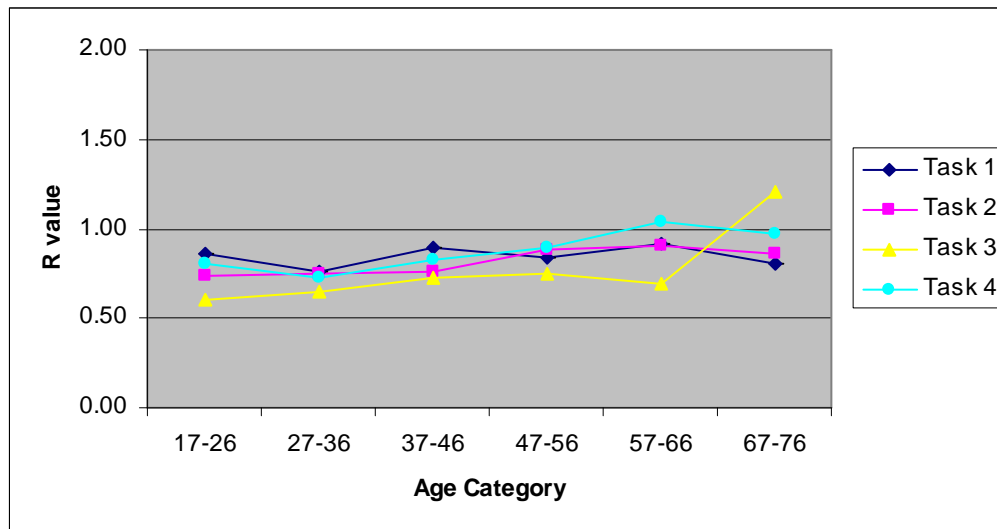
Figure 9: Tasks 1-4, R value against Age Category

As Figure 9 above shows, no single task had a consistently lower R scores than the other tasks across all the different age groups. Compared with the other tasks, Task 3 had the lowest R score across most of the age categories (so perhaps Task 3 can be said to be the task which is most easily resumed). Conversely, Task 4 had the highest R score of the four different tasks across many of the age categories (especially the older ones). From first principles, Task 4 was the most demanding task; the results obtained here support this.

In overall terms, there is not a huge difference in task performance with R as a function of age. However, there appears to be a slight trend towards greater R values with increasing age (especially for the more difficult Tasks: 3 and 4).

Table 3: R value by Age Category and Task

Age Category	Task 1	Task 2	Task 3	Task 4
17-26	0.86	0.74	0.61	0.81
27-36	0.77	0.75	0.65	0.73
37-46	0.89	0.76	0.73	0.83
47-56	0.84	0.88	0.75	0.89
57-66	0.92	0.90	0.69	1.04
67-76	0.80	0.86	1.20	0.97

An interesting point to note is that the majority of the mean R scores in Table 3 above (by age group and task) are less than 1. This point will be further discussed later in this report with respect to setting limits.

As with TSOT, another revealing metric is the spread of R values **within** each age group. The four figures below show this for the four different tasks.
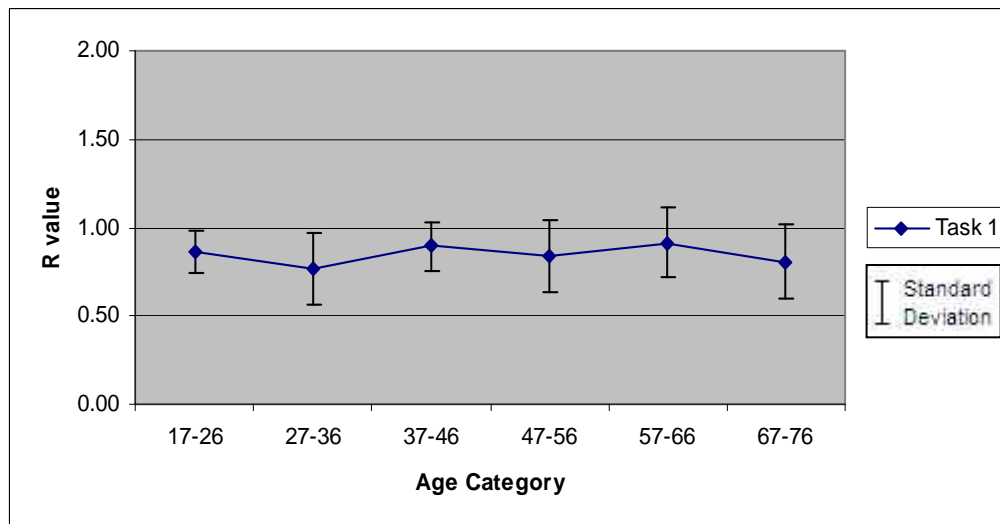
Figure 10: Task 1, R value against Age Category

As illustrated in Figure 10 above, the results for Task 1 showed no major differences between the age categories. This impression was borne out by an ANOVA which failed to produce a significant result ($MS = .031$, $df = 5$, $p = 0.79$).
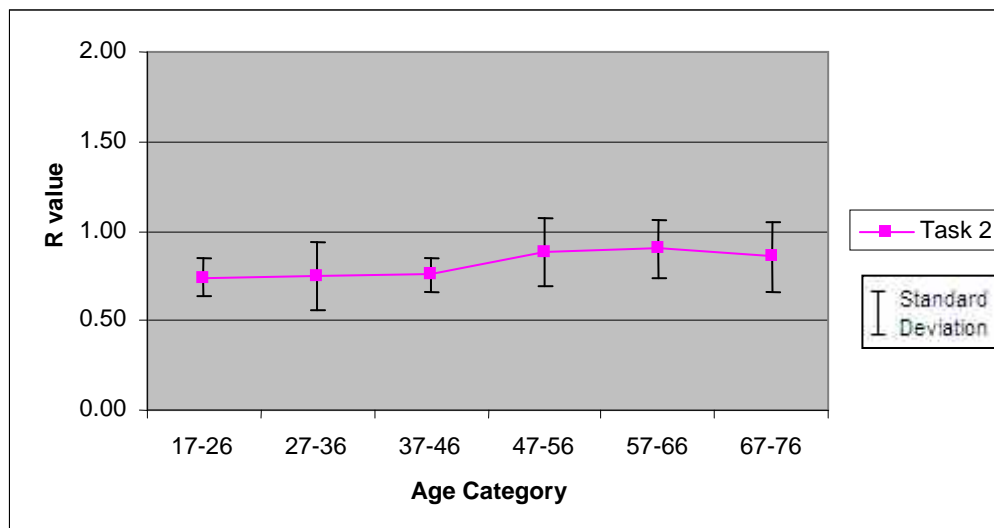
Figure 11: Task 2, R value against Age Category

The results for Task 2 were partly as expected; the three older age groups produced slightly more variation than the younger categories (see Figure 11 above). When an ANOVA was performed on these data it did not produce a significant result (MS = .053, df = 5, p = 0.09). However, a significance level of 0.09 was approaching significance; perhaps increasing the number of participants might have moved this figure in a more conclusive direction, but as it stands post-hoc, analyses were not merited.
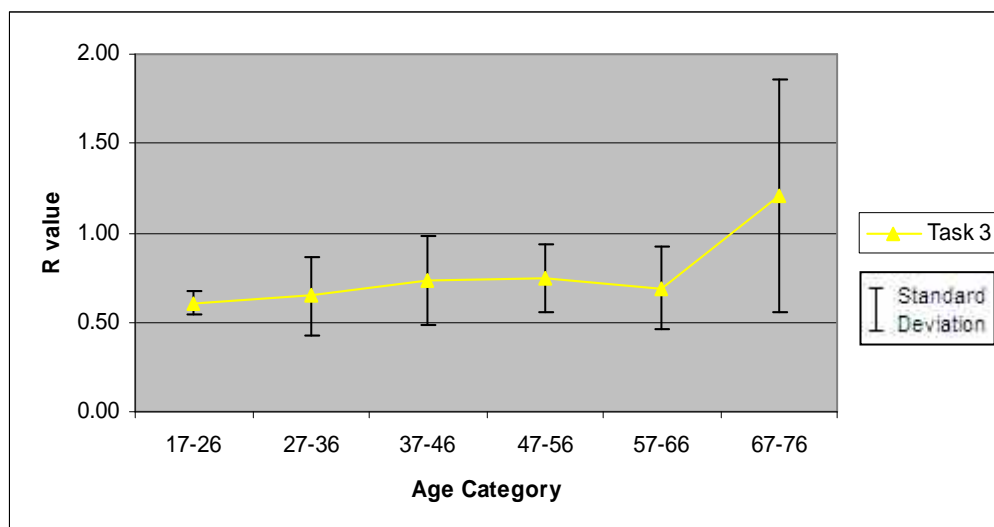
Figure 12: Task 3, R value against Age Category

The results for Task 3 were as expected; the age group 67-76 produced the largest amount of deviation compared to the other categories (see Figure 12 above). Unsurprisingly, given the sharp upward kick Figure 12 takes, at the 67-76 age category, the results of an ANOVA performed on these data produced a significant difference (MS = .460, df = 5, p = 0.002). The results of a Bonferonni post hoc-test showed the first four age categories (17-26, 27-36, 37-46 and 47-56) had significantly lower R values than the 67-77 age category. Note, the 57-66 age category was not also significantly different, although it was very near the significance level of .05 (p = .074).

Figure 13: Task 4, R value against Age Category

Task 4 also found a more expected pattern, where the three older age groups had a higher spreads of scores than the three younger groups (see Figure 13 above). The ANOVA conducted on these data did not reveal any significant differences ($MS = .120$, $df = 5$, $p = 0.47$).
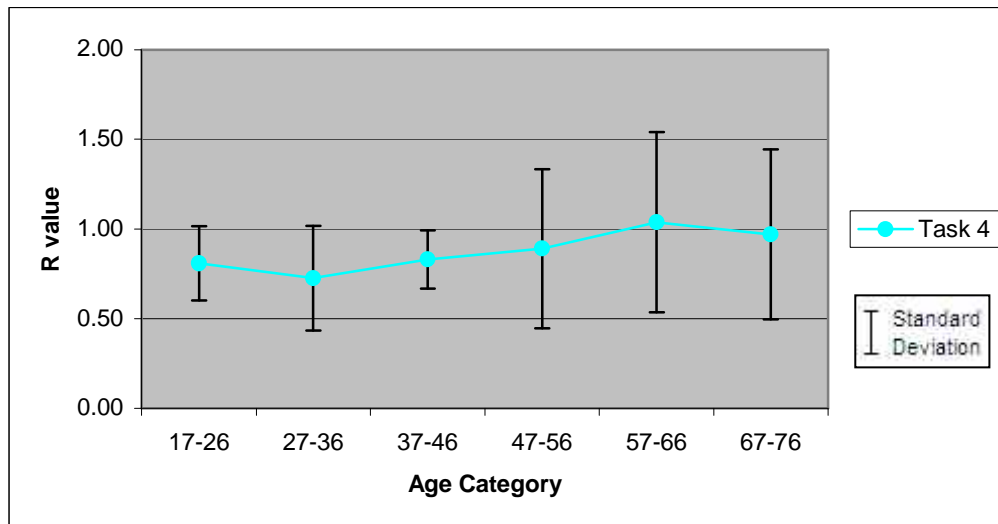
3.5.2   Data Sampling

As well as assessing the variations in TSOT and R due to age/types of task, the impact of age on performance of different IVIS tasks whilst applying the Occlusion technique needed to be ascertained in order to understand how age should be controlled for when producing a protocol for Occlusion. As seen in the literature review, a small number of studies have compared the age of participants within their experiments, but, as yet, no studies have analysed age directly with respect to the method outlined in the ISO standard.

As noted earlier, the ISO standard currently recommends there should be at least 20 % older people within an overall sample. What is the magnitude of the effect of randomly including 20 % of older drivers within the total participant sample? Randomly including participants and assessing the respective effect on the results allows conclusions to be drawn regarding the usefulness of the sampling method described in the ISO standard.

Therefore, the original Occlusion data was sampled by two different methods. Samples of 10 subjects were used for each method.
1. Random samples of ten participants were selected and the mean TSOT and R values were calculated. This was repeated ten times.
2. Random samples of eight participants aged under 50 were selected. Further random samples of 2 participants aged 50 and over were selected and were added to the sample of eight participants (so, following the ISO standard). Again TSOT and R were calculated for each of the ten samples used.

The differences between and among the two sampling strategies were compared to assess if there were any significant differences between the results obtained.

TSOT

In Figure 14 below, the blue diamonds relate to sampling strategy 1 and the pink squares relate to sampling strategy 2.

Figure 14: Mean TSOT value by sample number

The standard deviations of the two sampling strategy sets were calculated. For sampling strategy 1 the standard deviation was 1.43 and for sampling strategy 2 the standard deviation was 1.16. A t-test was performed on the data to ascertain if there was a significant difference between the two different sampling strategies; there was no significant difference recorded (t= 1.684, df = 18, p = 0.11).

R value

As shown in Figure 15 below, the dots shown in blue relate to sampling strategy 1 and those in pink relate to sampling strategy 2.

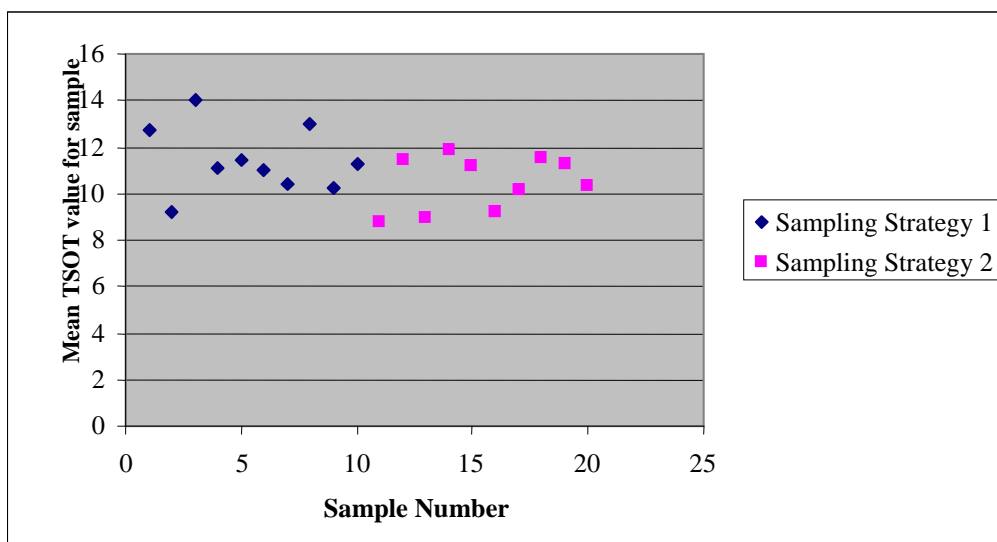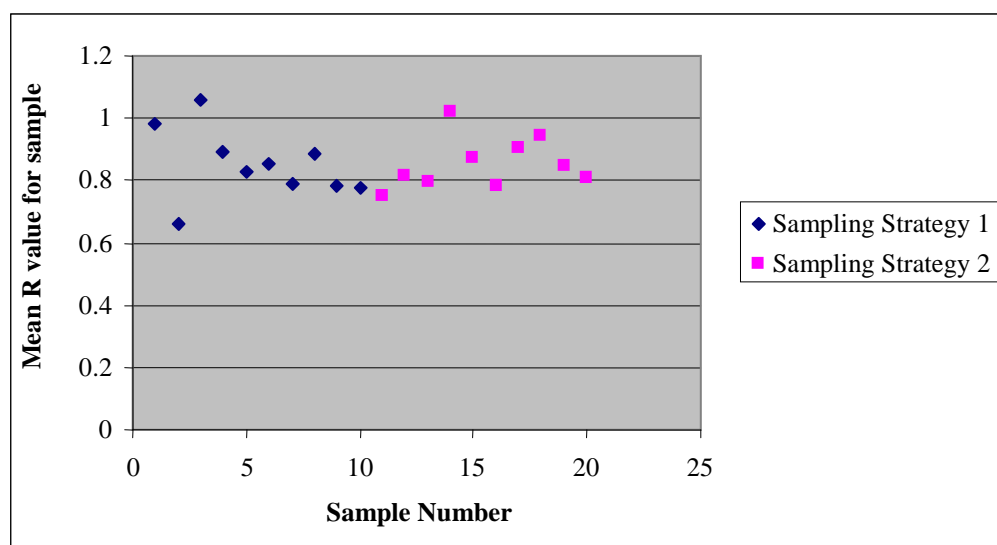**Figure 15: Mean R value by Sample Number**

The standard deviation of the two data sets was calculated. For sampling strategy 1 the standard deviation was 0.11 and for sampling strategy 2 the standard deviation was 0.08. A t-test was performed on the data to ascertain if there was a significant difference between the results of the

two different sampling strategies; there was no significant difference recorded (t = -.0.79, df = 18, p = 0.94).

Gender comparison

Finally, a comparison of the TSOT results was made for male and female participants. The results of these *t*-tests showed no gender effects for any of the four different tasks; Task 1 (t=-1.094, df=57, p=0.279); Task 2 (t=-1.047, df=58, p=0.299); Task 3 (t=-0.529, df=57, p=0.599) and Task 4 (t=0.314, df=56, p=0.755). This implies that an experimenter in a future Occlusion study does generally not need to control for gender (unless of course there are specific reasons for a representation of a population, such as for predominantly male lorry drivers).

3.5.3 Learning Effects

As described earlier, the study followed the ISO standard and used five trials for each participant using the Occlusion goggles (as well as five samples without the goggles). Before the actual testing, each participant was given appropriate practice both with and without the goggles - the intended purpose being to ensure that the participant is at "the top of the learning curve" when they start the trials proper. In other words, the participants' performance should not significantly change across the five trials.

The data for the four tasks (for all age categories) were analysed to assess if learning effects existed. If a significant learning effect was present then this may change the Occlusion protocol being developed in WP4 (for example, by requiring more practice trials to be given before the actual testing).

