

Evaluation and comparison of the Human Machine Interface in vehicles in the Australian market Final report



The Australian Automobile Association is the nation's peak motoring body, representing the state-based motoring clubs (the NRMA, RACV, RACQ, RAA, RAC, RACT and the AANT) and their over 10 million members. It is an apolitical and technology-neutral advocate for federal transport policy that improves safety, affordability, and mobility.



1 Table of Contents

2	<i>Acknowledgements</i>	6
3	<i>List of Abbreviations</i>	7
4	<i>Executive Summary</i>	8
4.1	Problem and Context.....	8
4.2	Objectives and Key Findings.....	8
	Practical Implications.....	10
5	<i>Introduction and Project Objectives</i>	11
5.1	Development of IVIS distractibility assessment protocols for an Australian context.	12
6	<i>Project Governance</i>	15
7	<i>Governance Structure and Research Team</i>	15
8	<i>Project Methodology</i>	16
8.1	General assessment protocol and data collection	16
8.2	Participants	17
8.3	In-Vehicle Infotainment Systems and Vehicles.....	17
8.4	IVIS functions - <i>Task Types</i>	19
8.5	Modes of IVIS interaction.....	21
8.6	Study Design.....	21
8.6.1	General IVIS Assessment Procedure	22
8.7	Data Analysis and Modelling.....	23
8.8	Data cleaning.....	25
9	<i>Phase 1: On-Track Drive - Assessment of cognitive and visual demand of IVISs in 'real-world' conditions</i>	25
9.1	Procedure	25
9.2	The Detection Response Task.....	26
9.3	Eye tracking	27
9.4	Vehicle Telemetry.....	28
10	<i>Results</i>	29
10.1	DRT Data preparation	29
10.2	On-Track Drive Pre-Corrected Data - DRT	29
10.3	Modelling On-Track DRT data.....	31
1.1.	Eye-tracking Data preparation	34
10.4	Eye-tracking data	34
10.5	Vehicle Telemetry Data preparation	34
10.6	Vehicle Telemetry Data.....	34
11	<i>Discussion – Phase 1 On-Track Test</i>	37

12	<i>Phase 2: Assessment of cognitive and visual demand of IVISs in a stationary setting.</i>	39
13	<i>Method – Phase 2</i>	39
13.1	Visual Occlusion Test	39
13.2	VOT Procedure	40
13.3	Simulated Driving tasks - <i>Tracking Task</i> with in-built visual DRT	41
13.4	Tracking Task Procedure	42
14	<i>Results</i>	42
14.1	VOT Pre-Corrected Data	42
14.2	Tracking Task Pre-Corrected Data	43
14.3	Modelling Off-Track Stationary Data	46
15	<i>Stationary Off-Track Tests as valid substitutes of On-Track assessment</i>	50
16	<i>Phase 3: A practical cost-effective solution for assessing distractibility of an IVIS</i>	53
17	<i>Method</i>	54
17.1	Defining the number of IVIS interactions (steps)	54
17.2	Procedure	56
18	<i>Results</i>	56
18.1	Validating the Step Metric Across On-Track and Off-Track Tests	56
18.2	Evaluating Step Counts as a Predictor of Driver Distraction	57
18.3	Model selection	58
18.4	Model prediction	59
19	<i>Discussion – Phase 3</i>	62
20	<i>Quantifying Interaction-Driven Distraction Costs to Inform an IVIS Safety Rating System</i>	64
20.1	Estimating the distraction cost of IVIS interactions	64
20.2	Deriving IVIS Interaction Thresholds and Their Mapping to IVIS Star Ratings	65
20.3	Procedure for Assessing IVIS Distractibility Using the Step Metric	69
20.3.1	Identify the IVIS task	70
20.3.2	Perform multiple repetitions of each task type	70
20.3.3	Count all driver–IVIS interactions	71
20.3.4	Document any system anomalies	71
20.3.5	Calculate the mean step count for each task type	72
20.3.6	Calculate the overall mean step count for the IVIS	72
20.3.7	Determine the IVIS star rating	72
21	<i>Conclusion, recommendations, and future steps</i>	73
21.1	Applications of Key Findings	73
21.2	Future Adaptability of the IVIS Distraction Assessment Protocol	74
21.3	Requirements for model re-validation	75
21.4	Implications of Task Type Consistency across IVISs for the Star-Rating Procedure	75
21.5	Step-Count Temporal Validity	76
21.6	Ensuring future impact	76
22	<i>References</i>	79

23	<i>Appendix 1 - Example “Run Sheet”</i>	82
24	<i>Appendix 2 - Full set of IVIS trials</i>	85
25	<i>Appendix 3 – Analysis of “Task Order”</i>	87
26	<i>Appendix 4 – Correlations of all Distraction Outcome Variables</i>	88
27	<i>Appendix 5 – The cost of IVIS interactions: step count Predictions of distraction outcomes</i>	89
28	<i>Appendix 6 – Linear mixed effects model for Drive DRT Misses</i>	94
28.1	Overall IVIS Effects (Collapsed Across Tasks)	94
28.2	Main Findings: IVIS Effects	94
28.3	Model Specification	94
28.4	Model Fit Statistics	95
28.5	Fixed Effects: IVIS Main Effects	95
28.6	Fixed Effects: IVIS × Task-Type Interactions	96
29	<i>Appendix 7 – Linear Mixed-Effects Model: Drive DRT Response Time</i>	99
29.1	Overview of Model	99
29.2	Main Findings: IVIS Effects	99
29.3	Model Specification	100
29.4	Model Fit Statistics	100
29.5	Fixed Effects: IVIS Main Effects	100
29.6	Fixed Effects: IVIS × Task-Type Interactions	101
30	<i>Appendix 8 - Linear Mixed-Effects Model — Visual Occlusion Test – Total Shutter Open Time (TSOT)</i>	104
30.1	Overview of Model	104
30.2	Main Findings: IVIS Effects	104
30.3	Model Specification	105
30.4	Model Fit Statistics	105
30.5	Fixed Effects: IVIS Main Effects	105
30.6	Fixed Effects: Vehicle × Task-Type Interactions	106
31	<i>Appendix 9 - Linear Mixed-Effects Model: Tracking Task Tracking Error</i>	109
31.1	Overview of Model	109
31.2	Main Findings: IVIS Effects	109
31.3	Model Specification	110
31.4	Model Fit Statistics	110
31.5	Fixed Effects: IVIS Main Effects	110
31.6	Fixed Effects: IVIS × Task-Type Interactions	111
32	<i>Appendix 10 – Linear Mixed-Effects Model – Tracking Task – visual DRT Response Times</i>	114
32.1	Overview of Model	114
32.2	Main Findings: IVIS Effects	114

32.3	Model Specification	114
32.4	Model Fit Statistics	115
32.5	Fixed Effects: IVIS Main Effects	115
32.6	Fixed Effects: IVIS × Task-Type Interactions	115
33	<i>Appendix 11 – Linear Mixed-Effects Model – Tracking Task vDRT Misses.....</i>	<i>118</i>
33.1	Overview of Model.....	118
33.2	Main Findings: IVIS Effects (Initiate Call Reference Task)	118
33.3	Model Specification	118
33.4	Model Fit Statistics	118
33.5	Fixed Effects: IVIS Main Effects	119
33.6	Fixed Effects: IVIS × Task-Type Interactions	119
34	<i>Appendix 12 - Vehicle Telemetry Model Output.....</i>	<i>122</i>
34.1	Linear Mixed-Effects Model: Mean Lateral Acceleration (g)	122
34.2	Linear Mixed-Effects Model: Mean Speed Delta (m/s change at 100 Hz).....	122
35	<i>Model Selection and Validation for the Step-Count Metric.....</i>	<i>124</i>
36	<i>Appendix 13 - Rationale for Establishing Core Task Domains.....</i>	<i>125</i>
36.1	The Four Core Task Domains	125
	1. Phone Operations	125
	2. Messaging / Notifications	125
	3. Media Control	126
	4. Navigation / Destination Entry.....	126
36.2	Why Core Tasks Are Defined Theoretically (Not Empirically)	126
36.3	What Happens When a Vehicle Lacks a Core Task?	127

2 Acknowledgements

This research was funded in the amount of \$1,233,443.97 AUD by the Australian Automobile Association (AAA) as a part of the AAA Road Safety Research Program and the research approach and findings do not necessarily reflect the views of the AAA or its member clubs. In accordance with the contractual arrangements for this research, the AAA has ownership of the Developed IP and De-identified data from the research. Any use beyond this report, for any purpose (including by the research team), requires the express written approval of the AAA.

We acknowledge and thank all participants for their time, interest and involvement in the project. We would like to acknowledge our research partners ABMARC who provided their engineering expertise and vehicle telemetry equipment, access to the Australian Automotive Research Centre proving ground, safety drivers, help with vehicle procurement and research participant recruitment. Special thanks to Andrew Freemantle, Simon McRae, David “Dumpy” Thompson, Duncan Cunningham, Marvin Holt and Adrian Castven for their invaluable help throughout the data collection stage at the AARC.

We would like to acknowledge our collaborators, Professor Dave Strayer and Dr Joel Cooper, whose expertise was central to this project. They played key roles in developing the test protocol, study design, and analysis plan, and provided valuable feedback throughout the data-collection phase. We also thank Professor Mike Regan, Dr Prasannah Prabhakaran, and their colleagues for their substantial contributions to the test protocol and study design prior to the University of Newcastle’s involvement. Their foundational work was instrumental in shaping the direction and quality of the project.

We thank the AAA, in particular National Program Manager Michelle Quester and Research Manager Dr Kate O’Donnell for their guidance and advice throughout the project.

The data-collection phase would not have been possible without the dedication and hard work of the University of Newcastle research team. We thank Augustine (Tin) Nguyen, Sophie Withers, Dr Alexander David Thorpe, and Dr Rachael Wynne for their valuable assistance. We acknowledge Dr Angus McKerral for his contributions, including conducting the pilot study and helping to develop the study design and test protocol, which were instrumental to the project’s success.

3 List of Abbreviations

AARC:	Australian Automotive Research Centre
ANCAP:	Australasian New Car Assessment Program
ADAS:	Advanced Driver-Assistance Systems
AOI:	Area of Interest
DMS:	Driver Monitoring System
DRT:	Detection Response Task
IVIS:	In-Vehicle Infotainment System
OEMs:	Original Equipment Manufacturers
OMS:	Occupant Monitoring Systems
RT:	Response Time
TSOT:	Total Shutter Open Time
TSCT:	Total Shutter Close Time
UoN:	University of Newcastle
vDRT:	Visual Detection Response Task
VOT:	Visual Occlusion Test
VOG:	Visual Occlusion Goggles

Disclaimer

1. The “Evaluation and comparison of the human machine interface in vehicles in the Australian market” research project undertaken by the University of Newcastle in conjunction with ABMARC and Red Scientific.
 2. Any information, data or recommendations etc. (Content) contained within this Australian Automobile Association (AAA) ACN 008 526 369 (AAA) research report (Report) is for information purposes only.
 3. As the AAA has no control or foresight on how a consumer of a Report may use or apply the Content, AAA nor its contractors will be responsible for any claims, damages, costs or losses following any reliance placed on any Content of the Report.
 4. The Content of the Report is based on extensive and careful research. As information is constantly changing, the Content of the Report is based on the best available information obtainable at the time. Due to factors beyond our control, information could change significantly, suddenly and without notice. The Content of the Report is therefore only relevant as at the time of publication. The AAA nor its contractors warrant the currency or accuracy of the Content thereafter.
 5. It is your responsibility to ensure that any Content of the Report is relevant and appropriate to you. The AAA nor its contractors provide any warranty nor carries any responsibility in relation to this.
-

4 Executive Summary

4.1 Problem and Context

Driver distraction is a major road safety challenge in Australia, contributing to approximately 16% of serious-injury crashes. Modern In-Vehicle Infotainment Systems (IVIS) are becoming increasingly complex, requiring multi-modal interactions that divert drivers' visual attention (looking away from the road), cognitive attention (thinking about the system rather than driving), and manual control (removing hands from the steering wheel). Despite this growing risk, IVIS design remains largely unregulated.