Task 1 and Task 2

Separate ANOVAs were performed on the data for Task 1 and Task 2 to see if there were any significant differences between each of the five samples and, therefore, evidence of a potential learning effect (see Figure 16 below for Task 1). There were no significant differences recorded for either Task 1 or Task 2 (Task 1 (*MS* = 22.8, *df* = 4, *p* = 0.165); Task 2 *(MS* = 17.1, *df* = 4, *p* = 0.31)). Therefore, these indicate no learning effect occurred in Task 1 or Task 2.
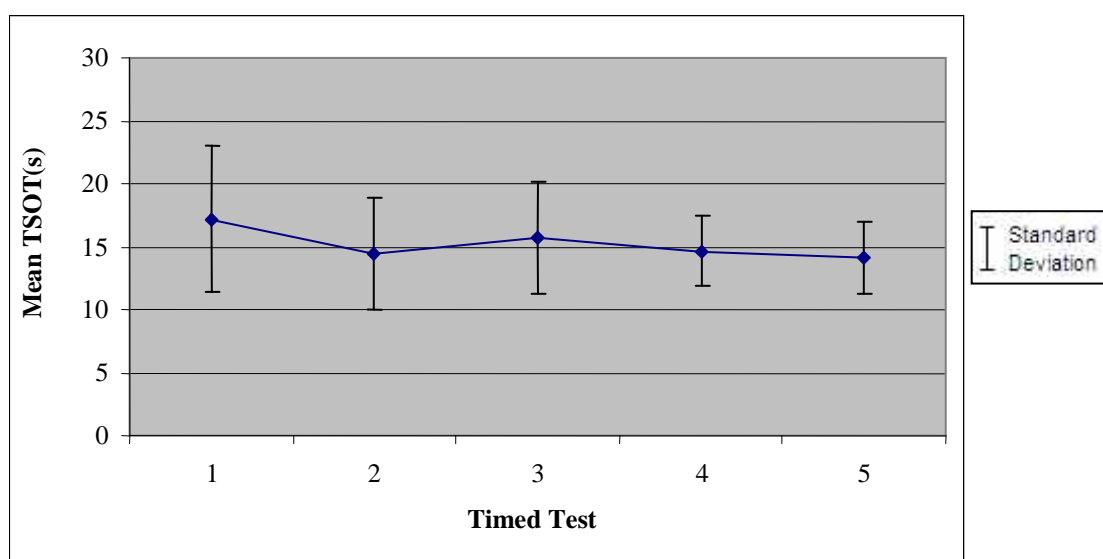

Figure 16: Task 1: Order of tasks by TSOT

Tasks 3 and 4

An ANOVA was performed on the data to see if there were any significant differences between the five samples and therefore a potential learning effect. Unlike for Tasks 1 and 2, here there was a significant difference recorded for tasks 3 and 4 ($MS = 399.46$, $df = 4$, $p < 0.001$ for Task 3, and $MS = 215.78$, $df = 4$, $p = 0.02$ for Task 4).

However, it must be noted that participants got on average *slower* on this task rather than *faster*. Therefore, although there are some significant differences within the order for this task, participants did not get quicker (i.e. displayed learning through improved performance times). The precise reason that they became slower is unknown, but it may be, in part, due to a vigilance decrement (as they were undertaking an experiment that took one hour overall).

Overall trial order

Differences in performance due to the order of presentation of Occluded or non-occluded trials were investigated[‡]. In other words, analyses were computed to establish whether performance was significantly different between those participants who performed the Occluded trials first followed by the non-occluded trials, and *vice-versa*. For all four tasks t-tests were employed between those participants who completed the Occluded trials first against those who did not. The results are displayed in Table 4 below:

Table 4: Differences in Performance Depending on Occlusion/Non-Occlusion Order (t-test results)

Task	t	df	p
Task 1	1.03	57	0.31
Task 2	0.02	58	0.98
Task 3	0.80	57	0.43
Task 4	-0.64	56	0.53

As shown in Table 4, no significant differences were found regarding overall trial order.

[‡] Following the ISO standard, the procedure employed was to test half of the participants on the Occlusion trial first and the non-Occlusion trial second, and the other half on non-Occlusion first and Occlusion second.

3.6   Key Implications

In term of participant age, our results found some differences between the age categories for TSOT and R, whereby the older participants showed a greater spread of scores (especially for TSOT).

Overall, this would imply that to obtain minimal inter-subject variability an experiment should ideally use younger/middle aged participants. However, it should be borne in mind that the exact task used will likely produce different results to the four tasks tested here (as all our four tasks produced different results), so ideally some piloting would be required by any experiment wishing to minimise variability due to participant age.

Ideally Work Packages 3 and 4 should use younger participants to minimise inter-subject variability (and especially avoid the 67-76 age group). Gender imbalance may be maintained for specific reasons (for example, to reflect a population - such as truck drivers being mainly male), but this does not appear to be a significant factor in Occlusion performance.

Regarding data sampling, the results found that there was no overall difference in either TSOT or R due to the sampling strategy used (either 10 random participants or stratifying a sample of 10 with two participants aged over 50 years). As such, this implies that experimenters in a future Occlusion study may not need to be too stringent with their criteria for selecting participants. For example, whether or not two participants over the age of 50 are included (as suggested by the ISO standard) might not greatly influence the results. However, as our results have shown, using subjects aged over 67 would increase the variability of the sample, so should be avoided (although, of course, this does not imply that IVIS tasks should only be designed for younger participants, instead it simply focuses on the process to minimise variability during testing).

Finally, our results found no significant performance improvements due to learning beyond the ISO standard-specified practice requirements. From the evidence obtained here it does not seem that learning is occurring through the 5 trials on any task, and that the practice sessions given are enough for a participant to reach the top of the learning curve. Likewise, no significant differences were found between those participants who undertook the Occlusion trials first, and those who did them second (ie after the non-occluded trials). As such, our results imply that the practice sessions and experimental order outlined in the ISO standard are sufficient[§], and therefore they will be used in the Occlusion protocol in WP4.

[§] In the case of experimental order it may be unnecessary, but not harmful.

4   Work Package 2 - Consideration of IVIS tasks and expert evaluation

4.1   Background

To help put the obtained results of WP1 into context, WP2 considered the results of the four different IVIS tasks from WP1 with reference to other data sources. In particular, WP1 produced a large amount of Occlusion related data, and WP2 provided additional information to help interpretation of these Occlusion findings. It examined what aspects of tasks may be distracting/lack usability, how the results relate to theoretical task performance time and what a large sample of the UK driving public consider "acceptable" tasks to be performed on the roads.

4.2   Study aims

The aim of this WP was to help understand the characteristics of IVIS tasks that are overly distracting. This assisted in later stages of this project to help establish external benchmarks for IVIS tasks using the Occlusion protocol.

4.3   Method

Work Package 2 was primarily conducted by means of a two day workshop with staff from TRL and University of Nottingham. In addition, some data (the social acceptability ratings) were obtained before the workshop, and others were obtained afterwards (the keystroke level task analysis).

4.3.1   Expert Opinion

The expert opinion/heuristic analysis was undertaken by four professionally qualified human factors researchers. As shown in the results, a matrix was designed to provide a structure for the outputs generated during a brainstorming. Using this matrix, the four experts assessed the characteristics of the four IVIS tasks that might be overly distracting, in particular examining the usability and likely demand from these tasks.

Before the brainstorming session, opinion was also obtained from two international experts (from CUT in Germany). The views of these two international researchers were incorporated into the expert analysis.

4.3.2   Keystroke Level Task Analysis

Based on previous 'keystroke level' task analysis protocol developed by members of the project team (Pettitt, Burnett and Stevens, 2006), such analyses were undertaken on each of the four tasks. The purpose was to quantify how long completion of the four IVIS tasks should theoretically take, and to compare findings to the data obtained in WP1.

The Keystroke Level Task Analysis model is usually used to predict and analyse specific tasks with a computer interface. It was developed in the 1980s and was designed to make predictions of task performance time by expert users on routine system tasks. The technique involves decomposition of a task into primitive actions, known as Operators, examples of this include pressing a key or button or movement of a hand between input devices. Based on that, total time taken for an "expert performer" is calculated, taking into account both mental and physical tasks (Pettit el al, 2006). Although primarily developed for desktop computing situations, the Keystroke Level Model (KLM) can also be used to predict the Occlusion measures TSOT and R, and thus for our purposes can be used to compare with the earlier data obtained for the four specific IVIS tasks for TSOT and R. These data can help explain the results from the empirical study completed in WP1, in particular why some interfaces are likely to be more demanding than others.

4.3.3   Social Acceptability

As will be seen in a later section of this project (WP5) there is a variation of existing Occlusion criteria around the world (for instance, between Japan and the USA, and even within the USA) Given this variation, it is argued here that to help develop appropriate UK Occlusion metrics (Demand Reference Levels) it is important to also take into account what tasks are considered to be social acceptable by a large sample of UK drivers.

Social acceptability ratings of performing the different IVIS tasks during driving were collected from 60 UK drivers of aged from 17-76 (the same subjects who were used for WP1, so all were familiar with the four tasks). For each IVIS task they were asked two questions:

> 1 – Do you believe drivers should be able to carry out this task while driving?

> 2 – Would you consider carry out this task while driving?

How the responses to these questions are linked to the participant's actual performance (under conditions of Occlusion) was also analysed.

Finally, ancillary data obtained from an on-line survey of 712 navigation system users concerning use of navigation system functions while driving were collected to put the obtained earlier results of 60 participants into a wider context (Forbes and Burnett, 2007).

4.4   Results

The results for the three data sources are presented separately; thereafter, key implications of the combined data are drawn.

4.4.1   Expert Opinion

To facilitate appropriate data collection, and to aid understanding the four tasks were assessed on four broad (and somewhat overlapping) categories, these were:

- the Input: how the driver enters information.
- the Task: what needs to be done.
- the Display: what is presented.
- the Output: what results are displayed by the system.

In Table 6 (see the following page), the results of all four tasks, for all these four categories, are displayed as a matrix. The results are coded whereby:

- those shown in red were judged to be negative features (i.e. potentially distracting and/or not having optimal usability),
- those in black were neutral features (but still important to note),
- and those in green were positive features for a specific task.

As can be seen in full details in Table 6, all tasks have some negative features (shown in red); however, in general tasks 3 and 4 have more negative features (except in the 'input' category as they are visual tasks that do not require the user to enter data). Likewise, tasks 1 and 2 have more positive features (shown in green) in the categories of 'task', 'display' and 'output'. To quantify this, by assigning scores of +1 for each green (positive) item, 0 for each black (neutral) item and -1 for each red (negative) item produces the following (Table 5):

Table 5: Summary of scores from the expert review matrix

Task	Green items (score of +1)	Red items (score of -1)	Overall score Green minus red scores
Task 1	14	9	**+5**
Task 2	8	7	**+1**
Task 3	3	11	**-8**
Task 4	3	12	**-9**

As such, there is a progression in task difficulty/usability difficulties, especially between those tasks that are driving related (tasks 1 and 2) and those non-driving related (tasks 3 and 4).

Table 6: Expert Review Matrix

Stage	Task 1	Task 2	Task 3	Task 4
Input	4 button pushes/ 4 screens	9 button pushes/ 5 screens	0 button pushes/ continually moving screen	0 button pushes/ continually moving screen
	require stylus	require stylus	No input required- purely visual	no input required- purely visual
	user paced	user paced	system paced	system paced
	learning required (experience effects)	learning required (experience effects)	less learning required	less learning required
			success depends on individual strategy	success depends on individual strategy
			depends on effective perceptive/cognitive processes/speed (but only one column to scan)	depends on effective perceptive/cognitive processes/speed (3 columns to scan)
Task	could need to use under conditions of high workload	could need to use under conditions of high workload		
	structured task	Structured task	no task structure	no task structure
	driving related	driving related	not driving related	not driving related
	requires cognitive categorisation (e.g. airport or shopping centre)		no reference points in screen - constantly changing visual image	no reference points in screen - constantly changing visual image
	task is chunkable / interruptable	task is semi-interruptable/chunkable (no info lost although additional glances may be	task is not interruptible	task is not interruptible
	sometimes better than stopping to re-route	user may need to act on partial info for example spelling of town names		
	scrolling sub-task too long			
	interferes with driving task due to holding information in working memory	interferes with driving task due to holding information in working memory	interferes with driving task due to holding information in working memory	interferes with driving task due to holding information in working memory
	requires taking hand off wheel	requires taking hand off wheel	does not require taking hand off wheel	does not require taking hand off wheel
	confirmation bias - look for similar e.g Bristol / Brighton	forgetting street/other address - need to hold info in memory		
Display	uses icons	some button pushes are on same screen	no differentiation between lines e.g. colours	no differentiation between lines e.g. colours
	size of selection items	Non-QWERTY keyboard	does not facilitate searches e.g. alphabetical	does not facilitate searches e.g. alphabetical
	Partly uses colour coding			
	predictability of display		moving target, speed of movement	moving target, speed of movement
	selection of final point of interest (fiddly)			
Output	auditory feedback possible	auditory feedback possible	picking wrong values no feedback	picking wrong values no feedback
	Visual feedback – immediate	visual feedback – immediate		
	reduces need to use additional distracters e.g. pieces of paper	reduces need to use additional distracters e.g. pieces of paper	might need to use additional distracters e.g. pieces of paper	might need to use additional distracters e.g. pieces of paper
	land on/off	land on/off		
	little system delay	little system delay but problem with fast input		

4.4.2   Keystroke Level Task Analysis

Each of the WP1 tasks (Tasks 1-4) were broken down into 'operators' as explained in Section 4.3.2 As a result the total task time taken for an "expert performer" is predicted which takes into account both mental and physical tasks.

The Operators used in this analysis are shown in Table 7:

<p align="center">Table 7: Operators used in analysis</p>

Operator name	Time value (s)
Reach far (Rf) – time for hand to move to device	0.45
Mentally prepare (M) – time to read information, verify action, hold information in memory, etc.	1.35
Homing (H) – time to move hand between controls (buttons) on device	0.40
Keying (K) – time to press a control (button)	0.20

The analysis can be used to predict and explain the TSOT and R values that were experimentally produced in WP1. For specific variants of each of the four tasks, analysis using the Keystroke Level Model (KLM) was conducted, and is reported below. Points are then made regarding differences in KLM predictions and observed experimental values for TSOT and R between tasks.

Task 1 – Point of Interest Search

Example task: receiving route guidance for journey from current location to Birmingham Airport
> [**]Mean observed static task time (unoccluded) = 9.55s
> Mean observed TSOT = 7.63s
> Observed R = 0.85 [††]

KLM for static task time
> Rf (Move hand from lap to control of device)
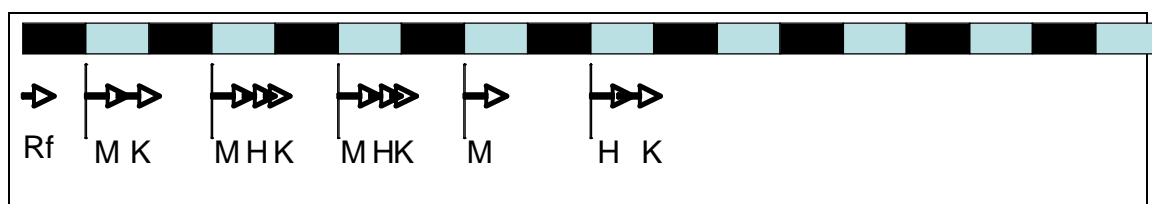> +M+K (Main menu-select "Navigate to")
> +M+H+K (Navigate to menu-select POI)
> +M+H+K (POI category menu-select category type)
> +M+H+K (List of POIs – select relevant one)
> = 7.85s

KLM under Occluded/unoccluded conditions (where the black areas represent the 1.5s Occluded periods and the blue ones correspond to the 1.5s shutter open periods):

[**] The observed values listed above and in other similar sections refer to the sample of 60 participants.
[††] As mentioned previously, R is calculated using the 'mean of ratios' approach; hence the apparent discrepancy in the calculation above (the "ratio of means" method would produce an R value of 0.79.)

From the above diagram, predicted TSOT = <u>7.5 s</u>
Therefore, predicted R = <u>0.96</u>

Note:
The first two H operators are unlikely to require full vision, as they will be automated movements associated with the top-level menus of the device. However, the final H (locating the specific POI name from a long list – 11 items) will require vision and therefore is modelled to start at the beginning of a shutter open period.

Task 2 – Address entry

Example task: receiving route guidance to address "12 Cedar Avenue, Runcorn"

Mean observed static task time (unoccluded) = 14.65s
Mean observed TSOT = 11.77s
Observed R=0.81

KLM for static task time
Rf (Move hand from lap to control of device)
+M+K (Main menu-select "Navigate to")
+M+H+K (Navigate to menu-select Address)
+M+H+K+H+K+H+K+H+K   (City – enter 4 letters)
+ M+H+K (City-select city from list)
+M+H+K (Street-select street from list)
+M+H+K (House number – say Done)
= <u>13.6s</u>

KLM under Occluded/unoccluded conditions (where the black areas represent the 1.5s Occluded periods and the blue ones correspond to the 1.5s shutter open periods):
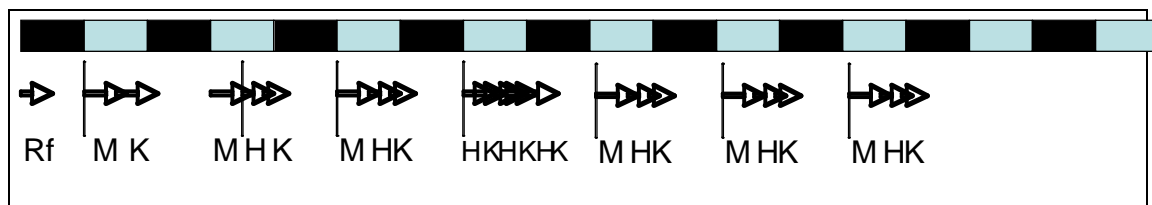
From the above diagram, predicted TSOT = <u>10.5 s</u>
Therefore, R is predicted to be <u>0.77</u>

Note:
It is assumed that only one M is required prior to entering the letters of the city name (as the City is one chunk in working memory). Also, the selection of a specific city and specific street are from short lists (max. 4 items). Therefore, it is assumed that the associated H operator does not require full vision.