The distractibility of infotainment systems has not been evaluated under Australian driving conditions. Local factors such as right-hand drive layouts, road geometry, traffic environments, and typical driver behaviour differ markedly from those in regions where most international research has been conducted. As a result, there is currently no reliable, evidence-based way for Original Equipment Manufacturers (OEMs), ANCAP, regulators, or consumers to determine how distracting a given IVIS may be under real-world Australian conditions. This lack of localised assessment limits informed decision-making, constrains effective policy development, and hinders the design of safer vehicle technologies.

EuroNCAP and ANCAP are actively considering how to expand driver-monitoring evaluations within their safety protocols. Throughout this research project, methods and preliminary findings have been shared with these agencies to help inform the development of emerging, evidence-based standards for assessing IVIS distraction.

4.2 Objectives and Key Findings

The overarching goal of this research was to develop, test, and validate a practical distraction-rating system for human-machine interfaces that could be applied consistently across both current and emerging IVIS technologies. The research demonstrated that infotainment systems become more or less distracting primarily based on how many “steps” or interactions a driver must perform to complete an IVIS task.

This study developed three objective methods to assess the IVIS distraction potential of vehicles available in the Australian market under Australian driving conditions. The methods comprise on-track testing, stationary vehicle testing, and a step-based interaction metric

developed through this study's statistical modelling. Each method differs in its instrumentation requirements and the level of technical expertise needed.

Practical Implications

The research has practical implications for different audiences.

Assessment Method	Description	Interested Parties	Cost	Skill	Potential Applications
On-track Testing	Real-world driving on a closed track to measure distraction.	Automotive manufacturers. EuroNCAP and ANCAP Human factors and road safety researchers.	High	Highly specialised researchers and data analysts	Methods enable testing to be repeated in different countries / regions. Findings can inform vehicle and equipment design. Findings can inform new and existing vehicle safety assessments.
Stationary Vehicle Testing	Simulated driving environment without vehicle movement to assess user interaction.	Human factors and road safety researchers.	Medium	Specialised researchers and data analysts	Methods and findings can inform future research.
Step-based interaction metric	The step metric quantifies an IVIS task by counting the number of discrete interactions required to complete it.	Fleet managers. Organisations. Drivers	45-50 Minutes	None	Method enables a member of the public to assess the distractibility of a vehicle's IVIS. Can be used as part of vehicle purchase decisions.

5 Introduction and Project Objectives

In Australia, road trauma remains a major public-health issue, with approximately 100 deaths occurring each month and a comparable number of people hospitalised each day as a result of road-crash injuries (Australian Institute of Health and Welfare [AIHW], 2023; Bureau of Infrastructure and Transport Research Economics [BITRE], 2023). Every year road trauma costs the national economy almost \$27 billion and brings tragedy into the lives of thousands of Australians (Steinhauser et al., 2022). Road safety is a national crisis that demands real leadership, collaboration and solutions.

Driver distraction is a significant challenge for road safety, resulting in a high number of crashes and fatalities (e.g., Young, Regan & Hammer, 2007), and secondary tasks have been identified as the most common (68.7%) distraction activity by road users (McEvoy, Stevenson & Woodward, 2006). Research has demonstrated that drivers spend between 30%-40% of their driving time engaged in secondary tasks such as handling or talking on a mobile phone or adjusting general vehicle controls such as opening or closing windows (Metz, Landau & Just, 2014).

As drivers, we have limited attentional resources, most of which must be devoted to monitoring the road environment and operating the vehicle. Modern Human–Machine Interfaces (HMIs), particularly In-Vehicle Infotainment Systems (IVIS), increasingly compete for these resources by providing access to communication, navigation, entertainment, and other information services through complex multi-modal interactions. Engaging in these secondary IVIS tasks draws attention away from driving, contributing to visual distraction (e.g., off-road glances), cognitive load (as mental resources are stretched), and manual distraction (from physically interacting with controls), all of which degrade driving performance (Choudhary & Velaga, 2017; Strayer, Watson & Drews, 2011). Although voice-activated dialogue systems were introduced to help reduce visual and visuo-manual demands, speech-based interaction itself imposes substantial cognitive load (Strayer et al., 2015). As the functionality of smartphones and portable devices continues to evolve, the volume of information available to drivers grows, further increasing vulnerability to distraction and impairing hazard detection, promoting task-neglect, and prolonging off-road glances (Regan et al., 2009; Crundall, Large & Burnett, 2016; Laing & Lee, 2010).

Importantly, the technologies incorporated into new vehicles in Australia vary considerably in their potential to distract drivers. Although many systems offer similar functions, manufacturers differ substantially in how these technologies are designed and implemented. These design choices mean that some vehicle cockpits place far greater demands

on a driver's attention than others, making them more likely to divert focus away from tasks essential for safe driving. This variability has been demonstrated in a series of research studies conducted by Professor David Strayer, Dr Joel Cooper, and colleagues at the University of Utah in the USA. Their work has resulted in distraction safety ratings for more than 40 vehicle models and highlights the wide range of distraction potential across systems (e.g., Strayer et al., 2017; Strayer, Cooper, Turrill, Coleman, & Hopman, 2015).

While findings from studies conducted in the USA are valuable and provide the basis for the current study, they are not directly transferable to the Australian context. Australian driving conditions differ from those in other regions due to factors such as steering wheel position (right-hand vs left-hand drive), road design, traffic patterns, and driver behaviour. Furthermore, ongoing innovation by automotive designers, combined with consumer demand, ensures that IVIS-related distractions will remain prevalent in future vehicles. At present, there is no reliable means for Original Equipment Manufacturers (OEMs), ANCAP, or drivers to assess how distracting a particular system may be under real-world Australian conditions. This gap presents a significant challenge to informed decision-making.

5.1 Development of IVIS distractibility assessment protocols for an Australian context

Several research initiatives have taken early steps to standardised methods and rating systems for assessing IVIS distractibility within the Australian context. In 2017, the Department of Transport (Road Safety, Victoria), with funding from the Victorian Government, conducted a study to identify methods for assessing driver distraction and to develop options for a distraction rating system (Imberger et al., 2020). The project generated two primary outcomes: the identification and recommendation of robust methods for assessing driver distraction, and the development of a framework for a distraction rating system.

In 2019, the Australian Automobile Association (AAA), together with the Department of Transport (Road Safety, Victoria), supported by funding from the Fédération Internationale de l'Automobile (FIA, France), collaborated with Professor Michael Regan and Dr Prasannah Prabhakaran from the University of New South Wales Sydney (UNSW) Research Centre for Integrated Transport Innovation to initiate a proof-of-concept study. The aims were to (1) replicate, extend, and refine methodologies developed by the University of Utah for assessing the distractibility of in-vehicle information systems (IVISs) and (2) evaluate the applicability of these methods in the Australian context to establish a foundation

for a broader program to improve human-machine interface (HMI) design and develop a distraction rating framework. The project was ultimately unable to proceed because a suitable controlled access test track (within available budget and compliant with COVID-19 restrictions) could not be identified. While the 2019 study was being scoped, VicRoads also funded a related project in 2020 with Red Scientific Inc. (Dr Joel Cooper's research consultancy) to further develop and streamline the HMI distraction assessment protocol established by Professor David Strayer and Doctor Cooper. This complementary work revisited the power and sensitivity of the Utah team's measures and explored alternative methods to evaluate the distraction potential of vehicle HMI interactions, identifying more efficient approaches suitable for large-scale assessment programs such as ANCAP.

In 2021, with a new funding source and removal of COVID-19 restrictions, the AAA awarded the feasibility stage of the HMI Distractibility Project to ABMARC who worked in collaboration with the University of Newcastle. The feasibility phase was used to refine and test proposed protocols, project timeframes and budget for the main research project. Following the final reporting of the feasibility stage in February 2022, the AAA awarded the contract for the full HMI Distractibility Project to the University of Newcastle in collaboration with ABMARC and Red Scientific. ABMARC is an independent Australian engineering consulting company with significant experience in automotive research including but not limited to ADAS and fuel efficiency testing. The HMI Distractibility Project began in February 2023 and ABMARC was responsible for engineering and technical components of on-track testing, specifically the installation and monitoring of vehicle telemetry hardware that provide data on driver inputs during their on-track test component. ABMARC is based at the Australian Automotive Research Centre (AARC; see facility [website](#)) automotive proving ground in Wensleydale, Victoria, which served as the test location for the pilot and main phase of the study.

The aim of the project was to establish a protocol that validly assesses the driver-distractibility of in-vehicle infotainment systems used in vehicles in the Australian market. To be effective, the method must distinguish between systems that are relatively "safe" to interact with and those that impose excessive visual demands or cognitive workload, thereby increasing distraction risk. The current study draws on the foundational research of the University of Utah team (Strayer, 2017) and the applied work conducted by Professor Mike Regan and Dr Prasannah Prabhakaran from UNSW. These earlier efforts provided the theoretical and methodological basis for assessing the distraction potential of vehicle HMI interactions in the present study. The authors would like to acknowledge and thank Red

Scientific and Professor David Strayer for their consultancy support and expert advice, which informed the development of the test protocol, methodology, and analysis plan.

The study consisted of a three phase, tiered evaluation framework:

- **Phase 1 – On-Track Drive:** IVIS distractibility was evaluated on-track with multiple participants tested within each vehicle, under conditions designed to mirror real-world driving. This is equivalent to Strayer et al's (2017) on-road assessment, with the addition of an eye-tracking measure. High-fidelity behavioural data from the Detection Response Task (DRT) and eye-tracking glasses were captured alongside vehicle telemetry data. The on-track drive provided the highest fidelity data and the closest to real-world driving as possible, while the DRT represents the industry-standard measure of cognitive workload in real-world tasks. This standardised task (ISO 17488:2016) measures residual cognitive capacity, or those resources not currently utilised by other tasks such as driving. Under states of low cognitive workload, residual capacity is plentiful, and DRT responses are fast and accurate. During periods of high cognitive workload, residual capacity is diminished, and DRT responses become slower, with some trials missed altogether.
- **Phase 2 – Off-Track Stationary test:** Our aim for Phase 2 was to explore if computer-based tasks, completed in stationary cars, could provide valid approximations of IVIS distractibility, as an alternative to expensive on-track testing. IVIS distractibility was assessed in stationary vehicles, in off-track conditions. Participants completed two, separate stationary tasks; a simulated driving task (computer-based Tracking Task and in-built visual DRT) used to measure cognitive workload and visual attention, and a visual occlusion test (VOT) that measured the visual demand of IVIS tasks. Stationary testing occurred in a car park at the AARC, without the need for a safety driver and with short-term access to test vehicles (approximately half the time of on-track testing).
- **Phase 3 – Modelling of IVIS distractibility:** The goal of this phase was to leverage statistical models to uncover the underlying relationships in the data collected during Phases 1 and 2—data that required substantial cost and technical expertise to obtain. Our aim was to identify more accessible, user-friendly measures of IVIS distraction by testing how well these models could predict new or unseen cases. This phase represented a novel approach, and if validated, it would enable IVIS distractibility to be assessed easily ranging from system designers with direct access to IVIS interfaces to consumers without access to a vehicle.

Through this phased approach, the project balanced scientific validity with long-term feasibility, to establish a sustainable methodology for ongoing evaluation of IVIS distractibility

in vehicles available in Australia. Importantly, the tiered structure of the assessment protocol provides options for several stakeholder groups; those with technical expertise and capability to conduct on-track testing, those with sufficient technical expertise to run stationary testing with computer-based tasks all the way to consumers who need a simple, cost-effective assessment method that allows them to make an informed decision when considering vehicles with IVISs.

6 Project Governance

7 Governance Structure and Research Team

Figure 1 shows the key members of the project management and research team. The project design and development teams worked on the test protocol and analysis plan up until the completion of the data collection phase in September 2023. ABMARC and the University of Newcastle data collection teams completed the data collection protocol from June through until September 2023. The University of Newcastle steering committee consisted of the University of Newcastle's project design and development team which met fortnightly. The University of Newcastle project management team was responsible for providing monthly updates to the AAA Project Management Team.