Task 3 – Short scrolling task

Example task: finding YHR company code and reading out associated share price

Mean observed static task time (unoccluded) = 14.12s
Mean observed TSOT = 10.94s
Observed R=0.78

In conducting this calculation, the specific search target with a mean static task time closest to 13.9 s was used – this was S8 "YHR". This was done because of the clear differences between variants of a task according to when the search term would first appear on the screen.

KLM for static task time

[‡‡]System response = 10s
+ M (Search for target on screen)
+ M (Read accompanying share price)
= 12.7 s

KLM under Occluded/unoccluded conditions (where the black areas represent the 1.5s Occluded periods and the blue ones correspond to the 1.5s shutter open periods):
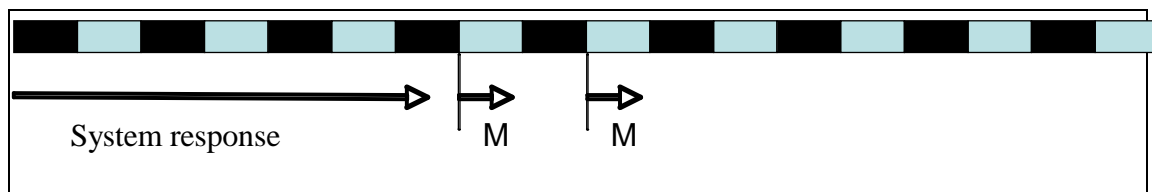
From the above diagram, predicted TSOT = 7.35 s
Therefore, R is predicted to be 0.58

Notes:
This task is problematic to model because KLM assumes that participants should have been able to detect the target during the first full 1.5 sec shutter open time, once the target appeared on screen, and then read the accompanying price during the next shutter open time (before the information went off screen). However, from observation of individual observed TSOTs, it is clear that this did not always happen.

Indeed, it is interesting to note that there was much more variability in TSOT values (in the observed data) than static task times. This shows the problem with resumability for this style of interface which should have been reflected in relatively high R values. However, R was quite low (also shown by the KLM). The KLM indicates that R was lower than it should have been because of the system response time (where no interaction could occur) which would significantly reduce TSOT in relation to static task time.

Task 4 – Long scrolling task

Example task: finding RLQ company code and reading out associated share price

Mean observed static task time (unoccluded) = 13.6s
Mean observed TSOT = 11.62s
Observed R=0.87

As for the short scrolling task, a variant was chosen with a similar static task time to the overall mean (L8= "RLQ").

KLM for static task time

System response = 5.3s
+ M (Search for target on screen)
+ M (Read accompanying share price)
= 8.0 s

[‡‡] This is the time taken for the scrolling letters 'YHR' to appear on at the bottom of the screen

KLM under Occluded/unoccluded conditions (where the black areas represent the 1.5s Occluded periods and the blue ones correspond to the 1.5s shutter open periods):
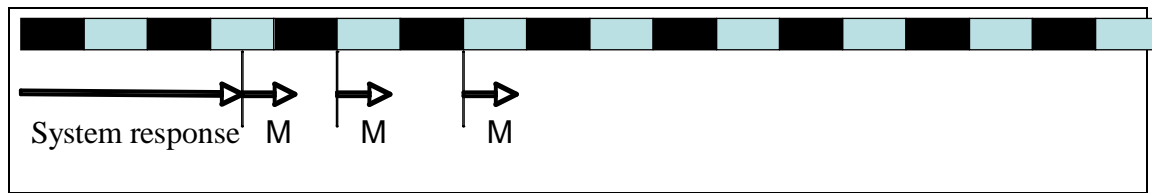
From the above diagram, predicted TSOT = <u>5.85 s</u>
Therefore, R is predicted to be <u>0.73</u>

Notes:
Because the first M operator cannot be completed during a shutter open time, it has to start again at the next opportunity.

As for the short scrolling text task, the results are affected considerably by a) the basic assumption of error free performance and b) the "wait" for the search target to appear. Observation of standard deviations for TSOT values versus static task times for this task show even greater variability, indicating why observed R was relatively high.

Summary of results

Table 8: Results of observed and predicted TSOT and R

TASK	Observed TSOT (WP1)	Predicted TSOT (KLM)	Observed R (WP1)	Predicted R (KLM)
Task 1	7.63	7.50	0.85	0.96
Task 2	11.77	10.50	0.81	0.77
Task 3	10.94	7.35	0.78	0.58
Task 4	11.62	5.85	0.87	0.73

General points/conclusions about the KLM

- KLM modelled TSOT performance well for tasks 1 and 2. Relative results were preserved and absolute errors were low.

- There was a tendency for the KLM to underestimate static task time and TSOT. For tasks 1 and 2, it is possible that the KLM model needs adjusting to become more accurate.

- KLM indicates that the differences between tasks 1 and 2 was largely due to the number of stages involved in the interaction. This explains why there were differences for TSOT/static task time (but not for R).

- The analysis for tasks 1 and 2 assumes instantaneous response from the system. Although, the TomTom system does respond quickly, there will still be some delay moving between screens, especially when information within a database is accessed.

- KLM was inaccurate for tasks 3 and 4. The KLM indicated the importance of system response time (waiting for information to appear) on the resumability ratio. However, further work would be required in expanding the M operator for use in modelling tasks where participants may not be able to find targets in single 1.5 second bursts of visual attention.

4.4.3   Social Acceptability

In this section the social acceptability of the four IVIS tasks used in WP 1 are considered.

The responses to the questions (Questions 1 and 2) which obtained information on social acceptability are shown below.

Question 1 – Do you believe drivers should be able to carry out this task while driving?

Question 2 – Would you consider carry out this task while driving?
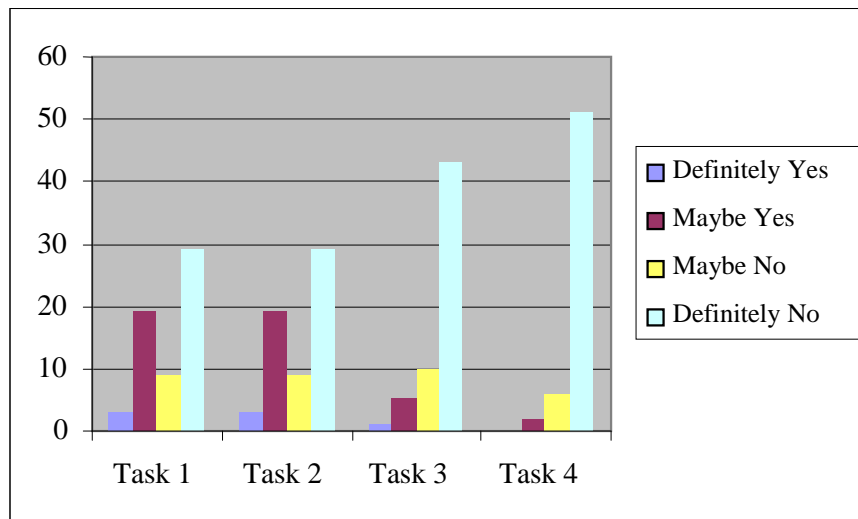
Figure 17: Frequency of Response to Question 1 by Task
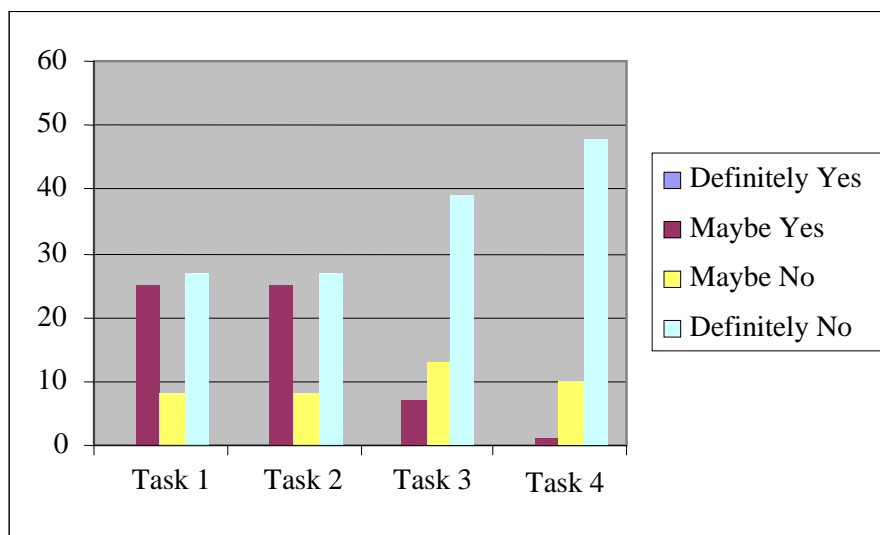
Figure 18: Frequency of Response to Question 2 by Task

Both questions produced an overall similar pattern of results, although there was a tendency for drivers to judge they may personally carry out the tasks, but such tasks may not be appropriate for drivers in general. That is, drivers stated that they would be generally more likely to engage in the task (question 2) whilst driving as compared with their judgements of whether the task should be allowed to be performed by drivers as a whole (question 1).

Purely focusing on question 1, about a third said that Tasks 1 and 2 should be allowed (either 'definitely yes' or 'probably yes' answers), about 10% said Task 3 should be allowed and less than 5% said Task 4 should be allowed (with nobody saying 'definitely' allowed). As such, the four tasks vary on how socially acceptable they are to perform whilst driving.

The figures below further analyses the responses and breaks them down by age categories.
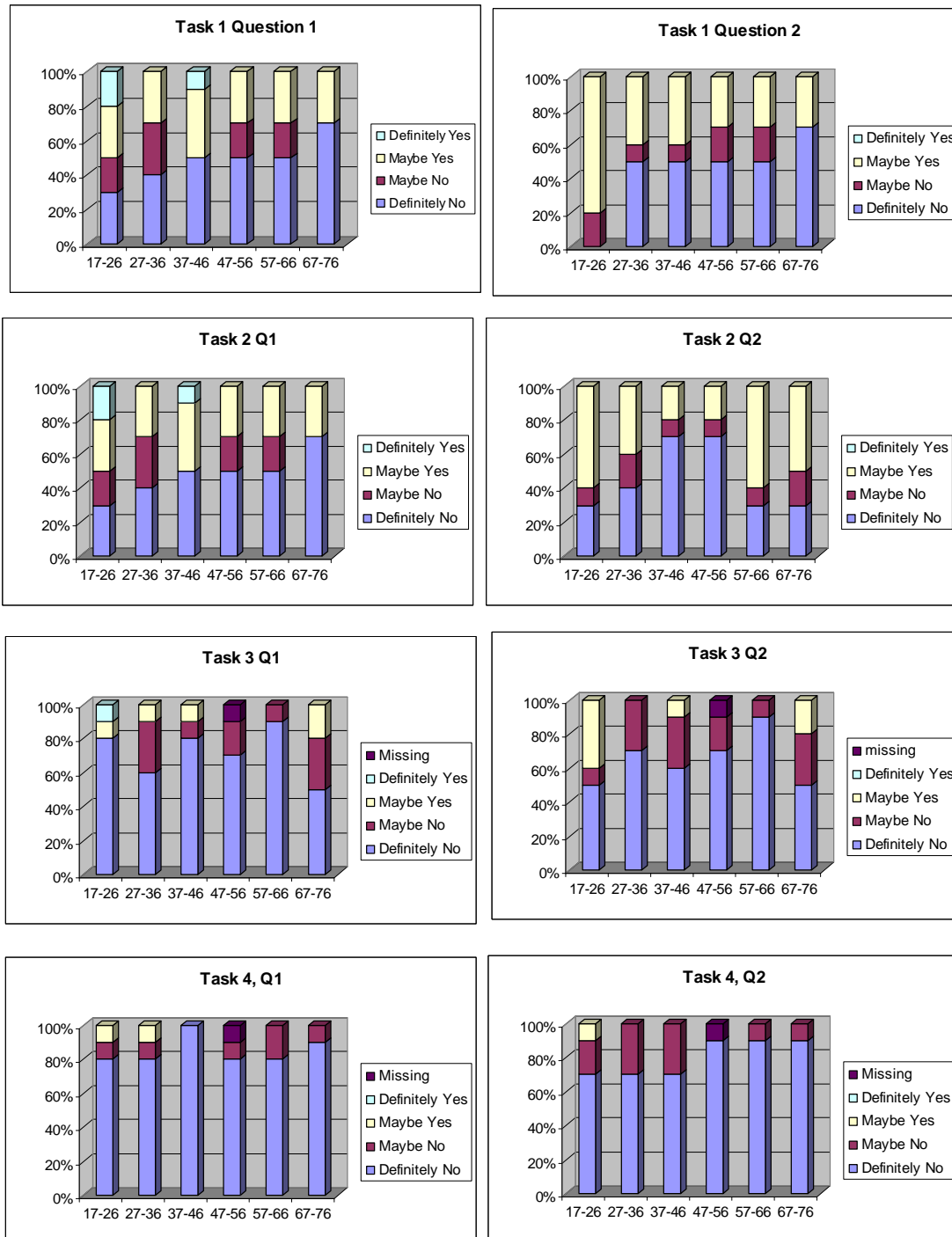
Figure 19: Frequency of response to research question by task and by age category

Again, there is quite a strong correlation between the results in questions 1 and 2 for all age categories. Focusing one question 1, generally the younger participant thought the tasks were more socially acceptable to do whilst driving compared with their older counterparts. For example, for Task 1, approximately half the 17-26 said it should be permitted (either

'definitely yes' or 'probably yes' answers), whereas only about a quarter of the 67-76 group thought it should be allowed (with no 'definitely yes' answers).

Research Question results compared to WP1 Data

To compare the social acceptability ratings to the actual performance results from WP1, the results of research Question 1 (Do you believe drivers should be able to carry out this task while driving?) was compared with their average TSOT values for each of the different tasks. The scoring for the research question is as follows: 1 = 'Definitely No', 2 = 'Maybe No', 3 = 'Maybe Yes' and 4 = 'Definitely Yes'.
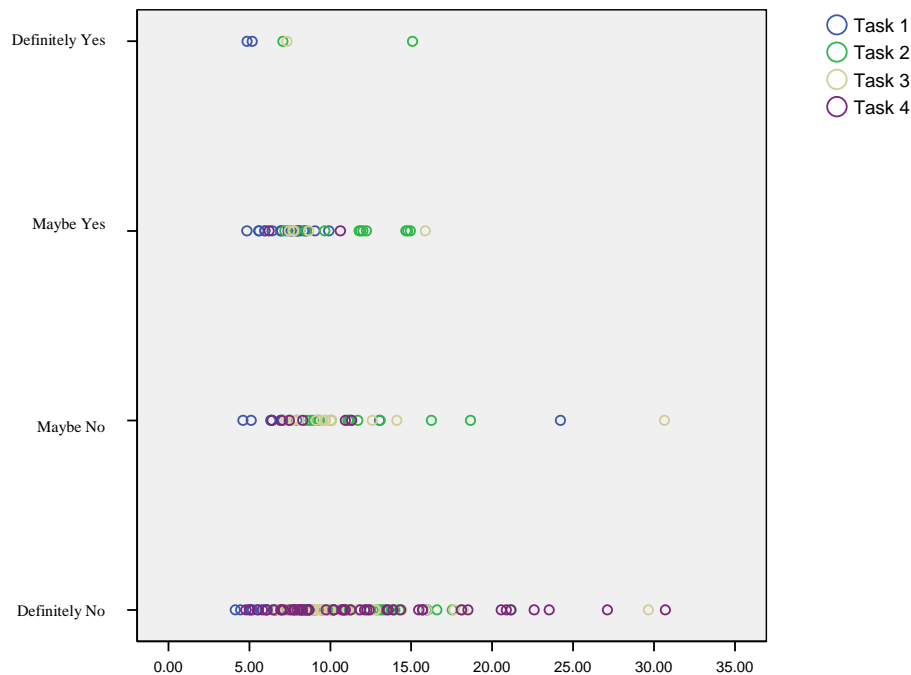

Figure 20: Results of Research Question 1 by Mean TSOT value

Of particular interest was to explore if there was a relationship between the obtained TSOT values and the research question ratings. For example, did those participants who stated a task could be undertaken whilst driving obtain a faster (ie, lower) TSOT score than those participants who stated a task should not be undertaken whilst driving? To explore this, a correlation between the obtained TSOT values and research question ratings was performed.

The results found no significant association (correlation = - 0.138, p= 0.293); this means that the TSOT score obtained by a participant is not linked to their rating of the acceptability of the task. As such, participants' opinions of whether an IVIS task should be allowed to be performed whilst driving is not significantly linked to their actual IVIS task performance (as measured by the Occlusion metric TSOT).

Age Effects

To explore age effects of social acceptability, the data set was subdivided by the age categories used in WP1. None of the results (for the specific age categories) found a significant association between the TSOT scores obtained by WP1 and the rating of the acceptability of the task.

Therefore, whilst the Demand Reference Level developed in WP5 should certainly include consideration of social acceptability ratings, more objective measures (e.g. Occlusion scores

for different aged participants, Lane Change Task results, other worldwide metrics in use, expert tasks reviews, Keystroke Level task analysis and alcohol impairment benchmarks) will assume an equal importance.

Ancillary Data: Survey of Navigation System Users

The results below show the response to three questions from an on-line survey of 712 navigation system users (novice and expert) concerning the use of navigation system functions while driving. The data presented below allows the results of the above 60 participant survey to be put into a broader context (for example, to confirm that the above results from 60 drivers were not abnormal, they were compared to a larger sample).

Novices were defined to be those navigation system users who had:

 a) Only ever owned one navigation system, and

 b) Had owned it for less than 12 months (n=160).

Experts were defined to be those navigation system users who had:

 a) Owned more than one navigation system, or

 b) Had owned one for more than 12 months (n=552).

Figure 21 shows that experts admit to entering destinations while driving more frequently than do novices. In fact, approximately 17% of the experts admit they enter destinations while driving quite often or frequently.
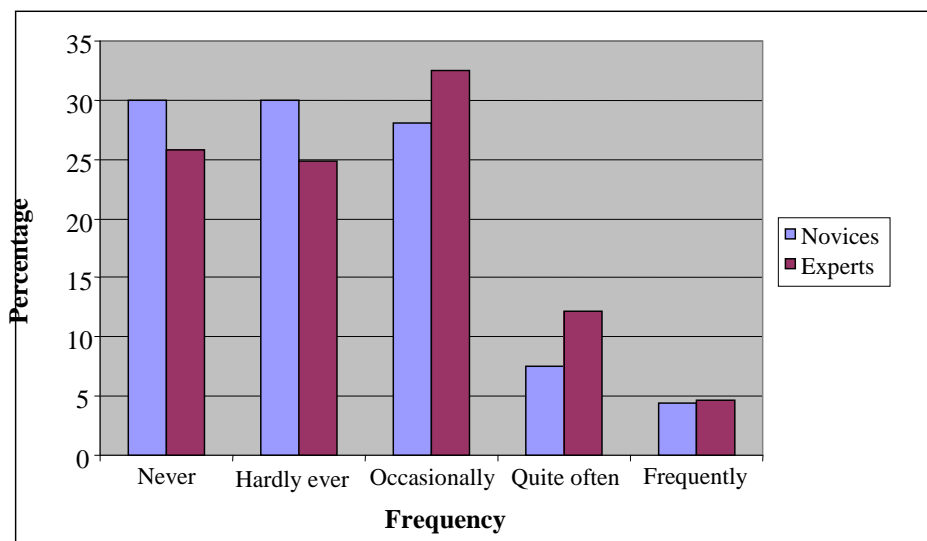
Figure 21: The frequency that participants admit to entering destinations while driving

Figure 22 is most directly applicable to the aims of this Occlusion project and shows that destination entry by stored location is considered most acceptable followed by destination entry by Point Of Interest (POI) and then entry by postcode/address. In relation to the tasks used in the Occlusion study, approximately 33% of the experts believe destination entry by POI is acceptable and 27% of experts believe destination by address is acceptable.
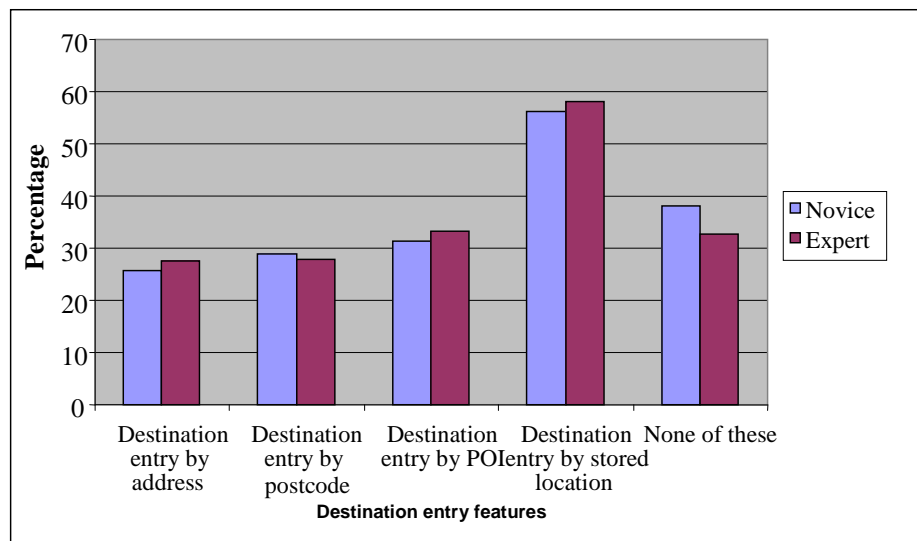
Figure 22: Destination entry features that participants think should be allowed while driving

Finally, Figure 23 shows that there is a big difference in the perceived acceptability of manipulating an IVIS map while driving - approx. 37% of experts believe it is acceptable for someone to conduct this task while driving compared with 20% for novices. There were few differences for other functions.
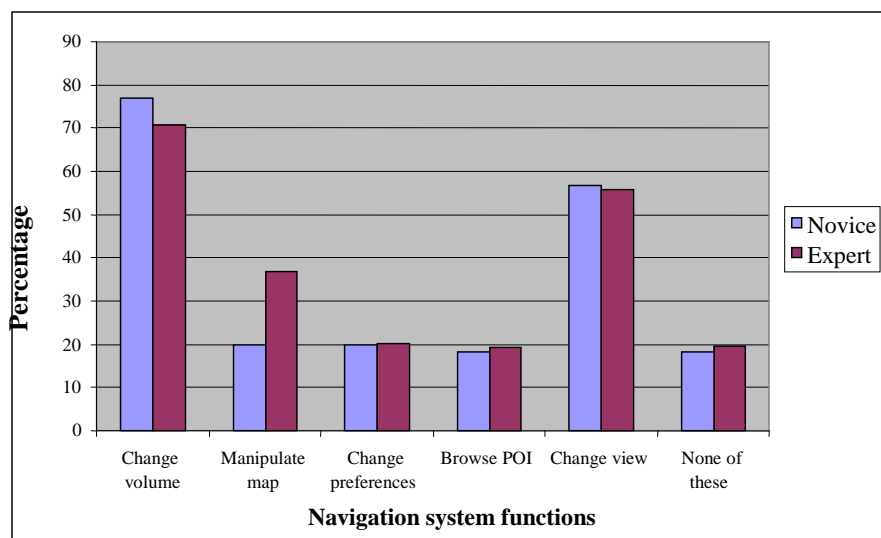
Figure 23: Navigation system functions that participants think should be allowed while driving

The characteristics of participants, who admitted to entering destinations while driving, were compared with those of participants who did not admit to this. It was found that they:

- Were significantly younger ($p < 0.001$).

- Considered themselves significantly more skilled at using computers ($p < 0.001$).

- Were significantly more experienced drivers (based primarily on miles driven in the last 5 years) ($p < 0.001$).

- Were significantly more experienced navigation system users ($p < 0.05$).

The characteristics of participants, who considered some form of destination entry while driving to be acceptable, were compared with those of participants who thought that any form of destination entry while driving was unacceptable. It was found that they:

- Were significantly younger ($p<0.01$) and had held their driving licences for significantly less time ($p<0.05$).

- Considered themselves to be significantly more skilled at using computers ($p<0.001$).

- Used their systems actively (i.e. with route guidance) in familiar areas significantly more frequently ($p<0.005$).

As such, the results of this ancillary study broadly support the results of the survey with 60 participants for the four IVIS tasks. In particular, younger drivers are generally more likely to opine that IVIS tasks are more acceptable to be performed whilst driving than older participants. Furthermore, related to the key characteristics of the tasks that are overly distracting (uncovered in the earlier-described Expert Review) those IVIS tasks that are perceived to be 'easier', quicker to perform and more driving task related are more likely to be socially acceptable. For example, destination entry by a stored location is more acceptable than destination entry by full address, which in turn is more acceptable than finding a share price on a scrolling display.

4.5   Key Implications

This Work Package undertook a fundamental review of our range of four IVIS tasks used in the Occlusion trials. In particular:

- By using a panel of subject matter experts, it examined the likely usability issues and demands (perceptual, cognitive and physical) from these tasks while driving, and uncovered the different types of distractions associated with each of the tasks.

- By undertaking a decomposition of each task, it revealed how the results of the Occlusion trials from WP1 related to the theoretical time needed to complete them.

- By conducting two surveys (one specifically for the four IVIS tasks, and one covering a wider area), it produced quantifiable data about which tasks, and which components of the tasks, are considered acceptable by UK drivers.

These three areas produced a rich array of results that helped the overall objective of this project to develop Demand Reference Level criteria using the Occlusion protocol. In particular:

- The expert review showed from first principles the characteristics of the four tasks that are overly distracting and/or may cause usability difficulties - these results both help understand and predict obtained Occlusion scores, and help establish what particular features of the task make it acceptable/unacceptable to the UK driving population.

- The matrix demonstrated that in addition to scores achieved for TSOT and R (related to the visual demand of a task) there are other factors that may impact upon acceptability of an IVIS task. It is clear from the results of WP1 that R scores were relatively low, yet from the social acceptability results, participants still noted that they felt the tasks were possibly not appropriate while driving. The matrix demonstrated that the issue of social acceptability is complex and other factors that impact on acceptability include: manual effort (e.g. number of button pushes, need to remove hands from steering wheel), interaction with equipment (stylus, screen, vehicle control) and cognitive distraction (such as, working memory required for the task).

- Such results help demonstrate, from first principles, the characteristics of the four tasks that may be overly demanding and/or may cause usability difficulties. When developing limits (in WP5), the earlier list of positive and negative task characteristics

can theoretically help explain why a new IVIS task has obtained its specific TSOT and R scores. Such an approach may be useful in future for developers, policy makers and researchers to understand Occlusion scores. Thus possibly it could be used to also assist designers (by maximising the properties of a task so that it has more 'green' features). This approach was applied to the quantification of the four tasks used here, and it showed that there was a progression in task difficulty (with task 1 being the easiest and task 4 being the most difficult, and the largest difference was between tasks 1 and 2 (driving related) and tasks 3 and 4 (non driving related).

• The Keystroke Level task analysis showed that the theoretical task completion times (both Static and Occluded) were closely linked to the results obtained from WP1 for two of the four IVIS tasks. This does not, of course, directly imply that only a Keystroke Level task analysis is sufficient to demonstrate which IVIS tasks may be acceptable, as it was not a good predictor for the other two IVIS tasks. However it does help support, augment, and explain the results obtained. Specifically, the Keystroke Level analysis provided important information regarding why R was not as high as was expected for the scrolling text tasks. Furthermore, the data demonstrated the primary interface characteristics that led to differences in TSOT values between tasks 1 and 2.

As such, the findings of this Work Package provide a valuable pool of data to support development of a Demand Reference Level for Occlusion: it shows what aspects of tasks may be distracting/lack usability, how the results relate to theoretical task performance time and what a large sample of UK driving public would consider acceptable to be performed whilst driving.

5   Work Package 3 - Benchmarking IVIS tasks to another impairment inducing factor (alcohol) and performance measures (Lane Change Task)

5.1   Background

In addition to the Occlusion method, other techniques exist to assess visual demand from IVIS tasks; one is the Lane Change Task (LCT). LCT measures performance decrement on a driving-like task whilst a participant is also engaged with an IVIS task. The LCT and Occlusion are both comparatively new techniques, so assessing which one is most sensitive to detect driver visual demand decrements is timely.

Impairment from alcohol is well-known to reduce driver performance, and limits on the maximum amount of alcohol permitted in a driver exist throughout the world. As such, comparing the performance decrements from alcohol intoxication at a 'set' level (the UK maximum permissible level for drivers) against the decrements from engagement in different IVIS tasks could help identify which IVIS tasks would be acceptable to perform whilst driving. Of course it should be noted that alcohol has other (perhaps more important) negative effects on driving, e.g. risk taking, reaction times and lane keeping, plus it has an impact over a much greater period, so it cannot be exactly compared to impairment from IVIS use. Instead, an overall comparison of IVIS impairment against alcohol was undertaken.

5.2   Study Aims

WP3 aimed to examine the four representative IVIS tasks in relation to other demands (the Lane Change Task) and known driving impairments (alcohol). By comparing the results obtained from these aspects to those found from Occlusion they help to further clarify what would be an appropriate Occlusion Demand Reference Level.

5.3   Method

5.3.1   Participants

Fifteen participants (7 females, 8 males) were selected at random from the TRL volunteer database, a pool of 1,300 drivers selected to represent a cross section of the driving population. Participants were required to have a full United Kingdom driving licence and normal or corrected vision.

Age boundaries for younger and older participants were established on the basis of the results from the age trials in WP 1 (for example, older participants were excluded). The average age of the selected sample was 44 years 6 months (range 23 to 56), they have held a driving licence for an average of 25 years and 9 months (range 5 to 38) and travelled an average of 7,269 miles per year (range 2,000 to 15,000).

Participants with severe alcohol problems were excluded; however, participants were required to be regular consumers of alcohol. Teetotal drivers would be unsuitable for this experiment as they would have little or no tolerance to the effects of alcohol. Experienced drinkers are able to tolerate increased levels of alcohol in the body without demonstrating the outer symptoms associated with alcohol consumption (such as loss of concentration, impaired vision and loss of balance) and as the vast majority of the population are drinkers it was decided to sample from the largest proportion of the driving population.

Participants were paid £30 for involvement in the study. Due to the nature of this study they were also provided with transport to and from the experimental facility.

5.3.2   Materials

Alcohol

Appropriate ethical approval was obtained for the proposed method of inducing intoxication. Participants were required to drink an alcoholic beverage. The beverage constituted vodka (40%) plus a disguising mixture (e.g. cream-soda) mixed using the adjusted Widmark formula (Stowell & Stowell, 1998) so that participants become intoxicated at the legal limit of Blood Alcohol Content (BAC) (80mg/100ml) with the volume of the mixer adjusted to maintain a 20% volume.

TRL has previously produced conversion charts based on weight, height and gender for when using the Widmark formula to simplify the task for the experimenter in determining the dose and mix of alcohol for each participant.

For male participants: $(3.82-(0.143*A) + (0.168*H) + (0.525*W))* T/100$

For female participants: $(-3.67+ (0.187*H) + (0.432*W))*T/100$

Where:

A = Age in years
H = Height in cm
W = Weight in kg
T = Target concentration

Participants consumed an average of 132ml (range 92- 202ml) of alcohol (184-404ml including mixer). This produced a mean intoxication of 38.7µg per 100ml of breath with (range 29-50µg) within a 40 minute waiting period (before the alcohol trial was completed).

Breathalyzer

The breathalyzer used in this study was a Lion laboratories alcometer SD-400. It is a hand-held instrument with automatic breath sampling, data storage, PC compatibility and digital display giving read-out of the subject's breath alcohol concentration.

PDA/ TFT screen

In the IVIS conditions participants were required to complete 4 LCT trials. Each of these trials was dedicated to one of four IVIS tasks previously used in WP1:
  • PDA POI – entering a destination by selecting a "point of interest" using the PDA

  • PDA Address – entering a destination by "address" function using the PDA

  • Shares Short – searching for a share price from a single scrolling column of text using an LCD screen

  • Shares Long – searching for a share price from three scrolling columns of text using an LCD screen

The equipment used in this experiment was an after-market Hewlett Packard iPAQ Pocket PDA, operating on a Windows platform running the Tom-Tom satellite navigation system and

a 8 Inch TFT LCD Monitor produced by LinITX. Further details were previously given in WP1.

Health Questionnaire

The health questionnaire contained items relating to the participants' current and past medical history which may have been relevant when considering the nature of this study. The inclusion of questions concerning whether female participants are or could be pregnant again reflects the nature of experiments that require participants to consume alcohol. It contained further questions related to participant's alcohol consumption that helped decide whether or not participants should have been excluded from the experiment on grounds of acute alcohol consumption.

For the purpose of this study, the health questionnaire was purely a screening tool to select appropriate subjects; as such, no further information will be reported here.

5.3.3   Lane Change Task (LCT)

The Lane Change Task is a laboratory based combined control and event detection metric based on the dual task paradigm. The dual task paradigm posits that primary task performance will degrade with the introduction of a secondary task. In this case, LCT performance can be viewed as the primary task and it is designed to be analogous to the driving task. It was developed as part of the ADAM project (Advanced Driver Attention Metrics; Mattes, 2003).
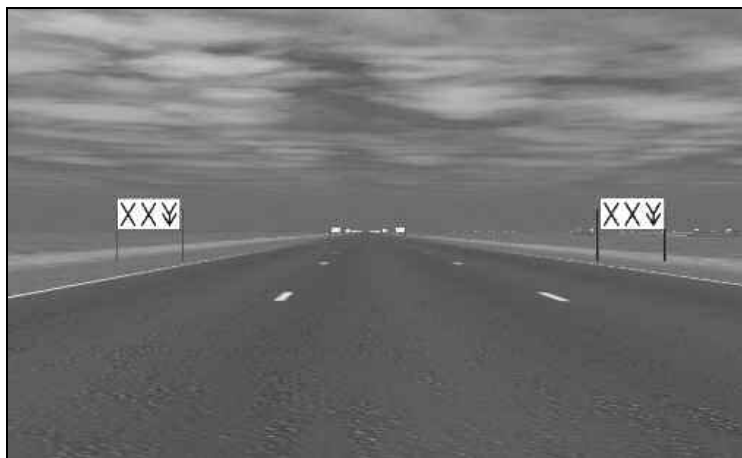

Figure 24: Screen shot from the LCT. In this instance the driver has to change from the centre lane to the right lane.

The LCT requires participants to negotiate a 3000m long section of three lane highway. Participants are instructed by signs on the roadside (150m apart) to perform a lane change manoeuvre. During this task participants are required to perform a specific secondary task. To avoid speed confounding the results it is controlled by the program and is kept at a constant 60 km/h. Illumination levels were set to approximate daytime driving with a constant light level. Low level engine sound was provided in order to provide the driver with auditory feedback more akin to what would they would hear in the driving task. Visual information was presented using an egocentric (front) view; no visual information was presented regarding side or rear views. The vehicle dynamics are such that the simulated car behaved as a standard passenger car. Participants were required to change lane when instructed; when not performing a lane change manoeuvre they were required to maintain a central position within

the lane. Performance of the lane change task by itself was used as a measure of baseline performance for comparison with performance of the LCT and a secondary task.
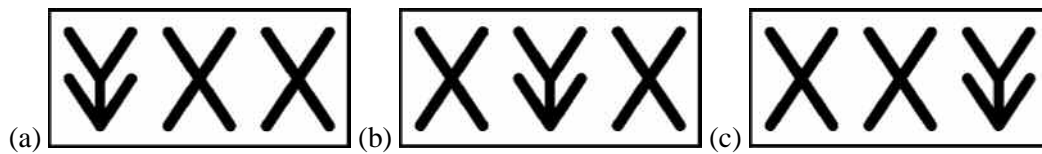

Figure 25: Signs; (a) left, (b) centre, (c) right.