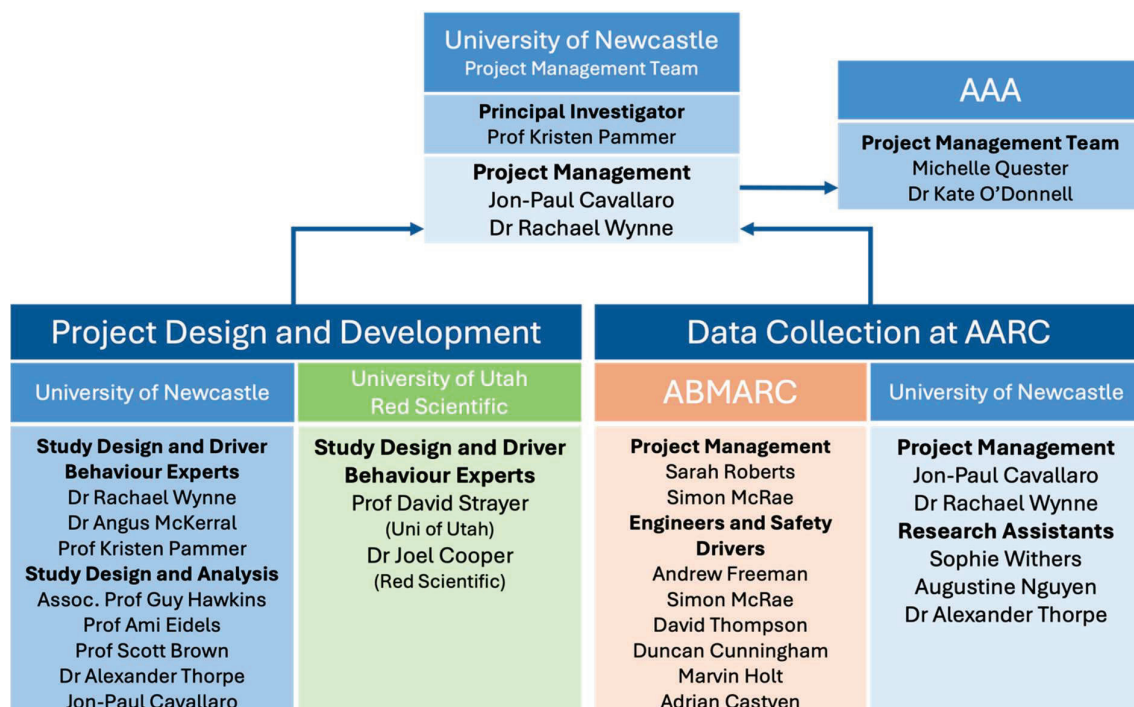


Figure 1. The AAA-HMI Distractibility Project governance structure.

8 Project Methodology

The following section outlines methodological details common to both on-track drive (Phase 1) and off-track stationary (Phase 2) tests. Methodological and procedural aspects unique to each phase are described in their corresponding section.

8.1 General assessment protocol and data collection

Data collection occurred at the AARC over a 12-day period in the Australian winter of 2023. The assessment protocol consisted of three test sessions (morning, midday and afternoon) per day. The on-track drive (Phase 1) took one session to complete, and the off-track stationary tests (Phase 2) were completed in one session (See Figure 2 for an outline of the two-day test protocol completed by each participant). Our data collection target was to test 20 participants and 13 vehicles; therefore, on average we had each group of participants complete the test protocol in three vehicles over two consecutive days.

We scheduled two-and-a-half hours per test session. First, participants completed a 45-minute familiarisation task to learn how to interact with an IVIS. This familiarisation stage occurred in a car park at the AARC while the vehicle was stationary. The remaining hour and 45 minutes were dedicated to the test component (track drive during Phase 1 or VOT and tracking task during Phase 2).

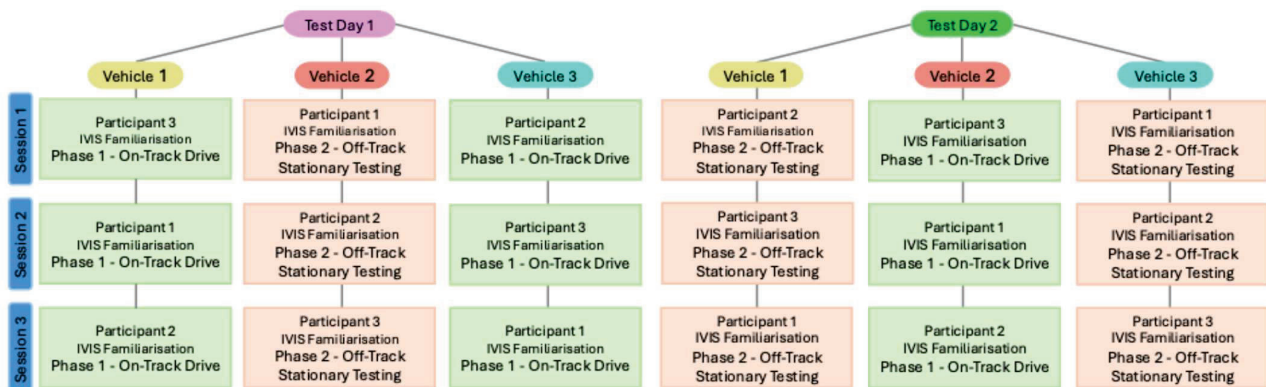


Figure 2. The test protocol completed by each participant at the AARC. All participants completed an on-track drive (Phase 1) and stationary testing (Phase 2) in three vehicles (i.e., assessing three IVISs) across a two-day period. The first test day involved two on-track drive sessions and one stationary test session. The second test day involved one on-track drive and two stationary test sessions.

8.2 Participants

Twenty (15 male) participants between the ages of 26-50 years ($M = 32$ years) completed the assessment protocol at the AARC. Participants reported an average of 5 hours of driving per week. Participants were recruited through the AAA's social media channels. They were screened for any potential issues that may compromise testing including but not limited to motion sickness and driving anxiety assessments. We recruited participants with a valid full Australian driver's license who drove a car more than once a week. Participants with corrected vision were only permitted to participate if they were able to wear contact lenses, as glasses could not be worn during on-track testing in Phase 1 where participants were required to wear eye-tracking glasses. Participants completed demographic questionnaires prior to attending the track site. The AARC is based in a remote location and the project demanded significant time commitment from participants, therefore, participants were reimbursed \$60 per hour. Informed consent was obtained from each participant before they began testing on their first day at the AARC. We acknowledge and thank all participants for their time, interest and involvement in the project. This study was approved by the University of Newcastle Human Research Ethics Committee (protocol H-2022-0014).

8.3 In-Vehicle Infotainment Systems and Vehicles

The UoN research team and ABMARC considered several options to procure vehicles including leasing vehicles from manufacturers, purchasing vehicles outright, and hiring vehicles from car hire companies. We chose to hire all vehicles from Melbourne Airport car hire companies as they offered a wide range of vehicles at reasonable cost.

The research team and engineers at ABMARC created a "wishlist" of vehicles to be tested. The vehicles were chosen based on several criteria: their market share in Australia, representative sample from a broad range of car categories (e.g., electric vehicles, European cars, small/medium/large SUV, 4WD etc.) and their availability at time of hire. With a 12-day test schedule at the AARC, we initially planned to test 16 vehicles; however, we tested 13 IVIS and vehicles, three fewer than planned because a schedule conflict during the first week of testing led to reduced track time. The 13 IVIS and vehicles are shown in Table 1. One vehicle, the Toyota Hilux, was provided by ABMARC, our research partner at the test facility. The Toyota Hilux was the best-selling vehicle in Australia by volume in 2023 and was included in the test fleet as a representative example of a widely used Australian vehicle equipped with its manufacturer's native Link infotainment system. We also tested the two most common smartphone systems; Google's Android Auto® and Apple's Apple CarPlay®. Both smartphone systems were assessed in the same vehicle; a 2022 Toyota Corolla Hatch, to ensure

fair comparison. The remaining nine native systems and vehicles were chosen to represent the car categories listed above but were ultimately chosen based on their availability at the time of hire. One vehicle, the MG XS, was assessed during the first week of testing, but its data were removed from the analysis because its IVIS had limited functionality (i.e., fewer than 4 IVIS functions). Therefore, a total of 12 IVIS and vehicles were included in the analysis.

We used Samsung Galaxy A54 smartphones to assess all calling, messaging and media functions in the native IVISs. All phones contained an identical contact list that consisted of randomly generated contact names (e.g., Tim Newton). Each contact was assigned to Sim cards in a separate set of phones held off-site. The media playlist contained 15 songs that were selected from Rolling Stone's 500 Greatest Songs of All Time. We connected the smartphones to each IVIS via Bluetooth rather than through a USB cable because connecting via USB forced IVISs to 'Android Auto' mode. We used the same model Samsung Galaxy A54 to assess Android Auto and an Apple iPhone X to assess AppleCarplay® platforms. We chose to connect these two smartphones via USB cables to ensure connection stability and eliminate the risk of system 'dropouts' over the course of testing.

Although specific systems were tested during the study, the purpose was not to provide a "product review" or evaluate their commercial performance. Instead, the focus was on examining how the inherent characteristics of each system contributed to varying levels of distraction. This approach allowed for a comparative understanding of the mechanisms behind distraction, rather than an assessment of the systems themselves.

Infotainment System	Vehicle Make/Model	Vehicle Category
Android Auto®	Toyota Corolla 2021	N/A
AppleCarPlay ®	Toyota Corolla 2021	N/A
Audi Multi-media Interface	Audi Q7 2020	Large SUV/European
BMW iDrive 7.0	BMW 330i 2023	European
Cupra MIB3	Cupra Ateca 2022	Medium SUV
SYNCR4A	Ford Ranger 2023	4WD
Mazda Connect	Mazda CX-9 2021	Japanese/SUV
MG iSMART*	MG ZS 2022*	Small SUV
NissanConnect	Nissan X-Trail 2023	Medium SUV
Android Automotive OS	Polestar 2 2022	Electric
Tesla Touchscreen	Tesla Model Y 2022	Electric
Toyota Link	Toyota Corolla 2021	Small car
Toyota Link	Toyota Hilux 2019	4WD

Table 11. The fleet of 13 test vehicles. We selected systems from a broad range of car categories; however, all cars, except the Toyota Hilux, were selected based on their availability at the time of hire. * The MG ZS 2022 was excluded from the analysis because it had fewer than 4 IVIS task functions.

8.4 IVIS functions - *Task Types*

We chose seven IVIS task types to assess across all systems; however, some systems had limited functionality (see Table 2 below for available task types in each IVIS or system), so we decided that a system must have four or more IVIS task types to be included in our analysis. Our IVIS task types were chosen to cover the same IVIS task types used by Strayer et al. (2017) i.e., audio entertainment, calling and dialling, text messaging and navigation. However, we decided to differentiate between types of audio entertainment such as playing a song from a playlist (i.e., “Media” IVIS task), and selecting a radio station (i.e., “Radio”). We also decided to differentiate between types of phone calls and text messages. We separated phone calls into two IVIS tasks: (1) “Receive Call” and (2) “Initiate Call”. Similarly, text messaging was separated into “Receive Text” and “Initiate Text”. We made these distinctions because the time and operations required to complete those tasks differ (e.g., a driver simply presses a button or touches the IVIS screen to answer a call, whereas a driver must locate a contact with a voice command or a visual search of the contact list on the IVIS before initiating the call) seven IVIS task types were:

1. ***Initiate Call***: Participants were instructed to call a contact on the car’s phone connected to the IVIS, or to dial their own phone number. The contact list consisted of 15 contacts. Initiate call trials were deemed complete once the system had connected to the correct contact.
2. ***Initiate Text***: Participants were asked to send a text message to one of the car phone’s contacts. For example, participants were asked to “*Text Alex and ask if he could buy you a coffee*”. Initiate Text trials were considered complete once the participant had successfully sent the text message and had said “Done”.
3. ***Media***: Participants were asked to play a song from a pre-loaded playlist on the car’s phone. The playlist consisted of 15 songs (See Appendix 2 - Full set of IVIS trials for full song playlist)
4. ***Navigation***: Participants were asked to navigate to a location of interest within a 40-kilometre radius of the AARC (E.g., Please navigate to Bells Beach, Victoria). Navigation trials were deemed complete once the system had successfully found the route to the location of interest and the participant had said “Done”.