During a LCT trial the program automatically records data to the computer on which it is running. From these data the LCT analysis program calculates a number of performance measures. These include; Mean deviation from the normative model, standard deviation from the normative model, mean steering angle, as well as time course and distance information to allow for standardisation of experimental runs.
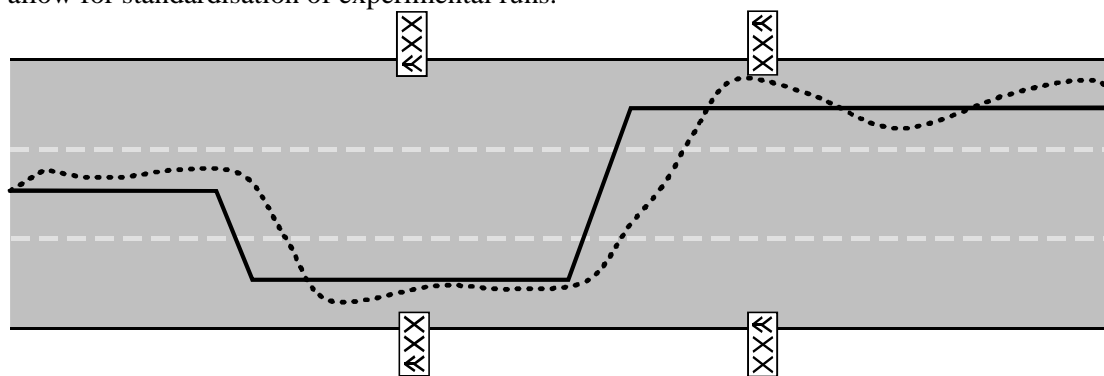

Figure 26: The LCT compares the normative model (solid line) to the participants driving course (broken line, driving from left to right).

5.3.4   Procedure

Informed consent was obtained from participants prior to commencement of the experiment. Upon giving consent, participants were required to complete the health questionnaire to ensure that they were fit to proceed with the study.

Participants were breathalysed before the experiment started to ensure they were not already intoxicated. If their blood alcohol level was already raised above zero then they were excluded from the experiment as it was problematic to calculate the correct dosage of alcohol to achieve the desired level (BAC 80mg/100ml).

Participants were also given the opportunity to familiarise themselves with the LCT by completion of practice laps and also were familiarised with the IVIS interfaces. Instructions were provided to participants about the overall study, as well as more specific experimental instructions concerning how to complete the trials.

The first five experimental conditions were counterbalanced. However, condition 6 (alcohol) was always performed last due to participants being intoxicated upon its completion. These trials were terminated once the tasks had been completed (c.f. conditions 1 and 6 where duration was fixed at 15 minutes). The six experimental conditions were:

- **Baseline (Condition 1)**
  Participants were required to complete 5 LCT laps lasting 15 minutes in total. These trials were completed without the presence of a secondary task and without the

influence of alcohol. These trials served to act as a baseline measure of driving performance.

- **The 4 IVIS Conditions**
  For each of the four IVIS conditions, participants completed five trials while performing the LCT (i.e. driving), and another five trials were performed with the LCT paused (i.e. while not driving). The order of these trials was counterbalanced. The four IVIS conditions were:

    o **Condition 2:** PDA POI – entering a destination by selecting a "point of interest" using the PDA.

    o **Condition 3:** PDA Address – entering a destination by "address" function using the PDA.

    o **Condition 4:** Shares Short – searching for a share price from a single scrolling column of text using an LCD screen.

    o **Condition 5:** Shares Long – searching for a share price from three scrolling columns of text using an LCD screen.

- **Alcohol (Condition 6)**
  In the alcohol condition, participants were given 10 minutes drinking time in which to consume the mixture, followed by a brief waiting period (20 minutes from finishing the drink). The justifications for such a waiting period are that the effects of alcohol take around 40 minutes to reach their peak and it may take this long for any residual alcohol on the breath to disperse. Participants were breathalysed again prior to beginning the LCT section of the experiment to ensure that they were at or over the legal limit. There were 5 three minute LCT laps lasting a total of 15 minutes.

Participants were required to remain in the test area after completion until their BAC returned to normal (this was necessary because it took longer for BAC to return to normal than it took to complete the trials).

5.4   Results

5.4.1   LCT Results

Participants consumed an average of 132ml (range 92- 202ml) of alcohol (total quantity 184-404ml including mixer). This produced a mean intoxication of 38.7µg per 100ml of breath (range 29-50µg) within the 40 minute waiting period. This was just marginally above the maximum intoxication level for driving of 35µg per 100ml.

A one-way repeated measures ANOVA was calculated for mean deviation from the normative model on the LCT across the six conditions (Baseline, PDA POI, PDA Address, Shares short, Shares long, and Alcohol). There was a significant main effect by condition for mean deviation from the normative model [$F_{(5, 15)} = 14.421$, $p < 0.05$]. Figure 27 shows the mean deviation from the normative model by LCT treatment condition. It shows that there was no significant difference between baseline performance of the LCT and performance of the LCT under the influence of alcohol. This was supported by the post-hoc comparisons.
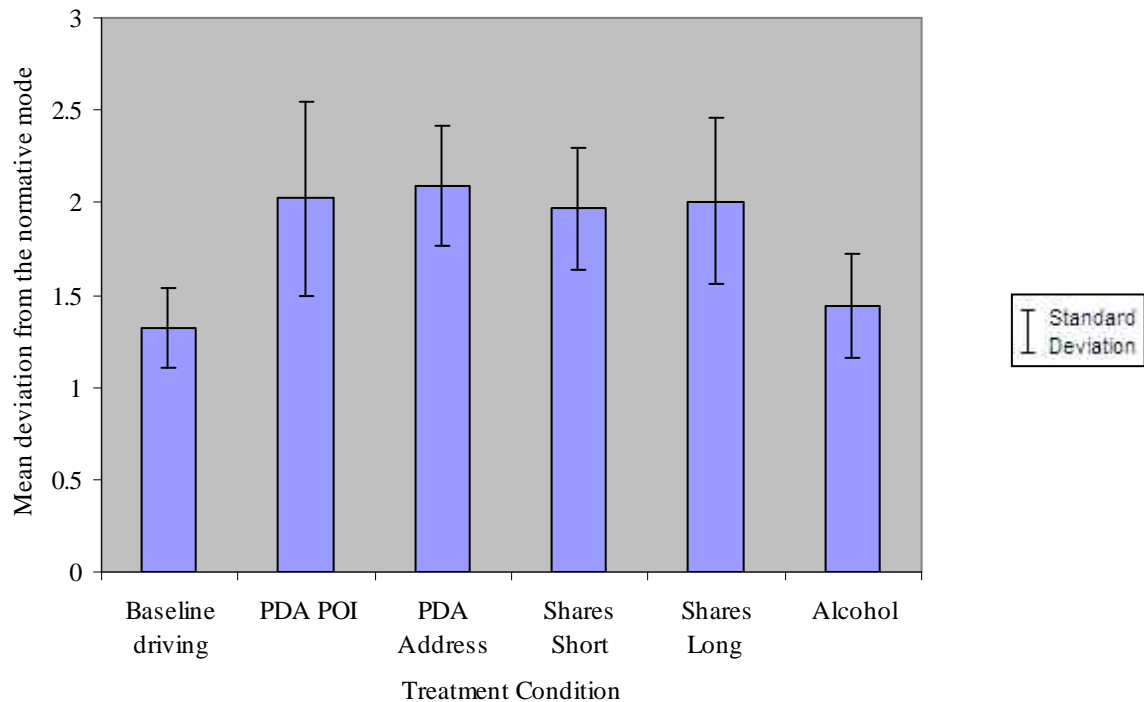
Figure 27: Mean deviation from the normative model by LCT condition

A Bonferroni pair-wise comparison of the six treatment conditions was conducted. There were a number of significant comparisons (at $p < 0.05$):

- Baseline vs. shares short.

- Baseline vs. shares long.

- Baseline vs. PDA POI.

- Baseline vs. PDA Address.

- Alcohol vs. shares short.

- Alcohol vs. shares long.

- Alcohol vs. PDA Address.

This implies that the majority of the IVIS tasks produced significantly worse results than the alcohol results of. Note: Alcohol and PDA POI scores were not demonstrated to be significantly different. This is likely to be due to the higher standard deviation present in the PDA POI data (see Table 9).

Table 9: Mean Deviations in performance from the normative model: mean and standard deviation results for each experimental condition

	Mean	Standard Deviation
Baseline LCT	1.32	0.22
Alcohol	1.44	0.29
PDA POI	2.02	0.52
PDA Address entry	2.10	0.32
Shares Short	1.97	0.33
Shares Long	2.01	0.45

Comparison of the four IVIS conditions revealed no significant difference in the mean deviation from the normative model for LCT driving between any of the tasks.

Whilst the mean deviation from the normative model was not different between all four IVIS conditions, there was a difference in mean total task times between the IVIS conditions. Additionally, mean total task time was found to be higher when participants completed the IVIS conditions in conjunction with the LCT. These results are presented in Figure 28. Note the PDA Address column is markedly higher than the other three (in addition the error bar also displays more variability, indicating a large variation in the ease with which participants were able to operate the PDA device and complete the LCT).
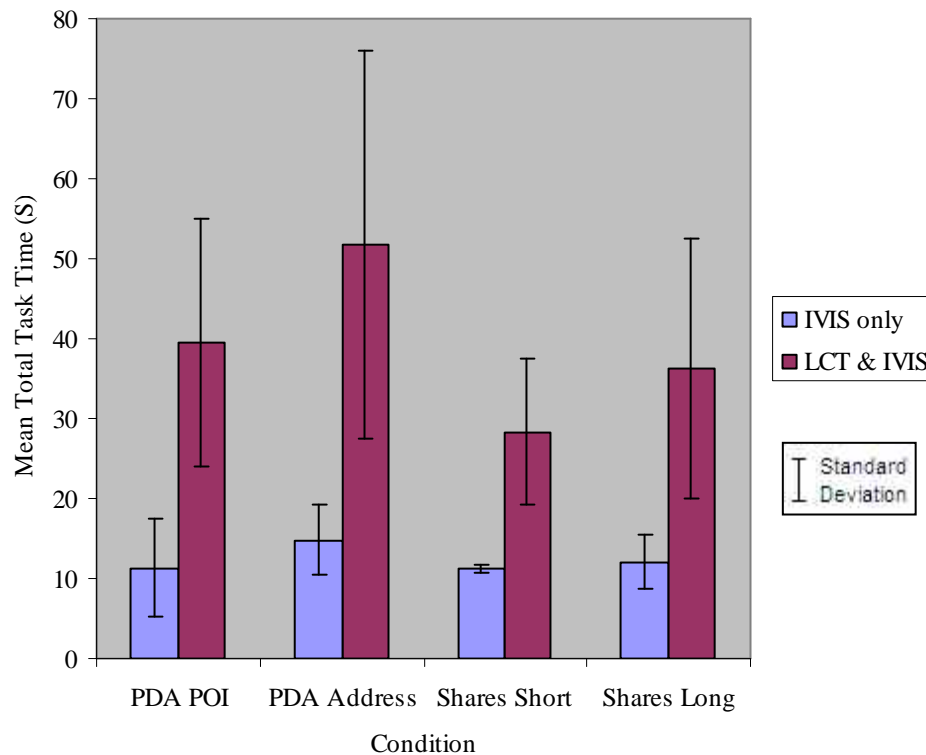
Figure 28: Mean total task time for IVIS conditions only, and in dual task (LCT) conditions

A repeated measures ANOVA revealed a significant difference between the mean total task time when only completing an IVIS task against completing an IVIS task while 'driving' though the LCT ($MS = 83.9$, $df = 1$, $p < 0.01$). Therefore, participants took longer to complete an IVIS condition when they were also completing the LCT. Furthermore, the ANOVA revealed total task time was significantly different between the four IVIS conditions ($MS = 990$, $df = 3$, $p < 0.01$).

5.4.2   Comparison with Occlusion Results

In terms of a broad comparison with the Occlusion results, Table 10 shows the performance of the four IVIS tasks for Occlusion (from WP1) and LCT.

Table 10: Summary comparison of Occlusion and LCT results for the 4 IVIS tasks

Task	Occlusion (Observed TSOT)	LCT (Mean MDev)
PDA POI	7.63	2.02
Address	11.77	2.10
Share Short	10.94	1.97
Shares Long	11.62	2.01

Note that WPs1 and 2, the social acceptability results, the usability evaluation and the pre-test pilot trials, all showed a relative difference between the four IVIS tasks. The Occlusion results also found this relative difference, in particular identifying that POI produced a significantly lower TSOT score than the other three tasks. However, the LCT failed to identify any significant differences in performance between the IVIS tasks; indicating a lack of sensitivity in comparison with Occlusion.

5.5   Key Findings

As demonstrated by the LCT results, driving whilst using any of the four IVIS tasks used in this study is significantly worse than driving at the United Kingdom legal alcohol limit.

There is an increase in the time taken to complete the IVIS tasks when performed in conjunction with the driving task.

Despite poorer performance in a dual-task situation, participants can maintain a consistent level of performance across the four IVIS tasks (as evidenced by no significant difference in LCT performance).

Compared with the Occlusion results, the LCT findings showed less ability to discriminate between the four IVIS tasks. This implies that Occlusion is a more sensitive measure for detecting performance differences due to varying levels of IVIS task demand.

The findings therefore allow benchmarking of the effects on performance of driving whilst undertaking the four IVIS tasks against a known, yet continuous, impairment factor (driving whilst intoxicated).

6   Work Package 4 – Protocol, Reliability and Loading Studies

6.1   Background

A key objective of this project was to address the issues required to establish an effective Occlusion methodology for the assessment of In-Vehicle Information Systems. It was the intention that the research would help develop a robust visual demand measurement protocol based on Occlusion. The ISO standard for Occlusion (ISO 16673: 2007) was published during 2007, but this research took an independent and critical view of all of its provisions. In particular, some aspects needed to be extended (e.g. precise experimental instructions for different IVIS task) and some elements modified (e.g. the age groups suggested in the standard) for an effective protocol to be developed.

6.2   Study Aims

The first aim was to develop an Occlusion protocol that took the form of a detailed method that contained all the necessary information for replication. Thereafter, a second aim was that the protocol should be evaluated by means of a reliability study. Finally, the protocol would be adjusted as necessary based on the findings of a study to examine if undertaking a primary task (in addition to the Occlusion task) made clear differences to the obtained results. The fundamental question concerned whether the results from an Occlusion-type study were modified by using a primary loading task during Occluded (non-visual) periods. Specifically, did a primary loading task affect the conclusions drawn from an Occlusion-related study?

6.3   Work Package 4(a): Writing the Protocol

The protocol was developed based on the Occlusion ISO standard, from the results of the first three work packages and by our practical experience of applying the Occlusion method.

The development process was iterative, whereby comments received by TRL from other organisations (e.g. University of Nottingham, UK and BASt, Germany) were used to further refine the protocol. Figure 29 shows how the protocol developed. The final version of the protocol is provided in Horberry et al. (2007; PPR 259).
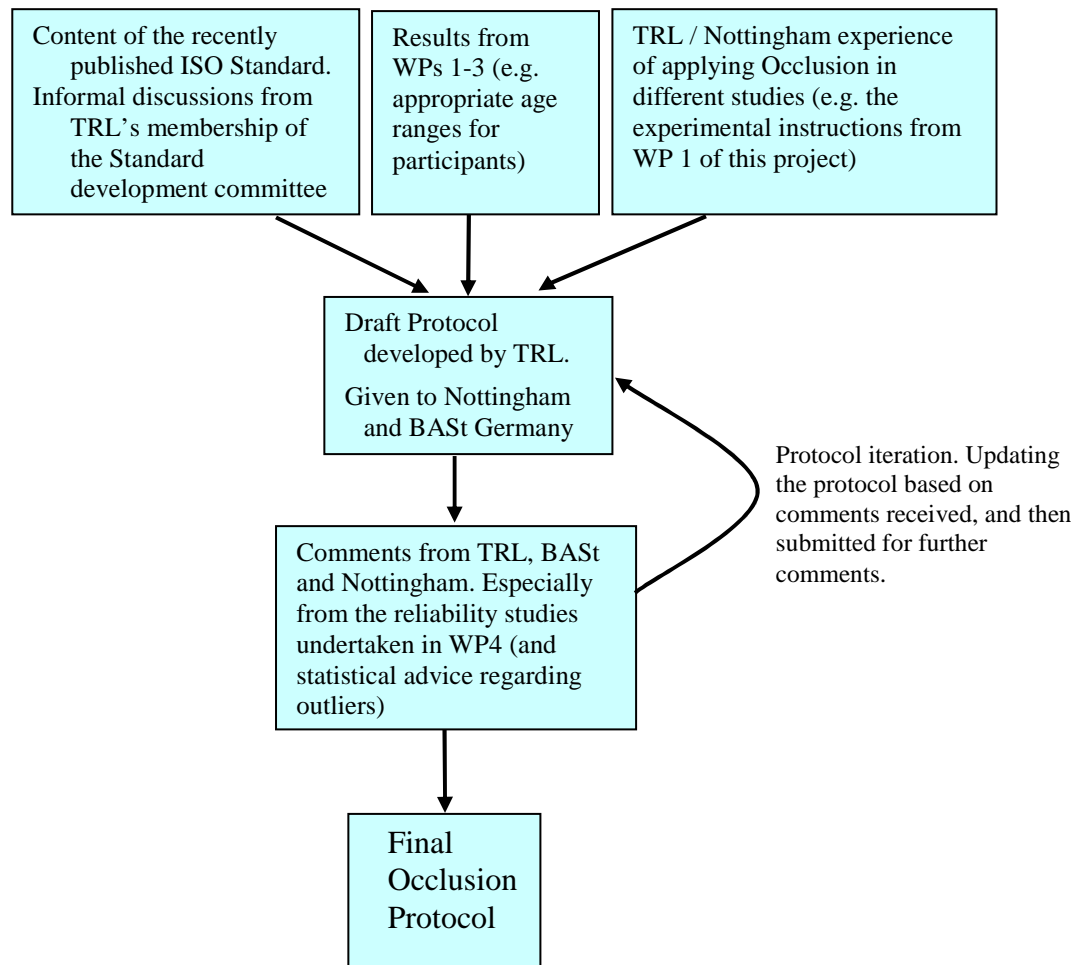
Figure 29: Protocol development

Compared with the ISO standard, the protocol largely expands and operationally defines the requirements for effective occlusion study measurement. Two more noteworthy changes are:

1. **Age and number of participants:** Although the ISO Occlusion standard recommends that 2 of the 10 participants employed should be aged above 50 years, the results of WP1 suggest that including subjects over 67 years may increase the variance of the sample. As such, unless there is a particular reason to focus on older drivers, it is recommended that the participants used are aged from 17 to 66 years old (for reliability reasons – refer to Section 3.6). This does not imply that the IVIS design and the DRL (discussed below) apply only to younger drivers. The restricted age range allows an increase in measurement confidence, and the DRL is appropriate for the whole age range.