5. **Radio:** Participants were asked to tune the radio to an AM or FM radio station in the local area. Radio trials were deemed complete once the system was tuned to the correct frequency and the participant had said “Done”.
6. **Receive Call:** Participants were required to answer a phone call from the researcher. Receive call trials were deemed complete once the call was connected and the participant had said “Hello”.
7. **Receive Text:** Participants interacted with the IVIS and had it read out a text message received from the researcher. Trials were deemed complete once the system had completed reading the text message and the participant had said “Done”.

Vehicle/System	Initiate Call (IntC)	Initiate Text (IntT)	Media (Medi)	Navigation (Nvgt)	Radio (Radi)	Receive Call (RcvC)	Receive Text (RcvT)
Android Auto	✓	✓	✓	✓	—	✓	✓
Apple CarPlay	✓	✓	✓	✓	—	✓	✓
Audi Q7	✓	✓	✓	✓	✓	✓	✓
BMW 330i	✓	✓	✓	✓	✓	✓	✓
Cupra Ateca	✓	✓	✓	✓	✓	✓	✓
Ford Ranger	✓	—	✓	✓	✓	✓	✓
Mazda CX-9	✓	—	✓	✓	✓	✓	✓
MG ZS*	✓	—	—	—	✓	✓	—
Nissan X-Trail	✓	—	✓	—	✓	✓	✓
Polestar 2	✓	✓	✓	✓	—	✓	✓
Tesla Model Y	✓	✓	✓	✓	✓	✓	✓
Toyota Corolla	✓	—	✓	—	✓	✓	✓
Toyota Hilux	✓	—	✓	✓	✓	✓	✓

Table 22. All IVIS task types available in all IVIS/Vehicles. We chose seven IVIS tasks to assess across all systems; however, some systems had limited capability. Systems with fewer than four task types (e.g., MG ZS) were excluded from the analysis.

We generated a run-sheet of IVIS tasks for each participant by randomly selecting 10 trials from a bank of 15–20 trials created for each task type (see Appendix 2 - Full set of IVIS trials for all task types). Each participant completed a unique run-sheet specifying the task-type trials for each vehicle or system. The task types included on a participant’s run-sheet for

a given IVIS were determined by the functionality available in that system; for example, the Navigation block was omitted if the IVIS did not support navigation. Within each IVIS, the order of the task-type blocks was randomised, resulting in participants completing the available tasks in different sequences (see Appendix 1 - Example “Run Sheet” for an illustration).

8.5 Modes of IVIS interaction

Three modes of interaction were used during testing:

1. the IVIS and centre stack (touch screen and physical buttons/dials),
2. voice commands, typically activated via a microphone button on the steering wheel, and
3. steering-wheel buttons (e.g., toggles for skipping forward or backward between songs).

Participants were encouraged to use whichever interaction mode they would normally use when driving, both during training and familiarisation. However, their choices were constrained by each IVIS’s functionality. Some interaction modes were only available for specific IVIS tasks (e.g., media selection could be completed using steering-wheel or IVIS-mounted controls, but not by voice), and some systems allowed multiple configurations of interaction modes within a single task (e.g., initiating a task via touch and completing it with a voice command). This variation created an impractically large number of potential interaction configurations. Exhaustively testing them was not feasible, and doing so would provide limited benefit, as any findings would not generalise to novel or emerging interaction modes (e.g., gesture control). Consequently, we defined the “mode of interaction” in a way that allowed us to quantify the number of IVIS interactions or “steps” required for each task in subsequent analyses (see Phase 3 testing for details).

8.6 Study Design

We used a repeated measures design; all participants completed an on-track drive (Phase 1) and stationary tests (Phase 2) in three vehicles. We made every effort to use counterbalance measures in the design of our test protocol; however, some design choices were governed by practicality and limits imposed by our test schedule. For example, to counterbalance the order of test phases each participant completed within a vehicle (e.g., in vehicle 1, participant 1 (P1) completes Phase 1 *then* Phase 2, P2 completes Phase 2 *then* Phase 1 and P3 completes P1 *then* P2) we would need to de-instrument and re-instrument a vehicle

(i.e., remove the telemetry system). This process took approximately 4 hours for one of the systems. Therefore, to minimise disruption to testing we kept a vehicle instrumented for a day of on-track testing, and all three participants completed their on-track drive (Phase 1) in “Vehicle 1” in one day and stationary testing (Phase 2) was completed in Vehicle 1 on the previous or following day, depending on when the vehicle was scheduled to complete on-track testing. In effect, all participants completed two on-track drives (Phase 1) and one stationary test session (Phase 2) on their first day of testing and one on-track drive (Phase 1) and two stationary test sessions (Phase 2) on their second day of testing in each of the three vehicles they tested in (see Figure 1. for a schematic outlining the two-day test protocol). We tested whether the sequence in which participants completed the on-track drives influenced DRT performance. Drive order had no meaningful effect on any model outcomes, confirming that IVIS differences were not confounded by practice or fatigue effects (see Appendix 3 – Analysis of “Task Order”). The schedule allowed us to counterbalance the order of stationary tasks within Phase 2 testing, where two participants completed the VOT *then* the tracking tasks and one participant completed the tracking task *then* the VOT in a vehicle.

Prior to testing, participants were familiarised with the IVIS. First, participants were encouraged to explore an IVIS’s functionality for the IVIS tasks outlined in Table 2, and to determine how to interact with the system in a way they felt most comfortable (e.g., via the centre console display, the audio assistant, or a combination of each). Once participants understood an IVIS’s functionality, they completed multiple practice trials of each task type until they were comfortable interacting with the system. Participants completed the familiarisation task at the beginning of their first test session within each vehicle, and they were allowed to re-familiarise with the system if they needed to at the beginning of their second test session within that vehicle.

8.6.1 General IVIS Assessment Procedure

For all testing in Phase 1 and 2, we used the following general test procedure. Researchers sat in the back seat of the vehicle to monitor participant performance, with clear vision of the participant and the IVIS. For each trial, the researcher asked participants to complete an IVIS task (E.g., For an ‘Initiate Call’ task, the researcher said, “Please call Alex.”) and then prompted the participant to begin the trial by saying “Go”. Participants were required to complete all trials as quickly and accurately as possible and to notify the researcher when they had completed a trial by saying “Done”, at which point the task, a single ‘experimental trial’, was complete. The researcher recorded notes for each trial (e.g., the trial was successful, participant errors or system failures) and allowed for 10 seconds after each trial so the

participant would be ready for the next trial. Overall, participants completed 10 trials per IVIS task, referred to as a “block” of trials, totalling up to 70 interactions for a test (i.e., on-track drive, Tracking Task, or VOT). Participants were told to take a break in-between each block. The researcher recorded the time each block began and ended. Each participant completed their own ‘run sheet’, which was a pred-determined order of IVIS tasks, across all vehicles and phases.

8.7 Data Analysis and Modelling

All outcome variables (i.e., our distractibility measures) were screened for outliers, skewness, and heteroscedasticity. Outcome variables are defined in the corresponding phase. Prior to modelling, data from the VOT were square root transformed to improve normality and stabilise variance.

It wasn’t feasible to test every participant in every car, which meant that only a small number of people were tested in each vehicle. A limitation of this approach is that individual differences can strongly influence the results. People naturally vary in how they perform tasks—for example, some may be more familiar with driving technology, more attentive, or more comfortable in test situations than others. With only a few participants per car, these differences can shape the outcomes in ways that may not reflect the distraction potential of an IVIS. In other words, part of what we observe may be due to *who* was interacting with the IVIS, rather than the IVIS’s functions. To minimise this issue, we used statistical models that account for both the overall trend across participants and the unique patterns for individuals. This helps us capture the big picture while still recognising that people differ in their performance. In this report, we make a clear distinction between pre-corrected data (i.e., the directly observed or empirical measurements) and corrected data, which represent model-derived estimates that account for systematic influences such as participant variability and measurement error.

We used linear mixed effects models for all model analyses. Linear mixed effects models are appropriate for repeated measures designs and are commonly used in IVIS distractibility studies to estimate cognitive workload (see Strayer et al., 2017; Strayer et al., 2019; McDonnell et al., 2021). All analyses were carried out in *R* (Version 4.3.2; R Core Team, 2023) using the lme4 package (Bates et al., 2015). We used the same model specification for all analyses in Phases 1 and 2; only the outcome variable differed between models. The full model outputs, including fixed- and random-effects structures, are provided in the corresponding appendices for each analysis. Each model included IVIS (i.e., the IVIS

or vehicle) and the IVIS $\times$ task-type interaction as fixed effects, with participants modelled as random intercepts to account for inter-individual variability. This specification allowed us to model both (1) differences in overall distractibility between IVISs and (2) how distractibility varies across task types within and between IVISs.

The fixed effect for task type was deliberately omitted. An overall “average” distractibility associated with a task type (e.g., the average distractibility of navigation interactions across all systems) is meaningless in this context, as it collapses over substantial design differences between IVISs. Our aim was to quantify how distractibility varies by task *within* each IVIS and to compare these patterns across systems. Excluding the task-type main effect ensures that the model captures IVIS-specific task profiles that are directly relevant to comparative assessment rather than estimating a generic effect that would obscure system-level differences.

The presence of IVIS $\times$ task-type interactions means that the IVIS fixed-effect coefficients correspond only to the predicted value for the reference task (Initiate Call) and therefore do not provide system-level summaries of distractibility. To derive interpretable IVIS-level estimates that appropriately average across task types while respecting the fitted interaction structure, we computed estimated marginal means using the *emmeans* package (Lenth, 2022). The estimated marginal means provide model-adjusted system-level predictions and form the basis for all IVIS comparisons and correlation analyses reported in the main text. The underlying fixed-effect coefficients are presented in Appendix 7 for completeness.

Although mixed-effects models allow for post-hoc contrasts between IVIS systems, these comparisons were intentionally not conducted. The purpose of the modelling was to demonstrate method sensitivity; that is, whether each distraction measure reliably detected increases in distraction across different IVISs and task types, rather than to statistically rank individual systems. Conducting pairwise comparisons across all systems and task types would require hundreds of contrasts per model, greatly inflating Type I and Type II error rates and producing results that are difficult to interpret and potentially misleading. Moreover, between-system differences are not stable across tasks, meaning that post-hoc comparisons could imply performance consistency that does not exist. Most importantly, the IVIS star-rating framework does not rely on inferential system-to-system differences, but instead on a separate step-based metric that reflects interaction demand. For these reasons,

post-hoc IVIS contrasts were considered both unnecessary and conceptually inappropriate for the aims of this project.

8.8 Data cleaning

We carried out thorough checks of data sets from Phase 1 and Phase 2. We found one participant had failed to follow the VOT procedure in Phase 2. Video recordings and data from the VOT showed the participant had repeatedly opened the VOGs during, before and after trials in two of three vehicles. We decided to remove all of their data (i.e., data from Phase 1 and Phase 2) from the analysis.

9 Phase 1: On-Track Drive - Assessment of cognitive and visual demand of IVISs in ‘real-world’ conditions

9.1 Procedure

The original iteration of this project was unable to proceed because lack of a suitable track within available budget and the constraints of COVID-19. Previous studies (Strayer et al., 2017) tested on public roads; however, the AAA would only support a closed track test for its safety benefits. Phase 1 testing was complete on the AARC’s highway



circuit. The highway circuit was chosen because of its likeness to a real highway; it was built to comply with Australia's standards for highways at over 80km/h. It is a 4.2 kilometre long, two-lane, 7.6 metre wide, closed oval circuit that is fully fenced, which prevents wildlife from entering circuit, with significant run-off areas around the track for added safety.

On-track testing offers data of the highest fidelity and replicates real-world driving as closely as possible. Participants drove the car (primary task) and interacted with the IVIS (secondary task), while they were required to engage in a tertiary task, the detection response task (DRT). Participants were asked to prioritise safe driving, interact with the IVIS and respond to the DRT (in that order) while maintaining a driving speed of 80 km/h. We trialled

two driving speeds during the pilot phase: 80 km/h and 100km/h. Higher speeds typically induce greater workload and come with an additional cost to safety in the event of an on-track incident. Our pilot study showed no apparent difference in the response data collected from participants performing the procedure at either speed, therefore we chose to run vehicles at the lower speed of 80km/h to maximise safety.