   Similarly, a sample size of 20, rather than a minimum of 10 is prescribed to increase the confidence of the results.

2. **Outliers:** An outlier was not specifically defined in the ISO standard. However, the standard does suggest excluding data where the TSOT value was more than 4 times the average TTT_{Unocc} for all trials completed by that participant.

   There are many ways to handle outliers and no commonly accepted approach; the method used in this study (WP 1 &4) was based on the outlier definition from the Lane Change Standard, whereby any data that is more than 2 standard deviations

above the mean is replaced with the value of mean plus two standard deviations. However, following statistical advice, the protocol was modified (Horberry et al., 2007; PPR 259) and now specifies that data should be excluded (and not replaced with another value) in cases where TSOT exceeds 4 times the average TTT_{Unocc} for all trials completed by that participant. This method is broadly the same as the ISO standard but the 'TRL/Nottingham' protocol explicitly defined these as outliers that must be removed. In practice, by comparing these two methods on the data sets obtained, the results produced by these two exclusion criteria differ very little. However, the authors believe that the method used in the protocol follows best statistical practice and hence is preferable.

6.4   Work Package 4(b): Reliability Study

In addition to development of the protocol, a reliability analysis was conducted. Piloting of the protocol was performed at University of Nottingham and compared with results from an identical, earlier study conducted at TRL. Comparative tests conducted at independent locations are very informative in the development of a protocol that is designed to be simple to administer at independent sites.

By comparing the results achieved in the independent tests, we also gained a clear indication as to the reliability of the protocol, and hence the whole process of assessing IVIS by means of the Occlusion technique. The fundamental research issue that needed to be accounted for was whether the protocol is reliable. That is, can independent research/design teams make use of the protocol to reach similar conclusions?

6.4.1   Reliability Study Method

The study undertaken at University of Nottingham followed the most recent version of the Occlusion protocol specified by TRL. The data collected were then compared to the earlier TRL data.

Participants

As required in the protocol, twenty participants were recruited (all aged between 17 and 66, specifically, mean age 29; range 20-40). Key characteristics of the participants identified in a pre-study questionnaire (following the protocol instructions) are listed below:
- 15 male; 5 female.

- All participants held full driving licences.

- Mean years of holding a driving license was 8.8 years (range 2 to 21).

- Mean number of miles driven in last year = 9,800 (range 2,000 to 22,000).

- All participants were very experienced with using technology. For instance, they all used computers everyday and had used computers for an average of 14 years (range 9-25). Furthermore, 16 of 20 had used a navigation system before.

Study Location and Tasks Used

The study took place in a car buck within a lit indoor environment (see Figure 30 & Figure 31) and lasted for approximately one hour. Four tasks were evaluated in the study and were identical to those described earlier in this report:
1. *POI* – Entering a Point of Interest (POI) using a PDA version of the Tom Tom navigation system.
2. *Address* – Entering a street address using a PDA version of the Tom Tom navigation system.

3. *Short scrolling* – Finding a 3 letter stock code from a scrolling list (one column) and reading out the accompanying share price.

4. *Long scrolling* – Finding a 3 letter stock code from a scrolling list (three columns) and reading out the accompanying share price.

Figure 30: Reliability study set up (for tasks 1 and 2)

Figure 31: Reliability study set up (for tasks 3 and 4)

Procedure and Study Design

All participants were fully trained with the four tasks, as specified in the protocol. In a repeated measures design with counterbalancing, each participant undertook five variants of each of the tasks under two experimental conditions:

- Static (i.e. unoccluded, with full vision).

- Occluded, using the PLATO goggles with the 1.5 second shutter open/closed intervals (as specified in the protocol).

The times that participants took to perform the different tasks were taken manually using a stopwatch and values were then calculated for TSOT (Total Shutter Open Time) and R (resumability ratio) using the procedure specified in the protocol.

6.4.2   Reliability Study Results

Table 11 shows the mean values (with standard deviations in brackets) for Total Task Time when unoccluded (TTT $_{Unoccl}$), Total Shutter Open Time (TSOT) and the Resumability ratio (R) obtained at University of Nottingham.

Table 11: Reliability study main results

	TTT $_{Unoccl}$	TSOT	R
Task 1 – POI	10.1 (4.08)	8.0 (2.02)	0.86 (0.249)
Task 2 - Address	15.2 (5.30)	11.4 (1.97)	0.80 (0.198)
Task 3 – Short scrolling	8.0 (2.57)	7.8 (3.57)	1.05 (0.543)
Task 4 – Long scrolling	9.3 (5.16)	8.2 (5.06)	1.01 (0.502)

Comparisons were made between the results of this study at Nottingham and those arising from the WP 1 study at TRL (using 60 participants); Figure 32 and Figure 33 show mean values for TSOT and R for each of the four tasks across the two studies at Nottingham and TRL.

Additionally, as demonstrated by WP1, variability in Occlusion performance increases with age, therefore, to account for any differences caused by the Nottingham data being sampled across a younger age range, a subset of the TRL data was extracted covering all participants from 19 to 37 (22 participants, mean age 27).
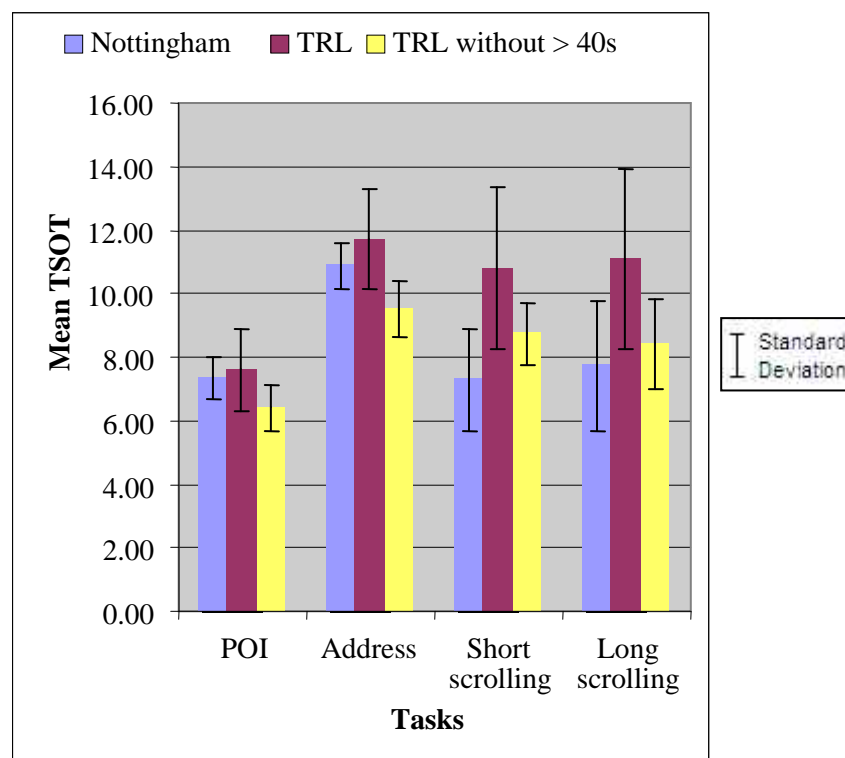

Figure 32: Comparison of University of Nottingham and TRL study results – mean TSOT
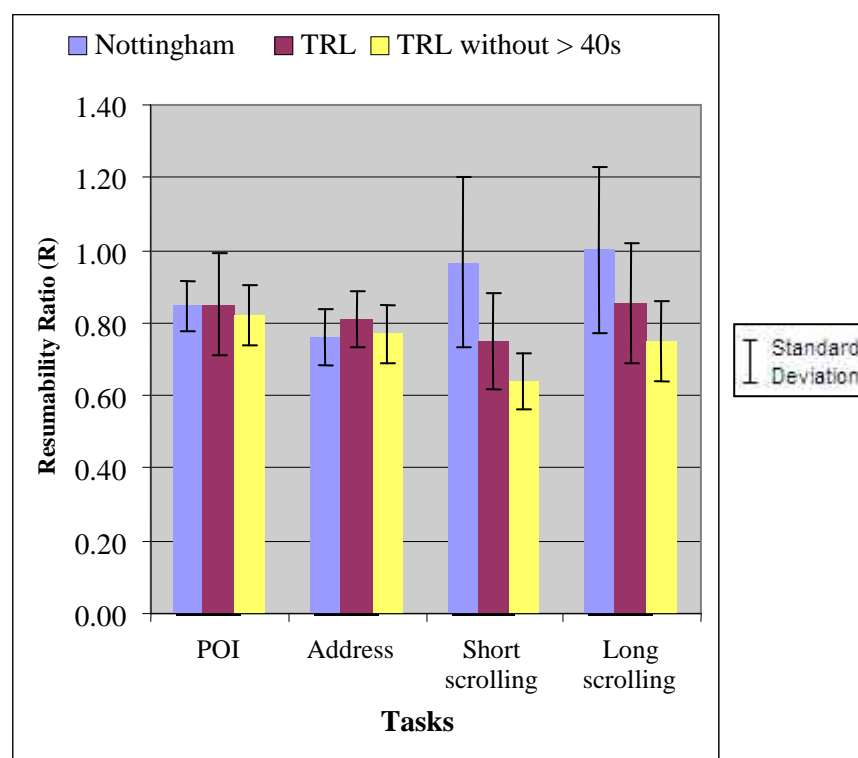

Figure 33: Comparison of University of Nottingham and TRL study results – Mean R

The graphs show that there is considerable agreement between the Nottingham and TRL results for Tasks 1 and 2 for both TSOT and R. This is encouraging as the primary use of the protocol is likely to be in the assessment of fully developed systems.

It is thought that differences in scrolling rates may have contributed to the differences for Tasks 3 and 4 (both across the studies and across the two tasks). In the Nottingham implementation, it was evident that the text scrolled significantly quicker with the short scrolling task (Task 3) compared with Task 4. This was quantified by measuring scroll time for 30 lines of share prices, the results revealed:

- Task 3 (single column scroll time for 30 lines): 12.8 seconds

- Task 4 (three column scroll time for 30 lines): 24.7 seconds

So in the Nottingham implementation the single column scrolling (Task 3) was comparatively quicker than Task 4, which caused some problems. Consequently, it was possible to find the stock code relatively quickly with Task 3 (as the information rapidly scrolled on to the screen). Conversely, it was also more likely for participants to miss the code. This may have contributed to the low TSOT value, yet a high value for R, compared with the WP1 study results.

Likewise, there were differences in scrolling speeds between computers used at Nottingham and TRL for Tasks 3 and 4. At TRL, these were:

- Task 3 (single column scroll time for 30 lines): 7 seconds

- Task 4 (three column scroll time for 30 lines): 14.5 seconds

So scrolling speed was markedly quicker at TRL compared with Nottingham.

Using Task 4 as a further example, the times to scroll 30 lines of share prices are shown below. At TRL, how long it took the 30 rows of shares to appear on the screen in Task 4 was measured across four different computer systems to compare with the Nottingham implementation.

- Nottingham computer = 24.7 seconds (Used in WP4)

- TRL older laptop, Win98 = 14.5 seconds (Used in WP1)

- TRL fairly old laptop, XP = 24.7 seconds

- TRL modestly specified desktop PC, XP = 24.7 seconds

- TRL very high specification laptop, Vista = 21 seconds

So, scrolling speeds can vary significantly between different computers (ranging from less than 15 seconds to almost 25 seconds on the five computers tested).

The overall result indicates that the protocol should require information to be recorded regarding the technical environment in which a task is implemented (additionally, the language in which a task is written should be selected carefully to ensure that a consistent speed of display is possible). This requirement (operationally defined in terms of performance speed) was added to a revised version of the protocol. This will better enable evaluation teams to make meaningful cross-study comparisons.

6.5   Work Package 4(c): Primary Loading Tasks and Occlusion Performance

This study involved consideration of the primary loading task on Occlusion parameters. There is a suggestion that Occlusion has limited validity because there is no apparent sensory or cognitive load during the Occluded periods. The participant is therefore able to consider the IVIS task and plan their next actions in this time. This does not necessarily reflect the conflicting demands of performing the IVIS task whilst driving a real vehicle since the participant must attend to the driving situation in the periods between doing the IVIS task. This raises the question of whether a simulated driving task or "primary loading task" is necessary during the Occlusion interval. The aim of this work package was to determine whether or not the Occlusion measure should be accompanied by a primary loading task. This study investigated the effects of two different primary loading tasks on Occlusion parameters.

In order to determine if a primary loading task (PLT) modifies Occlusion performance (and whether a PLT might be used to help approximate the demands of real driving), the work investigated:

1. The general effect of the presence/absence of PLT on performance of IVIS tasks under Occlusion. So, whether undertaking a loading task whilst performing an IVIS task made any difference to the Occlusion results produced.

2. The relative effects of two different types of PLT- one visual and one auditory. The nature of the PLT is likely to be important because different loading tasks will tap different mental resources. IVIS tasks exploit a faculty of the brain known as working memory, which has separate verbal and visuospatial elements. The PLTs were chosen that compete with the IVIS task for these processing resources. We examined the effect of PLTs that load working memory in different ways to investigate whether there is a differential effect on participant performance in completion of IVIS tasks.

3. The effect of the PLTs on performance on different IVIS tasks under Occlusion. Again (as example IVIS tasks) the four tasks were employed as previously detailed in WP1.

6.5.1   Loading Task Method

Participants

Ten participants were recruited for this study - a subset of the 20 participants used in the reliability study described above. The participants returned to the laboratory at University of Nottingham either one or two days after the reliability study to complete the PLTs.

Participant characteristics were as follows:
- Mean age 28.4 (range 22 to 40).

- 9 male; 1 female.

- All held full driving licences.

- Mean age of holding driving license = 9.5 years (range 3 to 21).

- Mean number of miles driven in last year = 9,500 (range 2,000 to 15,000).

- All were very experienced with using technology.

Study Location, Design and Tasks Used

The study lasted approximately one hour and required participants to undertake the same four tasks as used in the reliability study. In a repeated measures design with counterbalancing, each participant undertook five variants of each task under two experimental conditions:

- Primary Loading task 1 – Visual-spatial task (PLT1-VS)

- Primary Loading task 2 – Auditory-spatial task (PLT2-AS)

Participants were full trained in both of the primary loading tasks before data were collected.

For PLT1-VS, participants were requested to alternate their vision between the primary loading task (PLT) and the secondary tasks. The PLT occurred as part of a PowerPoint slide show that appeared on a 17" monitor directly in front of the driver (see Figure 34). The secondary display was always to the left side of the driver within the vehicle in a fixed location. The slide show progressed between variants of the PLT and a blank slide every 1.5 seconds.

Figure 34: PLT1-VS study set up (for tasks 1 and 2)

An auditory beep occurred as the PLT appeared notifying the participant that he/she must return visual attention to the PLT. Instructions made it clear to participants the importance of maintaining error-free performance on the PLT.

For this PLT (PLT-VS), participants were asked to search an array of four blocks of arrows and indicate verbally whether one is different or not (see the Figure 35). In piloting work, it was established that this simple visual-spatial task was consistently achievable in the 1.5 second period available.
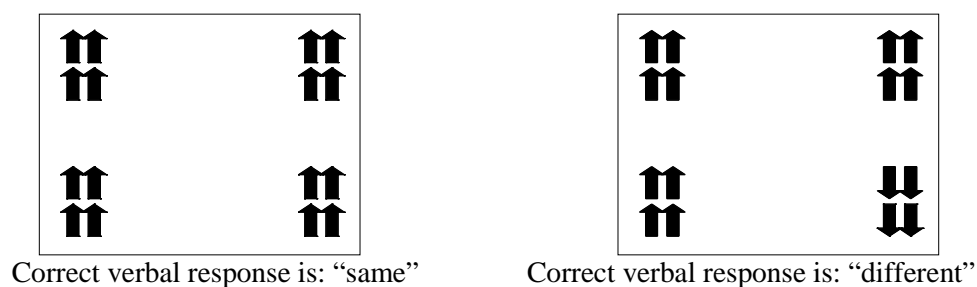

Correct verbal response is: "same"   Correct verbal response is: "different"

Figure 35: Example visual PLT task

For PLT2-AS, participants wore goggles as per the Occlusion protocol and 100ms into the Occluded interval a tone was presented (lasting 500ms). This tone was presented from two speakers in front of the driver and was programmed to be one of 3 possible pitches of equal intervals (262Hz, 330Hz and 392Hz or C4, E4 and G4 on a piano). Participants were asked to state whether they believed the tone they heard was higher, lower or the same in pitch as the one heard in the last Occluded interval. They gave their answer as a verbal response from the second Occluded period onwards.