Before testing, participants completed a sighting lap of the track to familiarise themselves with the road conditions. Participants wore eye-tracking glasses from the sighting lap to allow them as much time as possible to become accustomed to them; however, the glasses only recorded data during the test phase. The on-track drive involved four driving stages, three of which were familiarisation or training stages with the aim to prepare participants for the on-track drive test phase. We followed the training procedure outlined in the DRT ISO guidelines (ISO 17488:2016). All vehicles were fitted with two video cameras; a forward-facing camera placed behind the participant to record their interactions with the IVIS, and a rear-facing camera placed on the windscreen that recorded the driver's face to capture their head and eye movements. A research assistant sat in the rear of the car to monitor the participant's interaction with the IVIS and to monitor a laptop controlling the DRT and eye-tracking glasses. An ABMARC safety driver sat in the passenger seat to monitor the participant's driving performance and take over in the event of a driving incident (e.g., car leaving the roadway).

9.2 The Detection Response Task

The DRT is an industry standard driver distraction assessment method used to measure the cognitive load and attentional demands of secondary tasks. Strayer et al (2017) used the DRT in their study to assess cognitive load and we followed their lead by using the tactile DRT. We procured DRT microcontroller-based systems from Red Scientific.

In our Phase 1 on-track test, the DRT allowed us to measure the cognitive demands of IVIS interactions while participants were driving. The DRT is a simple task in which participants make a sequence of fast responses to a simple stimulus; in our case, a vibrating tactor which was attached to the participant's right collarbone. A small response microswitch was placed on the participant's right index finger. Both the stimulus and the response button are placed on the same side of the body so that reaction times reflect a single, consistent neural pathway. This improves measurement reliability and comparability between participants or conditions and leaves the left hand free to interact with the IVIS and other controls. The stimulus (i.e., vibration) occurs randomly every 3-5 seconds while participant goes about their primary task of driving the car.

We obtained two outcome variables from the DRT:

1. **DRT Miss Rate** – a DRT miss is a **failure to respond** to a DRT stimulus. Expressed as a proportion, DRT misses are defined as the sum of misses during a block of IVIS trials divided by the number of DRT stimuli presented during a block of IVIS trials. Participants miss a stimulus when their cognitive resources are taxed because they are interacting with an IVIS and their attentional process fail (ISO 17488:2016).
2. **Response time (RT)** – time (in milliseconds) taken to respond to a DRT stimulus. Time is recorded from stimulus onset until the participant responds to the stimulus. RTs were recorded for each trial within a block of IVIS task trials.

Both DRT outcome variables provide a measure of cognitive workload. For example, when cognitive capacity is taxed (e.g., when driving *and* interacting with the IVIS), participants are more likely to miss the DRT stimulus and if participants do respond to the stimulus, their response time is reliably slowed. Therefore, even short tasks such as receiving a call may be highly distracting if they require a substantial reallocation of attention. Conversely, relatively low-load tasks such as initiating a call via a voice command may still pose a significant distraction from systems that require substantial time to complete the task.

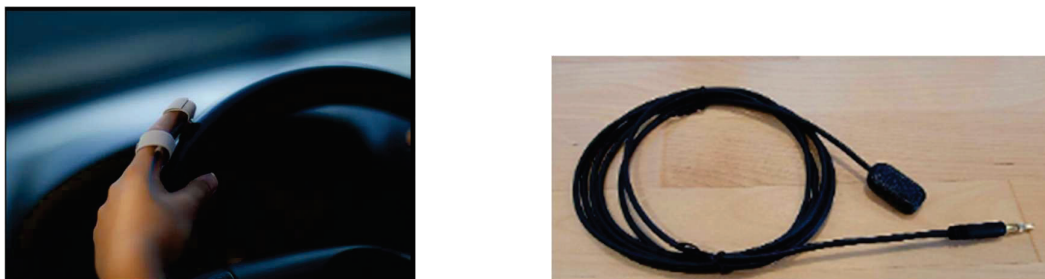


Figure 3. Equipment required to run the vibrotactile version of the Detection Response Task (DRT). The driver responds to a vibration stimulus with a simple microswitch (pictured, left) attached with a hook and loop strap on their left index finger. A vibrotactile motor (pictured, right) that delivers a vibration stimulus is taped to the driver's right collarbone.

9.3 Eye tracking

Eye tracking is a supplementary data collection methodology that helps to confirm how IVIS tasks impact eye glance behaviour. Our eye-tracking equipment, Tobii Pro Glasses 3, was used to accurately log both *where* drivers were looking and *what* they are looking at while they were driving and interacting with the IVIS. The eye tracking glasses 'track' the participant's pupils in each eye independently, and a small camera on the nose bridge captures video of the participants forward view. The eye-tracking data provide measures of time when a driver's eyes are 'on-road' or 'off-road'. Gaze-based driver monitoring systems are assessed and encouraged through ANCAP's safety rating criteria, but they are not required to achieve a 5-

star rating. Demonstrating the utility of eye-tracking to assess IVIS distractibility provides support for the expanded use of gaze-based DMS to capture system distractibility. Eye-tracking outcomes were collected to be compared with measures of visual attention captured during Phase 2.



Figure 4. Real-time eye gaze data was captured with sophisticated wearable eye tracker glasses. The Tobii Pro eye tracking glasses are lightweight, unobtrusive and do not affect driving performance.

9.4 Vehicle Telemetry

We collected vehicle telemetry data during on-track testing to provide additional behavioural context; however, the aim was not to model driving-performance risk. Instead, telemetry was used to examine whether workload-related changes in steering or speed control occurred alongside the cognitive and visual measures. Indirect performance measures such as telemetry are typically insufficient for accurate state detection on their own and instead serve as complementary indicators to direct measurement approaches such as eye-tracking, physiological monitoring, and performance probes. Nevertheless, given the industry emphasis on maximising value from existing in-vehicle signals, it was necessary to collect telemetry data to explore the potential for telemetry to complement direct measures of distraction.

We used two data acquisition units; a Racelogic VBOX and an Oxford OXTS unit because two test vehicles were driven simultaneously each day. Both systems were installed by ABMARC engineers and recorded the common signals including tri-axial acceleration, roll, pitch, yaw, and GPS position. Although the systems differed in the full range of signals available, only *mean lateral acceleration* (lane-keeping stability) and *mean speed delta* (longitudinal speed variability) were analysed, as these were the only telemetry variables consistently captured across both systems and are two of the most common vehicle control measures found in the distracted-driving literature related to workload degradation. These telemetry data were used primarily to provide behavioural context to cognitive and visual-attention measures.

10 Results

10.1 DRT Data preparation

We followed DRT data processing procedure as per Strayer et al., (2017) and Thorpe et al., (2020). We removed reaction times faster than 100ms and slower than 2500ms from reaction time calculations. Trials where participants failed to respond or responded after 2500ms were deemed misses. We removed all incomplete and interrupted trials from the data set prior to analysis.

10.2 On-Track Drive Pre-Corrected Data - DRT

Figures 5 and 6 show the mean pre-corrected DRT proportion of misses and DRT RTs from the track drive, respectively. Each panel within each figure shows the mean of those outcome variables (i.e., DRT proportion of misses or DRT RT) for each IVIS task type within each vehicle or IVIS. Error bars represent the standard error of the mean.

Figure 5 illustrates clear differences in the proportion of DRT misses across IVISs and task types. Systems such as the Mazda CX-9, BMW 330i, Tesla Model Y, Nissan X-Trail, Cupra Ateca and Toyota Hilux consistently show the highest miss rates, reflecting elevated cognitive demand and a reduced capacity to detect and respond to external stimuli during IVIS interaction. In contrast, the Audi Q7, Toyota Corolla and Polestar 2 display low and stable miss rates across their available task types, indicating comparatively lower workload. Android Auto also shows relatively low miss rates for most task types, except for more demanding text-entry interactions. Miss rates also vary systematically across task types. Simpler tasks (e.g., receiving calls or texts and initiating calls) tend to yield very low miss proportions, whereas more demanding tasks such as navigation, media browsing and initiating text messages produce substantially higher miss rates across most systems. This pattern aligns with theoretical expectations: as interaction complexity increases, DRT detection failures rise sharply. Collectively, these data demonstrate that DRT miss rate is a robust indicator of both IVIS-level design differences and the cognitive demands imposed by specific task types.

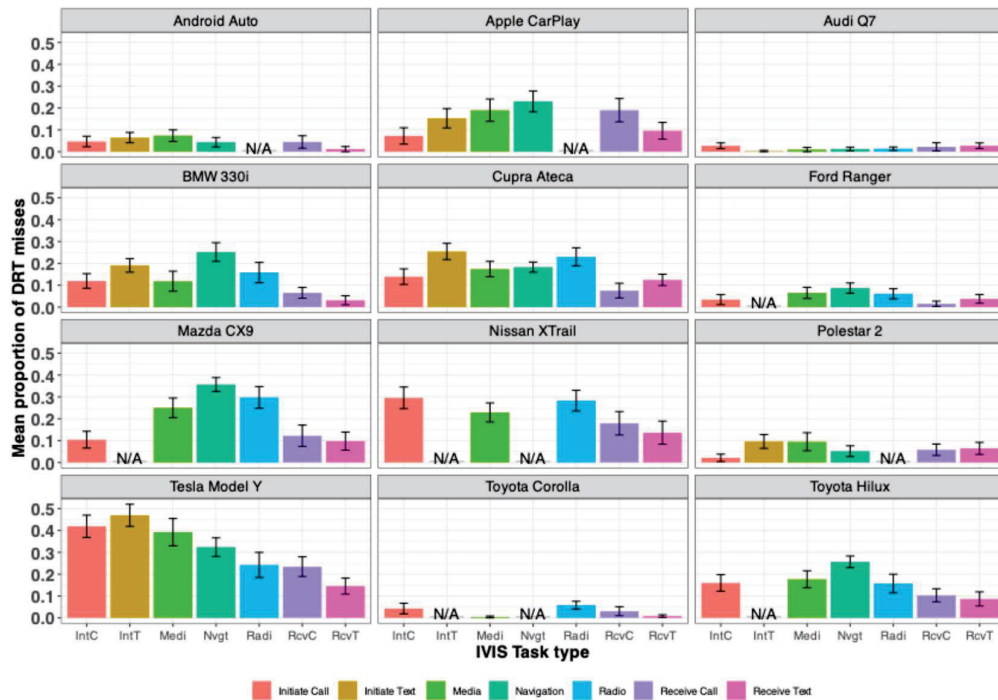


Figure 5. Measure of cognitive workload, expressed as the mean proportion of DRT misses associated with IVIS interactions during on-track testing. Each panel presents the pre-corrected data for a given IVIS or vehicle, with individual task types shown where applicable. Standard-error error bars indicate the variability of miss rates across participants. Task types labelled N/A represent functions that were not available within that IVIS.

Figure 6 illustrates that mean DRT reaction times vary across IVISs and task types, with some systems producing consistently slower responses than others. Systems such as the Mazda CX-9, BMW 330i, Tesla Model Y, Nissan X-Trail, Cupra Ateca and Toyota Hilux tend to show the longest RTs, reflecting generally higher levels of cognitive workload. In contrast, the Audi Q7, Toyota Corolla, Android Auto and Polestar 2 consistently produce relatively fast RTs, suggesting lower visual-cognitive demand. Reaction times also differ across task types in theoretically meaningful ways. Simpler interactions, such as receiving calls or texts and initiating calls, are associated with shorter RTs, whereas more complex tasks, including navigation, media interactions and initiating texts, produce longer RTs across most systems. These patterns indicate that RTs can capture broad differences in both IVIS design and task-type demands, although the magnitude of these effects is modest compared with other distraction metrics

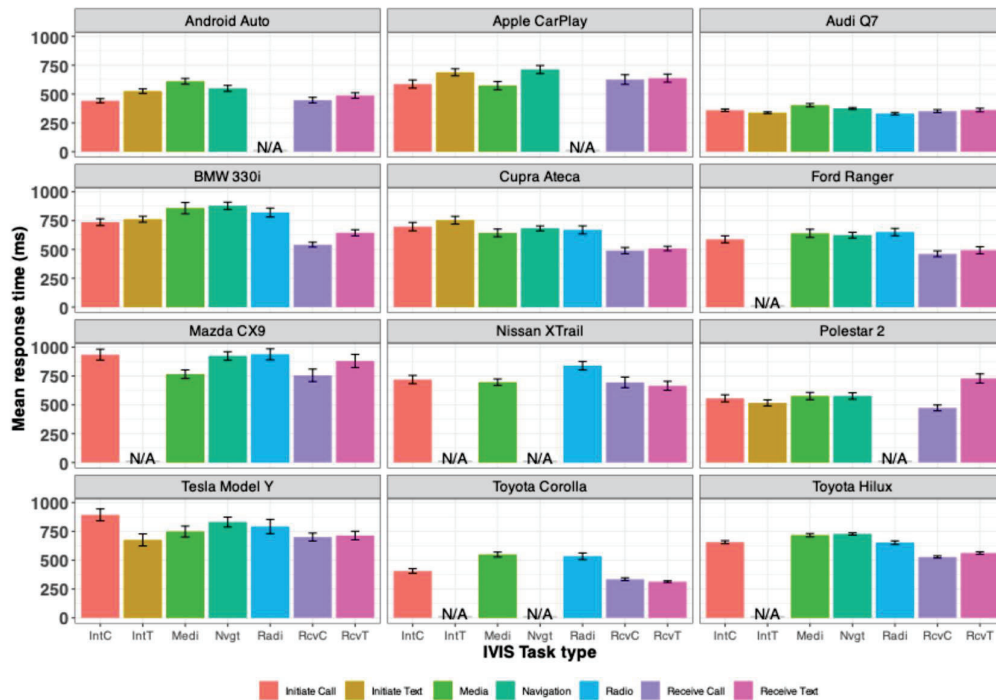


Figure 6. Pre-corrected vibrotactile DRT response times during on-track IVIS interactions. Each panel displays the mean reaction time for a given IVIS and its available task types, with standard-error error bars reflecting between-participant variability. The figure shows that reaction times vary modestly across IVISs and task types, indicating limited sensitivity of mean RT to differences in cognitive workload associated with IVIS interaction. Task types labelled N/A denote functions not available within that IVIS.

It is important to note that the data shown in Figures 5 and 6 are pre-corrected data before modelling (i.e., observed data) and these trends should not be over-interpreted prior to modelling. This is because our modelling provides a structured way to test whether these patterns are consistent, reliable, and generalisable beyond the specific people who completed the study in each vehicle. This is particularly important when we look at a broader level i.e., the ‘average’ distractibility of each IVIS system, as in subsequent modelling analyses.

10.3 Modelling On-Track DRT data

We applied linear mixed-effects models to our on-track outcome variables to statistically partition variance attributable to participants from variance due to IVISs. This approach estimates the average (fixed) effects of each IVIS while simultaneously accounting for individual (random) differences in performance. As a result, the model provides corrected, model-based estimates of distractibility that more closely reflect the IVIS’s underlying effect, independent of the specific drivers tested. This substantially increases confidence that the patterns we report represent true differences between IVISs rather than artefacts of who happened to be driving each vehicle.

Figure 7 shows the difference between pre-corrected DRT miss data (top panel) and corrected DRT miss model estimates (bottom panel). The mixed-effects model accounts for individual variability by estimating participant *and* IVIS-level effects. This produces model-based estimates that correct for differences in driver capability, yielding a more accurate approximation of each IVIS’s “true” distractibility. Consequently, while the same IVISs remain at the poorer end of the performance spectrum for DRT misses, others such as the Audi Q7, Polestar 2, and Toyota Corolla shift in their relative ranking *and* estimated level of distractibility. These shifts reflect the model’s ability to disentangle systematic participant effects from IVIS-level effects, enabling fairer comparisons between systems.

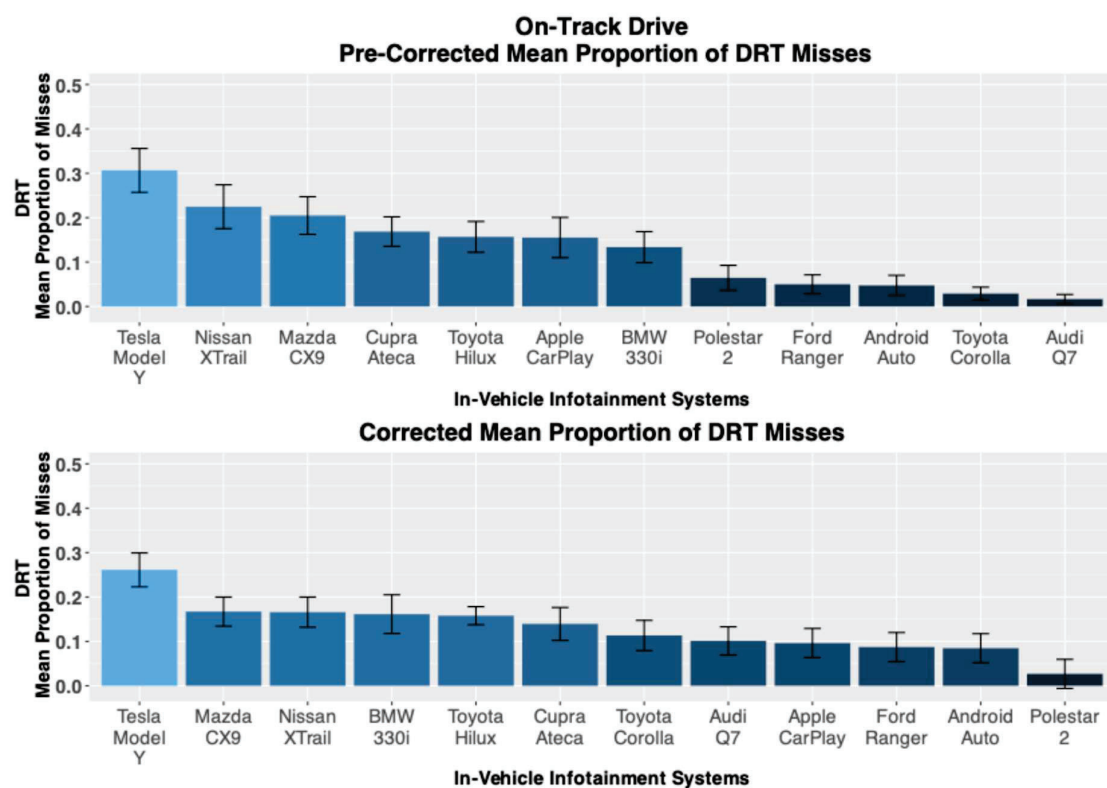


Figure 7: *On-track drive DRT misses. The top panel shows observed, pre-corrected data, and the bottom panel presents the corrected model estimates. Once individual variation is accounted for in the model, the ordering of IVISs changes, providing a more accurate or “true” ranking of IVIS distractibility.*

The DRT miss model estimates quantify the extent to which each IVIS reduces a driver’s available cognitive capacity while driving. As described earlier, the DRT measures a driver’s ability to detect a vibrotactile stimulus while simultaneously performing the primary driving task; thus, higher miss rates indicate greater cognitive load. The model outputs show clear differences between systems. For example, the Tesla Model Y produced an estimated miss probability of roughly 0.25, meaning that drivers interacting with this IVIS failed to

detect about one in four DRT stimuli. Although DRT misses do not translate directly to specific on-road crash scenarios, such elevated miss rates indicate a significant diversion of cognitive resources away from hazard detection. In practical terms, this reflects a reduced capacity to register and respond promptly to unexpected events and therefore signals a meaningful increase in safety risk for systems with high distractibility profiles.

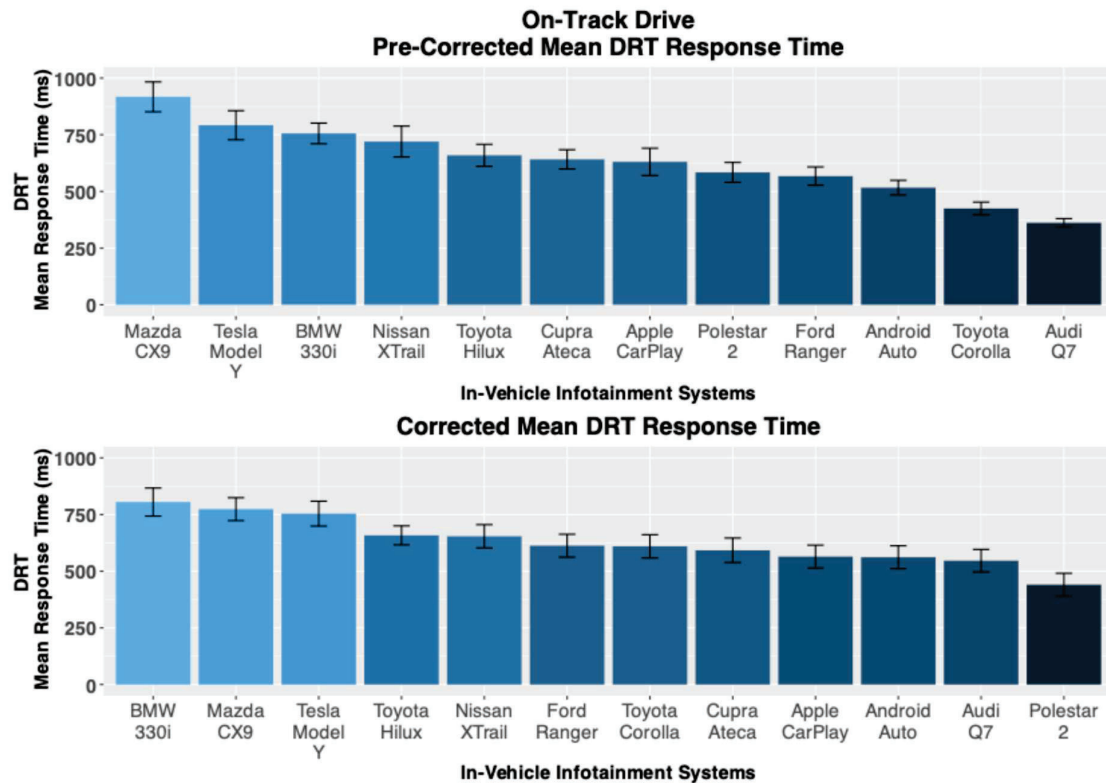


Figure 8. DRT Mean Response Time. The top panel shows the observed, pre-corrected mean response times, while the bottom panel presents the adjusted model estimates. Modelling the data accounts for individual differences in performance, which shifts the ordering of IVISs and provides a more accurate indication of their relative cognitive demand.

Figure 8 presents the pre-corrected DRT reaction time (RT) data (top panel) alongside the model-estimated marginal means (bottom panel). These corrected estimates adjust for participant-level variability and provide a more accurate representation of the underlying mean RT for each IVIS. The modelled values slightly modify the relative ordering of systems and reduce the overall spread of mean RTs, reflecting the statistical adjustment toward population-level estimates. Three systems, BMW 330i, Mazda CX-9 and Tesla Model Y, exhibit the slowest modelled mean RTs, typically in the range of 750–800 ms, indicating comparatively high cognitive demand during IVIS interaction. In contrast, systems associated with faster RTs, including Polestar 2, Android Auto and Audi Q7, produce mean RTs in the range of approximately 450–550 ms. Although a difference of around 200 ms between the least and most demanding systems may seem modest, at a travel speed of 80 km/h this

corresponds to an additional 4.5 metres travelled before a driver initiates a response. In practical terms, IVISs that impose greater cognitive workload may slow hazard detection sufficiently to influence whether a driver avoids or is involved in a collision.

1.1. Eye-tracking Data preparation

Protocol and processing pipelines for the eye-tracking data are currently being finalised. Detailed data cleaning and reduction procedures will be documented in this section in a future update.

10.4 Eye-tracking data

Results for eye-tracking are currently being processed and will be integrated into the final report.