Values for TTT_{Unoccl} were already available based on the results of the reliability study. In addition, based on data collected using a stopwatch, the following measures were calculated:

- Total Interrupted Vision Time (TIVT) – time required to achieve a secondary task with vision, when interrupted by visual-spatial PLT (so being comparable to the TSOT time obtained in the auditory-spatial condition)

- Total Shutter Open Time with auditory-spatial PLT ($TSOT_{AS}$) – time required to undertake a secondary task when vision permitted, under Occlusion conditions

- Resumability (PLT1) = $TIVT/TTT_{Unoccl}$ (PLT1)

- Resumability (PLT2) = $TSOT_{AS}/TTT_{Unoccl}$ (PLT2)

6.5.2   Loading Task Results

Table 12 shows the mean values (with standard deviations in brackets) for Total Interrupted Vision time (TIVT), Total Shutter Open Time with auditory-spatial PLT ($TSOT_{AS}$) and the Resumability ratios for PLT1 and PLT2.

Table 12: PLT study main results

	PLT1-VS		PLT2-AS	
	TIVT	R(PLT1)	$TSOT_{AS}$	R(PLT2)
Task 1 – POI	10.0 (3.77)	1.07 (0.27)	10.1 (3.36)	1.14 (0.40)
Task 2 – Address	11.6 (2.93)	0.82 (0.19)	15.5 (3.89)	1.11 (0.25)
Task 3 – Short scrolling	5.0 (2.48)	0.77 (0.43)	6.2 (2.54)	0.90 (0.40)
Task 4 – Long scrolling	6.4 (2.92)	0.96 (0.51)	7.2 (3.18)	0.96 (0.36)

Taken together, these results present an unclear picture, particularly with reference to the 'experimental' tasks 3 & 4. Therefore, until further research has been completed the results of such prototype tests are not informative; therefore they have been excluded from the analyses below.

Figure 36 shows the mean values for TIVT and $TSOT_{AS}$ for Tasks 1 & 2 when contrasted with the values obtained for TSOT with the basic Occlusion protocol. Similarly, Figure 37 shows the mean values for R(PLT1) and R(PLT2) for all tasks when compared with the mean value for R according to basic Occlusion.
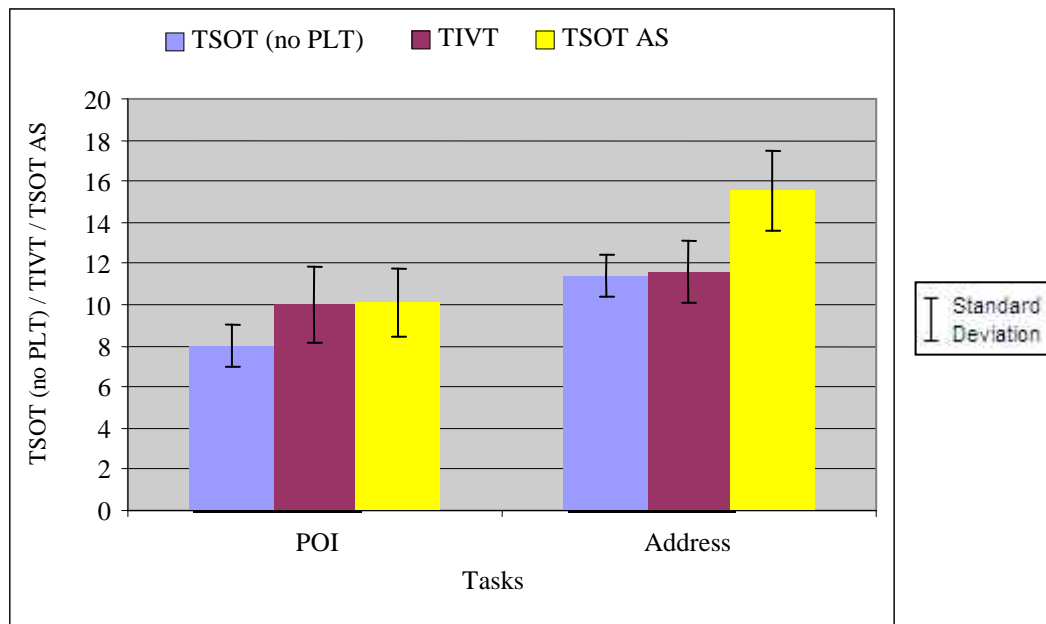
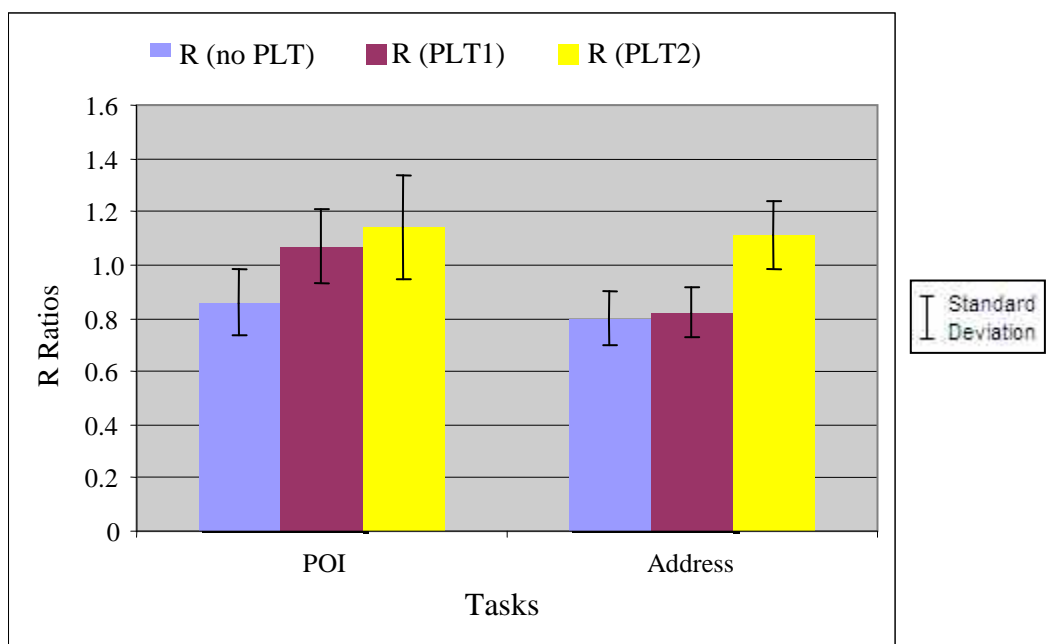**Figure 36: Comparison of TSOT-related results across conditions**

Figure 37: Comparison of R-related results across conditions

Taking Figure 36 and Figure 37 together, some interesting observations can be made:

For the two tasks (POI and Address) both PLTs led to higher values compared with traditional Occlusion – i.e. reduced performance. This effect seems reasonable in terms of the overall demands placed on participants, whereby a participant having to switch between two different tasks would generally be expected to show a lower level of performance than when they had to undertake just one task (separated by 1.5 second Occlusion intervals).

For these two tasks, the Auditory-Spatial PLT (PLT2-AS) generally was associated with higher values than the Visual-Spatial PLT (PLT1). This was particularly the case for the destination entry by address task where $TSOT_{AS}$ was markedly higher than the original TSOT value. This may be due to the Auditory-Spatial PLT interfering with participants' rehearsal of the address, thus disrupting the phonological loop (an address is more complex than a point of interest and, thus, require active rehearsal to be retained in working memory).

In comparing Tasks 1 and 2 based on the Occlusion protocol, it is concluded that the duration of visual demand (TSOT) was markedly higher for Task 2 (address) compared with Task 1 (POI), whereas there was little difference for intensity of visual demand (R). This was not quite the case with the PLT conditions. With PLT1 (visual spatial loading) there was little difference in the duration of visual demand across these tasks (Figure 36), whereas the intensity of visual demand was higher for Task 1 compared with Task 2 (Figure 37). For PLT2 (auditory-spatial loading), the duration of visual demand was significantly higher for Task 2 (address) compared with Task 1 (POI), whereas the intensity of visual demand was equivalent across the two tasks.

6.6   Key Implications

This work package undertook three tasks, the key implications for each are:

i)   The protocol has been successfully developed. This forms a detailed method that contains all the essential information for an Occlusion study to be reliability replicated (Horberry et al., 2007; PPR 259).

ii)   The protocol has been shown to be reliable. For tasks that are fully developed (tasks 1 and 2 in this study) it produced comparable results at TRL and Nottingham. To encompass tasks running on prototype hardware or software the protocol has been extended to ensure that the technical environment in which the testing takes place is more precisely specified (an example of this is differences in presentation speeds of the information presented by the IVIS).

iii)   The presence of a primary loading task whilst performing an IVIS task does make a difference to the Occlusion results. For the two navigation system tasks, performance was reduced (as expected) for both TSOT and R. However, for the two more 'experimental' tasks (scrolling) performance presented a mixed picture (possibly due to participants adopting more optimal search strategies) and future research would be required to elucidate this.

           Whatever, the exact reasons for the differences, it seems that introducing a loading task has few clear advantages in terms of Occlusion performance (and generally creates more inconsistencies). As such, the implication is that using a loading task with the Occlusion technique is not recommended, and the protocol was not modified to include the use of such an additional task. Further work in this area would be required to comprehensively establish the impact of different PLTs on Occlusion reliability and validity.

7   Work Package 5 – Setting Limits

7.1   Background and Study Aims

The objective of this WP was to develop a clear "benchmark" or Demand Reference Level (DRL) that can be used by DfT (and others) to identify in-vehicle devices, tasks or functions that involve an unacceptably high level of demand if used while driving.

Note that a Demand Reference Level using the occlusion technique is not the same as a safe/unsafe threshold for a number of reasons:

- Firstly, the demands of the driving task vary such that there are no absolutely "safe" situations in which secondary in-vehicle tasks can be carried out; in very demanding driving conditions, a brief glance away from the road scene could precipitate an accident.

- Secondly, different drivers have different capabilities for undertaking primary driving and secondary information tasks resulting in different performance and safety levels.

- Thirdly, distraction (and reduction in safety) results from a driver's engagement with the secondary task and different drivers have different strategies – an extreme example would be a driver that decides not to use an in-vehicle device, thus rendering its distraction to zero.

The Demand Reference Level (DRL) can be likened to a blood-alcohol concentration, above which society deems that driving is unacceptably unsafe. By analogy, the DRL would provide a target performance for industrial designers to achieve, and a criterion for the Department for Transport and others to use in legislation, guidelines or consumer information.

7.2   Method and Findings

In terms of a methodology to determine the DRL, the research initially considered direct measurement of driving safety: although performance reductions can be measured (e.g. in a simulator), there are no widely agreed measurement scales or limits of what represents safe (or unsafe) driving for a particular driver. Therefore, we consider it impractical to directly assess the safety of IVIS tasks in terms of driving performance.

We therefore concluded that the Demand Reference Level should be informed by including scientific, social acceptability and commercial views. In all seven data sources were employed:

1. **A meta-analysis of the literature review data,** presented in Section 1, especially focusing on the 'limits' used in other countries. This was largely restricted to Occlusion research, but other work was considered (e.g. the Society of Automotive Engineers (SAE) 15 second rule for the completion of IVIS tasks, as reported in the Driver Metrics Report, 2007). Many research groups have conducted research into completion of tasks with continual intermittent visual Occlusion. Care was taken with data quality as different researchers have used different Occlusion parameters (e.g. shutter open and closed times) so some data normalisation was necessary.

   This found only three major Occlusion cut-off criteria. The most stringent is the Japanese version (with a mean TSOT value of 7.5 seconds). Of course, a UK limit should not follow the most stringent international criteria just for the sake of it but as long as the value is realistic, there is an argument that the UK should use a similar

level to the Japanese; otherwise researchers/designers may simply choose to follow the JAMA value as the most restrictive/univeral. This suggests that the DRL for mean TSOT should be set at around 7.5 seconds.

There is far less agreement over the R value, but, as seen in the driver metrics international report, (where R is used) an R value of 1 is often considered the cut-off.

2. **To supplement the review of literature, personal contact was made with Occlusion experts** in the USA, Japan, Canada and Germany at an ISO TC22 SC13 WG8 meeting (concerning in-vehicle ergonomics) in Italy in May 2007. This was further supplemented by an email questionnaire to key individuals around the world. As such, these two approaches formed a virtual workshop with other world experts in Occlusion to try to reach an international consensus. Email opinion was received from Ford in the USA, Germany (a combined response from several German organisations, for example, BASt & Bosch), Dynamic Research (a leading contractor in this area to the USA government, and active in SAE and ISO committees), a recently retired representative of SAE and a representative of Toyota USA.

As expected, only limited international consensus was revealed. The main points were:
   - Most countries favoured their respective criteria (e.g. the Japanese using the JAMA guidelines).
   - There was no fundamental disagreement about having some kind of measure of the spread of TSOT scores.
   - Finally, there was little agreement for the role of R - some researchers (especially from Europe) value it, whilst others (mainly in North America) do not consider it useful.

3. **Experimental work conducted by TRL and Nottingham (WP1) concerning age**. Combinations of the two main parameters TSOT and R were studied using four tasks.

As seen earlier in this report, the study found some differences between the age categories for the two Occlusion performance measures. In particular the older participants showed a greater spread of scores (especially for TSOT). Overall, these results imply that to obtain minimal inter-subject variability an experiment should ideally use younger/middle aged participants, and especially avoid the 67-76 age group. Of course, the Design Reference Level described later should apply to all age groups. But when undertaking limited Occlusion studies (of 20 or less participants) then using a wider number of older subjects would simply create more variance in the results, so could potentially cloud any effects of comparing different IVIS tasks in different studies (that may be using different groups of participants).

4. **The social acceptability/tasks analysis (keystroke level model)/likely usability issues (in WP2).** IVIS tasks that are already considered acceptable (e.g. tuning a radio), those that are borderline acceptable (e.g. Point of Interest Search with a navigation system) and those that are widely considered unacceptable (e.g. reading scrolling text not related to driving) were considered to determine possible boundary points. Likewise, the usability analysis helped identify, from first principles, those tasks that are better designed to be performed whilst driving.

The expert review showed the fundamental characteristics of the four tasks that are overly distracting and/or may cause usability difficulties: Task 1 had the least usability problems and Task 4 had most. These results both help understand and predict Occlusion scores, and helped establish what particular features of the task make it acceptable/unacceptable to the UK driving population. The social

acceptability findings showed that none of the four example tasks had widespread overall public support for them to be undertaken whilst driving. However, Task 1 was considered the most acceptable, and often considered acceptable by younger drivers. This may imply that the TSOT boundary for whether an IVIS task is acceptable/ unacceptable should be set at approximately the usability score level of Task 1 (shown in Tables 5 and 6 above).

The Keystroke Level task analysis showed that the theoretical task completion times (both Static and Occluded) were closely linked to the results obtained from WP1 for two of the four IVIS tasks (and described why it does not predict scores for the other two tasks). These findings help support, augment, and explain the obtained results and provide a theoretical underpinning concerning, for example, why Task 1 has a lower TSOT score than Task 2. As such it provides another independent data source to illustrate why cut-off criteria set approximately at the level of the mean TSOT results for Task 1 may be appropriate.

5. **Using the Lane Change Task; the work undertook a comparison of impairment of visual demand from using an IVIS with impairment caused by alcohol intoxication at the socially agreed UK limit (WP3).**

This showed that, in general, impairment from an IVIS is greater than driving at the maximum permitted alcohol level. However the types of impairment, of course, differ; for example, alcohol impairment is continuous, whereas IVIS impairment is usually driver initiated and for limited duration. The results may imply that having strict Occlusion cut-off criteria may restrict the overall impairments from IVIS use to be no worse than the overall impairments from driving at the alcohol limit (in terms of a gross measure of driver impairment throughout a whole drive).

When compared with Occlusion task performance (from WP1), the LCT findings showed less ability to discriminate between the four IVIS tasks. This implies that the Occlusion technique is preferable to the alternative LCT approach, so strengthens the case for using the Occlusion protocol.

6. **The results of the reliability analysis (using the developed Occlusion protocol at two research centres) and the findings of the primary loading task (WP4).**

For tasks that are fully developed on specific hardware (i.e. Tasks 1 and 2 in this study) the reliability study found the Occlusion protocol produced comparable results across laboratories. This independent data source shows results are reliable, so provides further confidence of the judgement to set cut-off criteria around the level of Task 1.

With a primary loading task, performance was reduced for the two navigation system tasks using the Occlusion metrics. This result is broadly as expected in terms of task loading (where having more tasks to undertake within the same period of time can reduce performance). Introducing a PLT increases variability and it generates additional experimental difficulties without any obvious benefit; this indicates that the TSOT benchmark should be set without a loading task.

7. **Knowledge of the way in which industry uses design guidelines for the development of new In-Vehicle Information Systems.**

IVIS designers and manufacturers need a procedure that is reasonably easy to apply (as Occlusion, using the newly developed protocol, undoubtedly is) and that cut-off criteria are clear, simple to calculate and are unambiguous. The statistics required to calculate if an IVIS task is acceptable need to be reasonably undemanding. Also,

designers require criteria that can be applied early in the product design lifecycle, and that allows them to make any necessary changes to the IVIS task and then re-test it against the criteria. Finally, the criteria should be capable of being applied to original manufacturer fitted equipment, nomadic devices and other off-the-shelf systems.

7.3   The Demand Reference Level

A Demand Reference Level (DRL) has been requested by DfT (and others) to identify in-vehicle devices, tasks or functions that involve an unacceptably high level of demand if used while driving. The DRL proposed here is based on the Occlusion parameter TSOT. The diagnostic value of R is the subject of much debate, although there is a view amongst some researchers (e.g. Pettitt et al, 2006) that R greater than 1 is unacceptable. Given the unclear pattern for R from the results of this study, and the general lack of international support for its use, it was decided that R should not be used in the DRL, but that it should, nevertheless, be included in the protocol and reported by experimenters.

Before discussing the absolute level, it is necessary to explore the distribution of TSOT values. Typically, Occlusion data are not normally distributed. The skewed nature of the data occurs because most IVIS tasks cannot be completed in less than a certain amount of time[§§], whereas there can often be a long 'tail' where some participants take longer to perform tasks. A typical distribution example for TSOT is shown in Figure 38:
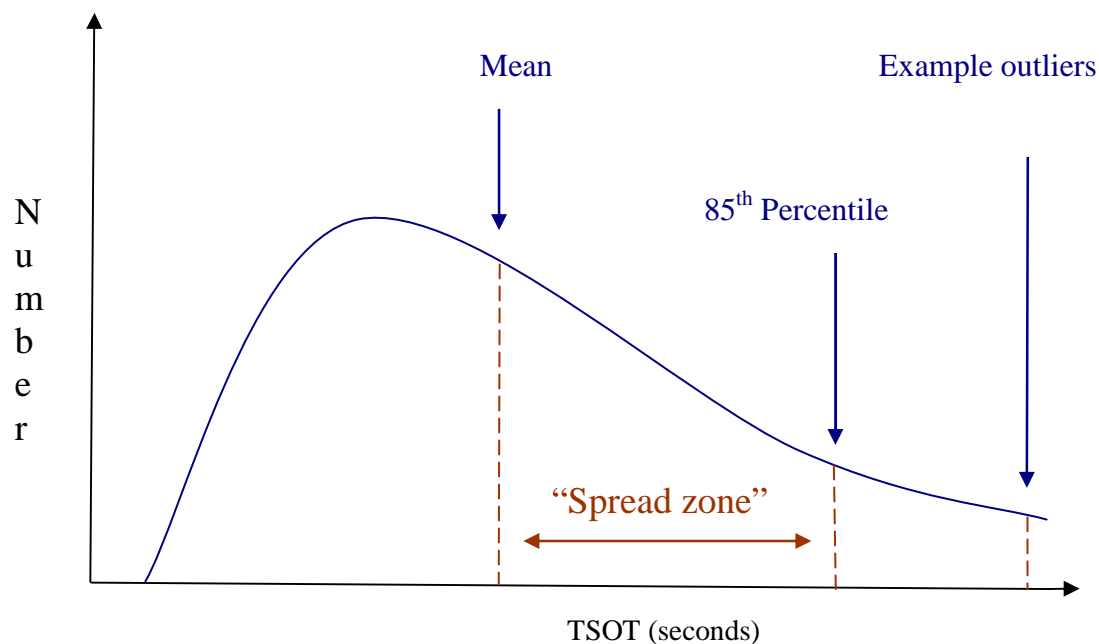

Figure 38: Typical distribution of TSOT for a task

Based on the discussions of Section 7.2 and the research team we believe that a TSOT at the JAMA level of 7.5 seconds (mean) is broadly appropriate as a reference level. However, from our appreciation of the distribution of TSOT, statistical considerations imply that the mean value is a poor representation of the data. Arguably the 85th percentile provides a 'better' single value as it emphasises the tail of the distribution more.

Our recommendation is that the DRL takes account both of the absolute values of TSOT and the spread of the distribution. As a single absolute value we acknowledge that the mean is the

---

[§§] Indeed, the occlusion method is inappropriate for testing tasks that take less than 5 seconds

most widely used and reported value in occlusion research and so (despite statistical arguments for using 85th percentile or transforming the data to remove the skew) we can agree with the use of the mean.

To take account of the distribution of TSOT values, we propose the use of a spread parameter which is the squared difference between the 85th percentile and the mean, normalised by dividing by the mean:

$$\text{i.e. } \text{spread} = \frac{(85^{\text{th}} \text{ percentile} - \text{mean})^2}{\text{mean}}$$

This can be thought of as similar to a standard deviation term but one more appropriate for a skewed distribution of values.

The demand level produced by adding the mean TSOT and the spread (DL_{Occl}) allows for a trade-off between absolute value and spread. Our proposal is below:

> For an IVIS task to 'meet' the Demand Reference Level, the TSOT values produced (mean + 'spread' combined) need to be below 8.0 seconds. The formula is:
>
> $$\text{Mean} + \frac{(85^{\text{th}} \text{ percentile} - \text{mean})^2}{\text{mean}} < 8.0 \text{ s} \qquad \text{(Equation 1)}$$

It is considered that using the mean (or, indeed, 85th percentile) alone would not adequately account for the spread of scores, so the spread is a valuable statistic, especially when only a low number of participants are used for Occlusion testing. But it is still simple to calculate, as it requires only two values: the mean and 85th percentile. Given that the data are skewed, purely using standard deviation as a spread is not appropriate.

Using these values, the process to assess if a task meets the Demand Reference Level (and is therefore considered to be 'acceptable' by this Occlusion criteria) is shown in the simple flowchart (Figure 39) below.
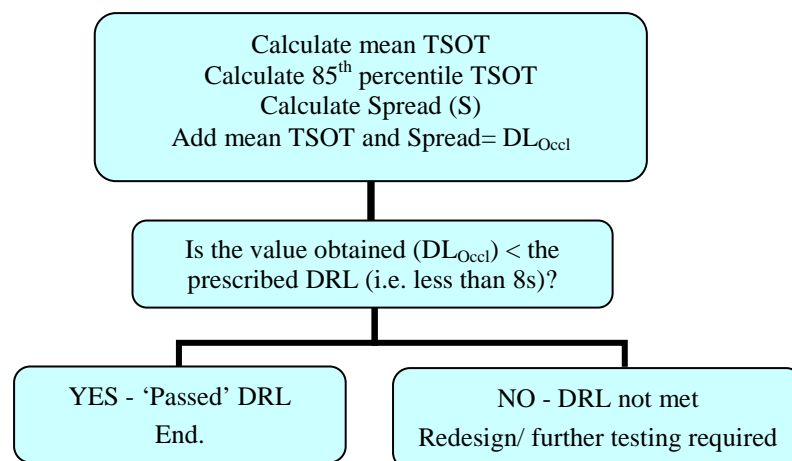
Figure 39: Graphical representation of the process to evaluate if an IVIS task is acceptable or unacceptable using Occlusion

Applying the DRL to the results obtained in WP1 with 60 people for the four example tasks used in this research shows that three of the four tasks would not have met the DRL. Task 1, the POI, is a borderline pass (with a TSOT with spread (i.e. DL_{Occl}) of just less than 8s), but

the other three tasks would have been clear failures, with all producing a DL_{Occl} of over 10s (see Table 13).

Table 13: Mean TSOT and R values for Tasks 1 to 4

Task	Mean TSOT (s)	85th Percentile (s)	Spread (s)	DL_{Occl} (s)
Task 1	7.63	9.06	0.27	7.90 (pass)
Task 2	11.77	14.96	0.86	12.64 (fail)
Task 3	10.94	13.21	0.47	11.41 (fail)
Task 4	11.62	16.79	2.3	13.92 (fail)

These results are fundamentally in agreement with the work undertaken earlier in this project. For example:

- The social acceptability finding in WP2 showed that none of the four tasks had widespread overall public support for them to be undertaken whilst driving (although Task 1 was considered the least unacceptable, and often considered acceptable by younger drivers).

- The expert review in WP2 showed the four tasks had a range of significant usability issues (but, again, Task 1 had the least severe problems, and was quantified as the most usable).

- The results of the alcohol study in WP3 showed that in some cases the effect of using IVIS can be more detrimental on simulated driving performance than driving at the highest legal level of alcohol intoxication. This suggests that the research and policy community need to focus more strongly on the safety implications of in-vehicle distraction.

- Given the potential safety issues with excessive visual demand from IVIS devices, there is growing international recognition that stringent criteria are important. The DRL proposed is in line with the most stringent guidelines used elsewhere in the world (the JAMA guidelines as evidenced in the literature review in WP1).

- There was a moderate degree of international consensus; for example, on the need to examine the spread of TSOT scores (as evidenced in WP5).

As such, there is converging evidence from other sources that supports a DRL for TSOT. The situation concerning R is mixed. Our work here does not support the use of R but we recommend that it is measured and reported.

7.4   International Opinion

As reported above, although there is little international agreement on any specific Occlusion limit, the response from Ford USA compared the proposed DRL to a major study undertaken in the USA with 13 different IVIS tasks (Tijerina, 2007, personal communication). These tasks included radio tuning, destination entry, manual dialling of a mobile phone and map reading. The results showed that the proposed DRL agreed with Ford's classification of whether a task had a lower or higher workload (roughly approximated to a pass/fail on the DRL) for **12 of the 13 tasks**. The one task that the two approaches disagreed upon was a 'mid-difficulty' task: the proposed DRL found it should not be allowed, whilst the Ford approach measured a lower workload (and so suggested it should be permitted).

This level of agreement with previous American results gives further confidence in the DRL developed here.

7.5   Key Implications

Using the Occlusion protocol and the DRL created by this project allows manufacturers, researchers and regulators to both undertake repeatable Occlusion research and to compare the visual demand from using an IVIS with a Design Reference Level (DRL). This process can be used both for original manufacturer fitted equipment and for nomadic and other off-the-shelf systems.

The DRL presented here is in the form of criteria and a flowchart, whereby the Occlusion metrics of TSOT and its distribution is employed. If an IVIS system exceeds the reference level its interface should generally be redesigned and/or further testing should be undertaken.

Such an approach could be included within a 'star rating' system, whereby the DRL rating could be used for consumer information which might also include factors such as button size, security of mounting (for a nomadic IVIS device) and crashworthiness.

An IVIS task that meets the DRL doesn't necessarily imply that it is 'safe' for a driver to perform whilst a vehicle is in motion, but it does suggest the visual demand required during the interaction is within a benchmark limit. IVIS tasks that do not meet the DRL generally involve an excessive level of visual demand. However, there might be situations where such tasks are still acceptable based on other grounds (such as long tasks that are easily interruptible and impose only a low intensity of visual demand). In such cases, the justification for such tasks being acceptable should ideally also be evidence-based, for example established by other repeatable metrics (e.g. within driving simulator evaluations). As such, the DRL is a guideline rather than an absolute limit, and it is recommended that more DRL data is collected on as many common IVIS tasks as possible to further support the values proposed (in part this will only come over time as use of the Occlusion method becomes more widespread).

The DRL was informed by many converging data sources and it is the contention of this project that it represents a reasonable criterion to assess if an IVIS presents an acceptable level of visual demand when driving.

8   Conclusions

This project has successfully collected and analysed a great deal of valuable Occlusion-related data. For example, it undertook the largest Occlusion study to date in the world (the age experiment reported in WP1) and reported that occlusion data were generally skewed). It analysed several illustrative tasks from first principles (e.g. their likely usability issues and a task analysis of the theoretical time to complete them) and it benchmarked the performance decrements of using an IVIS whilst driving to impairments from alcohol intoxication. The research also conducted the first assessment of reliability for the Occlusion protocol and explored the issues involved in utilising primary loading tasks for Occlusion.

The work has received a great deal of international attention (for example, invitations to speak about the research at international events) and has generally helped promote the Occlusion method. As such, it has helped develop scientific knowledge regarding the use of the Occlusion method to assess visual demand from IVIS tasks.

On a more practical level, the research has developed both a comprehensive Occlusion protocol and a Demand Reference Level. It is believed that using both the protocol and DRL will help identify IVIS tasks that require excessive visual demand. This should aid road safety by helping to reduce the use of such 'inappropriate' IVIS tasks by providing firm evidence to support policy, design guidelines and consumer information.

9   Recommended next steps

We recommend that the protocol is disseminated widely. One route is to publish the protocol as an ISO technical report (to support the ISO standard). This would require acknowledgement of an international consensus and it would give the protocol maximum exposure internationally.

Second, the DRL was developed using a wide range of data sources. Taken as a whole these largely agreed and produced converging evidence. Following the widespread dissemination of the DRL, it is recommended that it is periodically reviewed (especially as new IVIS devices appear on the market and new research findings emerge).

Finally, it is recommended that further data is collected to benchmark the visual demand for a wide range of IVIS tasks against the proposed DRL. Such research would also investigate the relationship between occlusion-related measures and those taken from more established methods, such as simulator and road trials. Such research could assist DfT in understanding the range of interface types that could be considered to be unacceptable.

Acknowledgements

References

Asoh, T., & Iihoshi, A., 2005. A study using the Occlusion technique to evaluation visual distraction, Japan Automobile Research Institute & Japan Automobile Manufacturers Association Inc, Human factors and Ergonomics Society Europe Chapter Annual Meeting 2005.

Baumann, M., Keinath, A., Krems, J.F., & Bengler, K., 2004. Evaluation of in-vehicle HMI using Occlusion techniques: experimental results and practical implications, Department of Psychology, Chemnitz University of Technology, Chemnitz D-09107, Germany, and, BMW Group Forschung and Technik, Munich D-80788, Germany, Applied Ergonomics, Volume 35, Issue 3, May 2004, Pages 197-205, The Occlusion Technique.

Bengler, K., 2003. Occlusion – A Method to Assess Dialogue Interuptability, A Summary of Results, BMW Group, Presented at HFES, Nov 11 2003, San Diego.

Burnett G, Pettitt, M., & Stevens. A., 2005. Defining Driver Distraction. 12th World congress on Intelligent Transport Systems, San Francisco, November 6-10 2005. CD Rom Custom Number: 2586, ERTICO, B-1050 Brussels. www.ertico.com

Burns, P. (Transport Canada), 2006. Occlusion research in Canada. Presentation during the driver metrics workshop, Ottawa, October 2006.

Driver Metrics Workshop, 2007. Workshop draft summary. Downloaded 23rd May 2007 from: http://ppc.uiowa.edu/drivermetricsworkshop/

Forbes, N. & Burnett, G.E., 2007. Investigating the contexts in which in-vehicle navigation system users have received and followed inaccurate route guidance instructions. In 3rd International Conference in Driver Behaviour and Training, to be held in Dublin, 12-13 November 2007.

Horberry, T., Anderson, J., Regan, M., Triggs, T. & Brown, J., 2006. Driver Distraction: The Effects of Concurrent In-Vehicle Tasks, Road Environment Complexity and Age on Driving Performance. *Accident Analysis and Prevention* 38(1), pages 185-191.

Horberry, T., Stevens, A., Cotter, S ., Robbins, R. & Burnett, G., 2007. Occlusion Protocol. TRL Published Project Report (PPR) 259.

International Organization for Standardisation. ISO 16673: 2007. Road vehicles - Ergonomic aspects of transport information and control systems - Occlusion method to assess visual demand due to the use of in-vehicle systems.

Mattes S., 2003. The lane change task as a tool for driver distraction evaluation. In H. Strasser, K. Kluth, H. Rausch and H. Bubb, (Eds), Quality of work and products in enterprise of the future. Ergnomia, Stuttgart.

Niiya, K., 2000. Trial evaluation results on navigation operation using Occlusion technique. Discussion paper: ISO/TC22/SC13/WG8 Visual distraction TF. Toyota Motor Corporation. November 13, 2000. ISO/TC22/SC13/WG8/N289.

Pettitt, M.A., Burnett, G.E. & Stevens, A., 2006. Extending the keystroke level model (KLM) to assess the visual demand of an in-vehicle information system (IVIS). Proc. World Congress on Intelligent Transport Systems 2006.

Pettitt, M. A., Burnett, G., Bayer, S. & Stevens, A., 2006. Assessment of the occlusion technique as a means for evaluating the distraction potential of driver support systems. IEE Proceedings Intelligent Transport Systems - Volume 153, Issue 4, p. 259-266 December 2006.

Stevens, A., Bygrave, S., Brook-Carter, N. & Luke, T., 2004. Occlusion as a technique for measuring in-vehicle information system (IVIS) visual distraction: a research literature review. Transport Research Laboratory (TRL) report no. TRL609. TRL, Crowthorne, Berkshire, United Kingdom.

Stowell, A.R. & Stowell, L.I., 1998. Estimation of blood alcohol concentrations after social drinking. *Journal of Forensic Sciences*. 43 (1), pp. 14-21.

Tijerina, L., Johnston, S., Parmer, E., Winterbottom, M.D., & Goodman, M., 2000. Driver distraction with wireless telecommunications and route guidance systems, National highway and traffic safety administration, Technical report No. DOT HS 809-069.

Tijerina, L., 2007. Personal communication.

Transport Canada, 2006. R Occlusion Research – Road Safety and Motor Vehicle Regulation, Transport Canada Presentation, AAM driver focus task force, Detroit June 2006.

Weir, D.H., Chiang, D.P. & Brooks, A.M., 2003. A study of the effect of varying visual Occlusion and false duration conditions on the driver behaviour and performance or using secondary task human-machine interface. ADE technical papers series 2003-01-0128. 2003 SAE world congress, Detroit, Michigan, March 3-6, 2003.

Abstract

This report addresses the issues required to establish an effective Occlusion methodology for the assessment of In-Vehicle Information Systems (IVIS). The research develops a robust visual demand measurement protocol based on Occlusion, and proposes criteria for the acceptability of different tasks. This provides information on which IVIS tasks should not be undertaken while driving. This work was undertaken in five work packages:

1. The impact of age on performance of IVIS tasks whilst applying the Occlusion technique

2. Consideration of IVIS tasks and expert evaluation

3. Benchmarking IVIS tasks to another impairment inducing factor (alcohol) and performance measures (Lane Change Task)

4. Production of a detailed protocol for assessment of IVIS tasks, performing a reliability study of Occlusion procedure and testing the influence of a primary loading task

5. Setting limits: determining performance limits based on the Occlusion measures