10.5 Vehicle Telemetry Data preparation

Driving telemetry data were extracted from the VBox system for all three phases, and from the Oxford system for Phases 1 and 3. Two metrics were included in the analysis:

1. *mean lateral acceleration* (g) reflects the average lateral g-force exerted on the vehicle, where 1 g equals the acceleration due to gravity (9.81 m/s^2). This metric represents the side forces acting on the vehicle during steering and indicates how well the driver maintained stable lateral control; higher g-forces correspond to greater side-to-side movement and reduced lane-keeping stability.
2. *mean speed delta* (m/s) captures the point-to-point change in vehicle speed at a 100 Hz sampling frequency, indexing how frequently and sharply the driver accelerates or decelerates.

Both measures were selected based on subject matter expert guidance, as both are expected to increase with higher cognitive workload.

10.6 Vehicle Telemetry Data

To ensure that differences in hardware did not bias the telemetry outcomes, we evaluated whether the two data-logging systems (Oxford and VBox) produced different values for our driving measures. The Toyota Hilux was the only vehicle instrumented with both systems, allowing a within-vehicle comparison while holding IVIS, task type, and driver sample constant. Mixed-effects models were fitted separately for each telemetry variable, with telemetry system and task type as fixed effects and participant as a random intercept. For mean speed delta, the telemetry-system effect was non-significant and trivially small (System = -5.34×10^{-5} , $p = .705$). Likewise, the mean lateral acceleration model showed no reliable

difference between systems (System estimate ≈ 0 , $p > .30$), indicating that neither device produced systematically larger or smaller lateral-movement readings. Across both models, participant-level variance was near zero and no task-type effects emerged, reflecting highly stable driving behaviour and minimal measurement noise. Together, these findings confirm that Oxford and VBOX yielded equivalent telemetry measurements for both speed and lateral-acceleration metrics, and therefore measurement-system differences do not influence the distraction-related analyses or conclusions.

The mixed-effects model examining mean lateral acceleration evaluated whether IVIS system or IVIS $\times$ task-type combinations were associated with differences in lateral-vehicle movement. The fixed-effect estimates were extremely small across all systems, and none of the IVIS or interaction terms reached statistical significance (p 's $> .15$). Participant-level variance was negligible, indicating that drivers behaved similarly on this measure regardless of system or task. Residual variance was also very small, demonstrating that mean lateral acceleration was a highly stable driving-performance metric with minimal variability. Overall, the model showed no evidence that IVIS interactions meaningfully influenced lateral vehicle control, and the measure was not sensitive to distraction-related differences between systems.

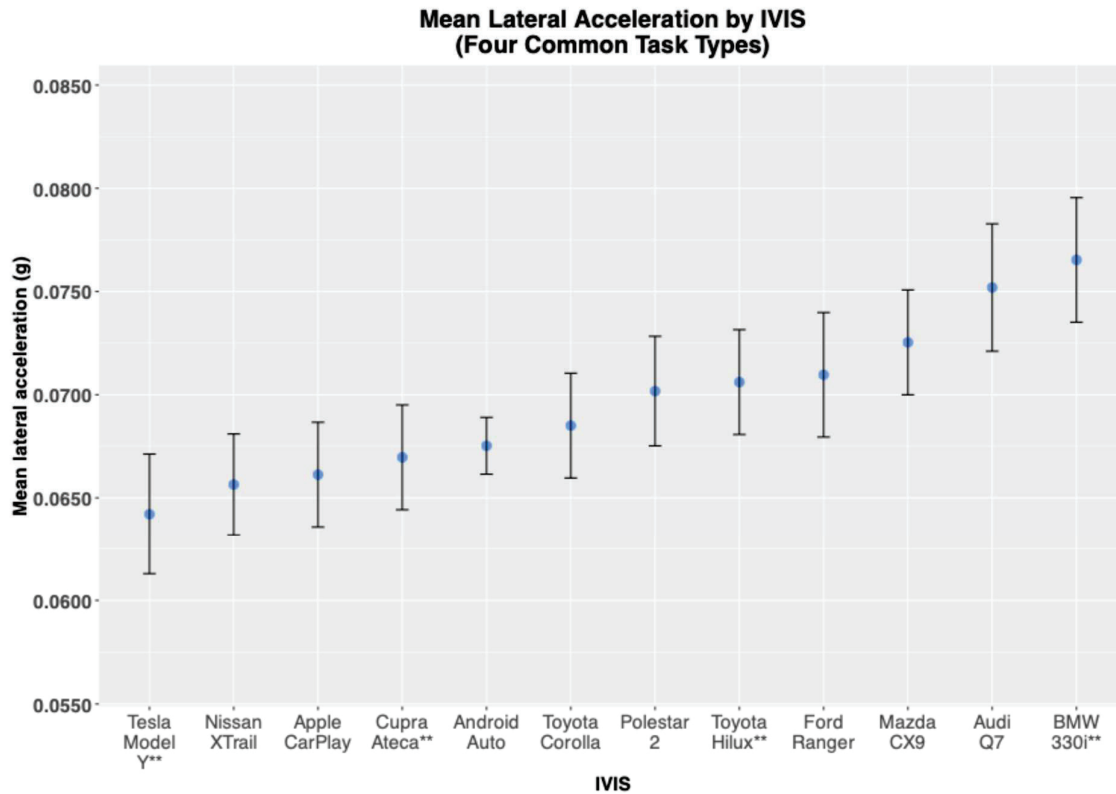


Figure 9. Estimated mean lateral acceleration (g) for each IVIS system with ± 1 SE error bars. All systems produced extremely small effects (typically < 0.01 g), and the substantial overlap of error bars indicates no meaningful degradation in lane-keeping performance. Systems marked with “**” were measured using the VBox device (BMW 330i, Cupra Ateca, Tesla Model Y, and Toyota Hilux). All remaining systems were fitted with the Oxford OXTS unit.

The mixed-effects model assessing speed-maintenance behaviour in metres per second examined whether IVIS system or IVIS $\times$ task-type combinations were associated with deviations from the target speed. The fixed-effect estimates were extremely small in magnitude (typically on the order of 10^{-4} m/s), indicating that differences between systems corresponded to fractions of a millimetre per second, far below any threshold of practical driving relevance (See Figure 10 for model estimates). Almost all IVIS and interaction terms were statistically non-significant, with the exception of Android Auto and the Android Auto $\times$ Initiate Text interaction; however, even these effects were trivially small and reflect statistical artefacts rather than meaningful behavioural differences. Participant-level variance collapsed to zero, demonstrating that drivers showed virtually identical speed-maintenance behaviour regardless of IVIS, and residual variance was minimal. Overall, the model provides no evidence that interacting with any IVIS meaningfully affects a driver’s ability to maintain a stable speed. See Appendix 12 - Vehicle Telemetry Models Output for a summary of vehicle telemetry modelling.

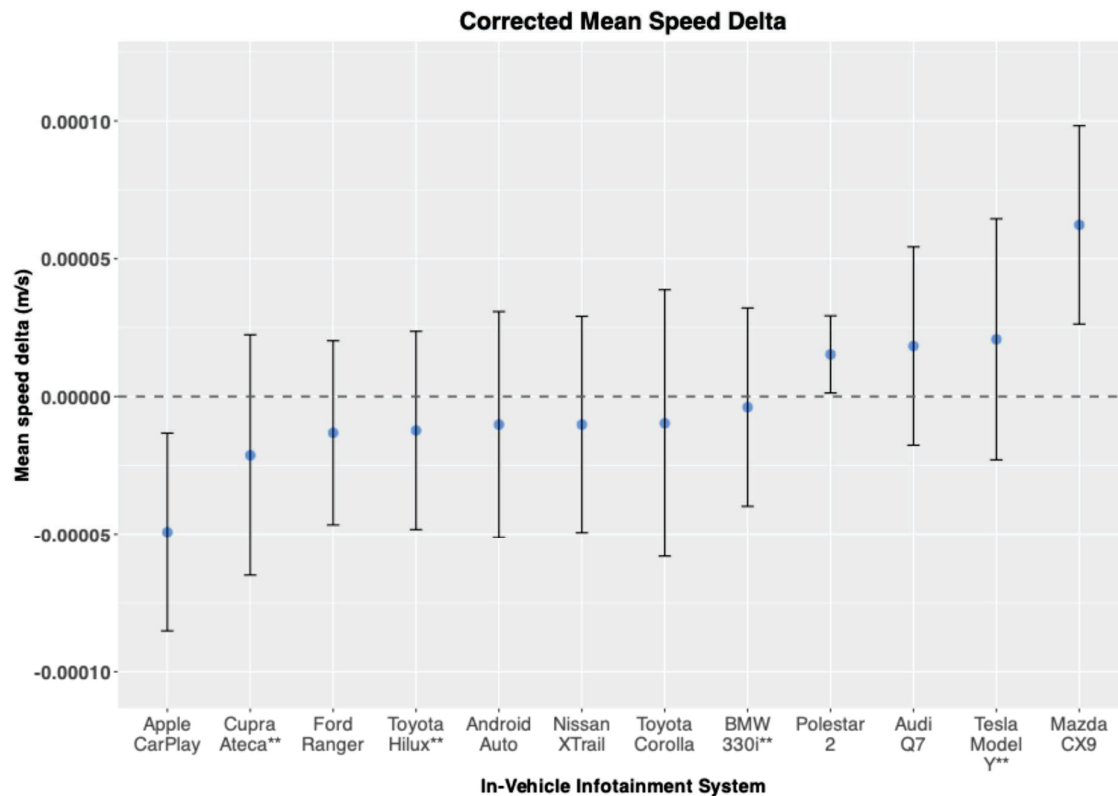


Figure 10. Mean speed delta (m/s) by IVIS system, with error bars representing ± 1 standard error of the model estimate. The small magnitude of the error bars and their broad overlap across systems indicate that moment-to-moment speed fluctuations were essentially equivalent, regardless of IVIS platform. Systems fitted with the VBox device are denoted by “**”; all remaining systems used the Oxford OXTS unit.

11 Discussion – Phase 1 On-Track Test

We conducted on-track testing to evaluate the effectiveness of the test protocol for assessing IVIS distractibility in 12 vehicles available in the Australian market. Using the Detection Response Task (DRT), an industry-standard measure of cognitive workload in real-world tasks, we assessed how different IVISs and IVIS task types influenced driver distraction. The DRT results strongly support its use as a key indicator of IVIS distraction. The mixed-effects model revealed substantial and statistically reliable differences in DRT miss rate between IVIS systems, demonstrating that the DRT is sensitive to system-level design features that influence cognitive workload. High-load systems such as the Tesla Model Y, Nissan X-Trail and Toyota Hilux produced significantly elevated miss rates, whereas low-load systems such as Apple CarPlay and Polestar 2 did not. The DRT also responded appropriately to task demands, with navigation and media interactions increasing miss rates for several systems. This pattern indicates that the DRT both discriminates effectively between IVIS designs and provides a valid measure of cognitive workload,

supporting its suitability as a core metric in the IVIS distraction rating framework. On-track tests provide 'ground truth' or benchmark for IVIS distractibility we aimed to predict from the phase 2 testing.

The analysis of vehicle-control data provided little evidence that driving performance was sensitive to changes in cognitive workload induced by IVIS tasks. According to the cognitive control hypothesis (e.g., as applied in Strayer & Cooper, 2017), cognitive load primarily disrupts behaviours that rely on controlled, non-automated processing, whereas highly practised behaviours, such as maintaining lane position and regulating speed on a predictable, low-demand roadway, are comparatively resistant to moderate increases in workload. Under such conditions, drivers rely on automated control processes that show limited sensitivity to secondary-task interference unless workload approaches extreme levels. This aligns with broader findings in the driver-state monitoring (DMS) literature: lane-keeping, steering variability, and speed regulation typically change only when attentional impairment becomes substantial, which is why contemporary DMS frameworks emphasise ocular and facial indicators over telemetry-based measures (Bergasa et al., 2006; Status of Driver State Monitoring Technologies and Validation Methods, 2020). In our study, participants were explicitly instructed to prioritise safe driving, further reducing the likelihood that moderate IVIS-related workload would manifest in degraded vehicle control. Moreover, detecting telemetry-level impairment would require drivers to operate close to a critical-event threshold; an ethically unacceptable scenario that cannot be experimentally induced. Consistent with this, our telemetry analyses did not reveal meaningful differences between IVIS systems, and the small variations observed were statistically unreliable and fell within normal driving variability. Taken together, these findings indicate that telemetry measures are not well suited to detecting moderate IVIS-related workload, and the report therefore places greater emphasis on alternative distraction metrics that demonstrated clearer sensitivity to IVIS demand.

On-track testing is essential to understand the parameters and metrics of IVIS distractibility. However, the on-track procedures used in Phase 1 are often too costly to serve as a practical, routine method for assessing distraction in the field. Phase 2 therefore examined whether a streamlined version of the Phase 1 assessment protocol could produce comparable results and valid indicators of cognitive workload. This phase enabled the evaluation of a simpler, shorter, non-driving approach for assessing the distractibility of infotainment systems, both now and in future applications. We removed all driving elements from the test procedure and instead used a stationary, in-vehicle setup incorporating secondary tasks designed to simulate the key cognitive components of driving.

Note: Detailed eye-tracking results and an accompanying discussion of visual attention are currently being finalized. These findings will be integrated into a subsequent revision of this report.

12 Phase 2: Assessment of cognitive and visual demand of IVISs in a stationary setting.

Phase 2 was one-step removed from Phase 1 testing; we explored if IVIS distractibility could be assessed while a vehicle was stationary. Rather than drive a vehicle, participants completed computer-based tasks and interacted with an IVIS while the vehicle was stationary in a carpark at the AARC. This is a novel approach as stationary testing and computer-based tasks haven't been used to assess IVIS distractibility. While our stationary tests require some specialised equipment and knowledge of statistical analysis techniques like those used during Phase 1, technical users or assessors would be more than capable of conducting stationary assessments with the added benefits of reduced test costs and safer operations compared to on-track testing.

13 Method – Phase 2

Participants completed two separate stationary tasks during Phase 2 testing; 1) the visual occlusion test (VOT) – a test that measures the visual demand of IVIS tasks, and 2) a simulated driving task (Tracking Task) that was used to measure cognitive workload and visual attention required to complete an IVIS task. We fitted vehicles with a forward-facing GoPro camera to record participants interacting with the IVIS and to record researcher verbal instructions. The VOT and the tracking task were controlled by a DELL XPS 13-inch 2-in-1 laptop. Participants followed the General IVIS Assessment Procedure outlined above.

13.1 Visual Occlusion Test

The visual occlusion test (VOT) is an ISO-standardised test that estimates the visual demand of driving or, as in our study, the visual demand of IVIS interactions. The VOT allows us to determine safe attention allocation to in-car tasks while 'driving'. Participants or 'drivers' sit in a stationary car and wear visual occlusion goggles (VOGs) that block their field of view. Participants operate the VOGs with a small micro-switch strapped to their index finger. When participants need to look at the IVIS, they press the switch to open the

VOGs for a pre-programmed interval. The VOG ‘open’ interval for a task (i.e., total shutter open time or TSOT) provides a measure of how much time a ‘driver’ spends looking at the IVIS, which is an indication of how long their eyes are “off-road”. Conversely, the length of the interval when the VOGs are closed (i.e., total shutter close time or TSCT) provides a measure of how much time a driver spends looking “on-road”.

13.2 VOT Procedure

To begin each VOT trial, participants were instructed to place their hands on the steering wheel to mimic real driving. The visual occlusion goggles (VOGs) were programmed to run the “peek” protocol, which differs from the “cycle” method (outlined in the NHTSA distraction guidelines) in that the peek method is controlled by the participant, whereas the cycle method “cycles” between open and closed automatically. When participants opened the VOGs, a single cycle progressed with a viewing interval (i.e., goggles open) of 1500ms and closed for a minimum of 1500ms immediately after the viewing interval. Participants could open the goggles as many times as required for a task and were instructed to *only* open the goggles when they needed to look at the IVIS or other controls. Given our instructions, participants could choose to keep the VOGs closed when completing an IVIS task. For example, when receiving a phone call, participants could press the steering wheel microphone button to answer the call and keep the VOGs closed, which would result in a TSOT of 0 seconds for that task. Once participants completed a trial, they were to place their hands back on the steering wheel in their driving position. We chose to initiate and end the VOT at the beginning and end of each block of trials (i.e., for each IVIS task type). The VOT logging system recorded continuous data over each block but did not mark individual trial boundaries. Each block was designed to include 10 trials. Therefore, we treated the block as the unit of analysis and normalised block-level totals (e.g., TSOT) by dividing by 10 to obtain per-10-trial-equivalent values. This approach preserves relative differences between IVISs and task types but introduces some uncertainty into the absolute per-trial estimates, which we acknowledge as a limitation.



Figure 11. Visual Occlusion Test — Visual demand was assessed using vision occlusion goggles (VOGs) that limit participants' access to the IVIS. A microswitch mounted on the index finger enabled participants to momentarily "open" the goggles to view the interface during task completion. The cumulative duration of these openings represents the visual-attentional demand imposed by each IVIS.

13.3 Simulated Driving tasks - **Tracking Task** with in-built visual DRT

The Tracking Task (as it will be referred to from here) is a computer-based task that induces cognitive workload to simulate the demands of on-road driving. The task ran on a laptop computer that was affixed to a vehicle's windscreen directly in front of the participant (See Figure 8 below for tracking task setup) and it is comprised of two components designed to assess different aspects of performance:

- **Tracking Task** - The "tracking" component is a continuous task that simulates driving. Participants control a needle using a keyboard and are to align the needle with a target that moves or 'tracks' left and right across the computer screen. Tracking the target demands visuo-spatial awareness and motor control as the target changes directions at random intervals. The participant's needle can easily become mis-aligned (i.e., in an "error" state) with the target when the participant looks at the IVIS (i.e., away from the computer screen or "roadway"). This error state is known as "tracking error" and it is our outcome variable of interest. **Tracking error** is recorded continuously over the course of the task at 60Hz and is expressed as the number of pixels (px) between the target and the needle.
- **Visual Detection Response Task (vDRT)**: a visual DRT is built into the tracking task to measure residual cognitive workload. The vDRT uses a visual stimulus of the DRT which is programmed to the same ISO standard as the DRT used in Phase 1. Standard DRT measures, **misses** and **RTs**, are derived from the vDRT.

13.4 Tracking Task Procedure

Participants were told to prioritise the tracking component as if they were driving a car, to interact with the IVIS (second priority) and respond quickly as possible to the vDRT (third priority). Participants controlled the tracking and vDRT components of the tracking task with their right hand and used their left hand to interact with the IVIS. Participants were told not to respond to the task during breaks between blocks.



Figure 12. Stationary testing with the Tracking Task in the AARC carpark. The Tracking Task is a computer-based task that requires a laptop computer, keyboard and a windscreen mount to hold the laptop in front of the participant. The “tracking” component of the task is a continuous task that simulates driving.

14 Results

14.1 VOT Pre-Corrected Data

We used total shutter open time (TSOT) in milliseconds as our outcome variable of visual attention. We recorded participant-level TSOT for each IVIS task type i.e., a block of trials. All VOT TSOT data reported below are square root transformed (see Data Analysis and Modelling for details).

Broadly speaking, systems such as those in the Tesla Model Y, Cupra Ateca, Mazda CX-9, Audi Q7 and Nissan X-Trail required substantially more visual attention in the VOT, as indicated by their longer observed TSOT values in Figure 11. In contrast, Android Auto, Apple CarPlay, the Polestar 2 and the Toyota Corolla native system produced comparatively short TSOTs, reflecting lower visual-attentional demands. Task-level patterns also followed expected trends: Navigation and Media interactions were among the most visually demanding, whereas receiving calls or texts required the least eyes-off-road time. While these empirical patterns provide useful descriptive insight, meaningful comparisons between visual-attention demand (VOT) and cognitive workload (on-track DRT) are based on the modelled outcomes rather than raw scores.

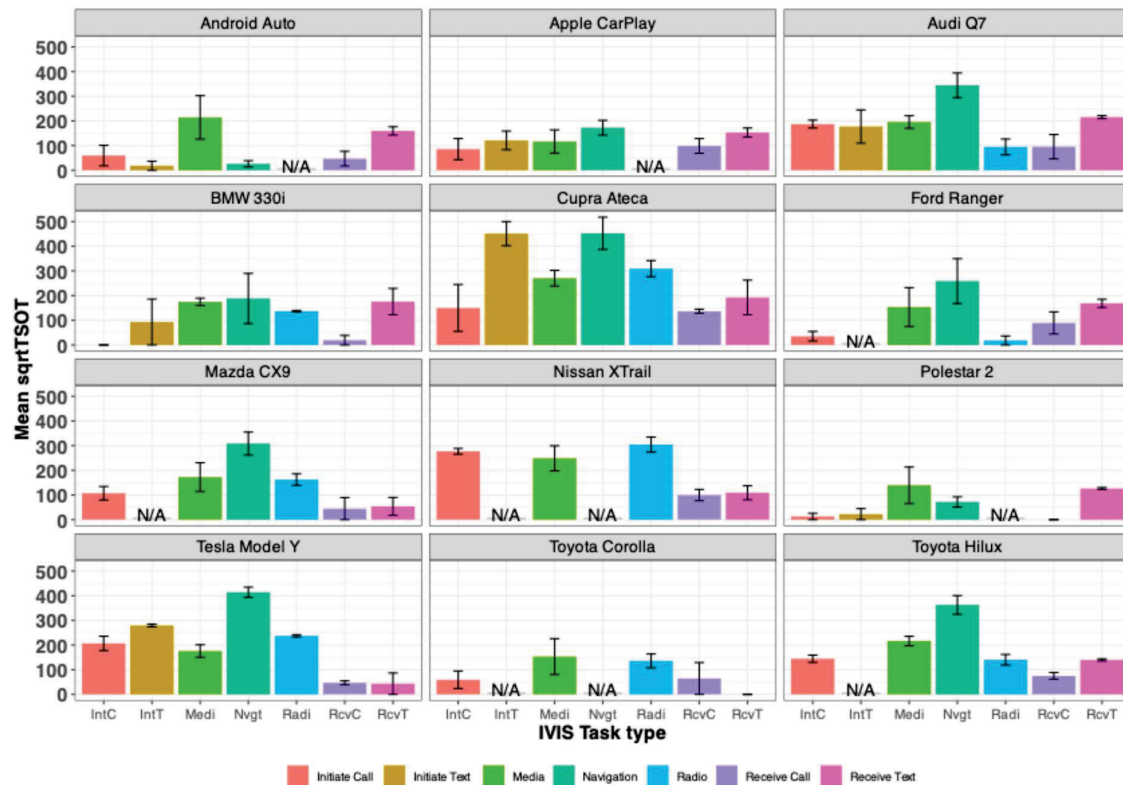


Figure 13. Square-root-transformed, pre-corrected Total Shutter Open Time (TSOT) for each IVIS and IVIS task type in the Visual Occlusion Test (VOT). The figure displays the observed TSOT values on the variance-stabilised (square-root) scale used for subsequent modelling, providing a baseline view of visual-attentional demand prior to statistical correction. Standard-error error bars show the between-participant variability in the observed TSOT data.

14.2 Tracking Task Pre-Corrected Data

We recorded misses and RTs from the vDRT and tracking task tracking error for each IVIS task type i.e., across a block of trials for each participant. All tracking data reported below are the mean for each IVIS task type within each IVIS. The error bars represent the standard error of those means.

Figure 11 shows the pre-corrected data for the visual DRT. Android Auto, Apple CarPlay, Polestar 2, Mazda CX9 and the Nissan X-Trail demanded more visual attention, which we infer from higher proportions of misses in Figure 11. In comparison, the systems in the Audi Q7, Toyota Corolla and the Tesla Model Y demand low levels of visual attention i.e., low mean proportion of misses. Navigation (Nvgt), Media (Medi) and Radio (Radi) IVIS task types demanded the highest levels of visual attention (i.e., higher proportion of misses). Conversely, receiving a call or a text demanded the lowest levels of visual attention, which is consistent with the low cognitive demand for those IVIS tasks during Phase 1 and for the VOT during Phase 2.

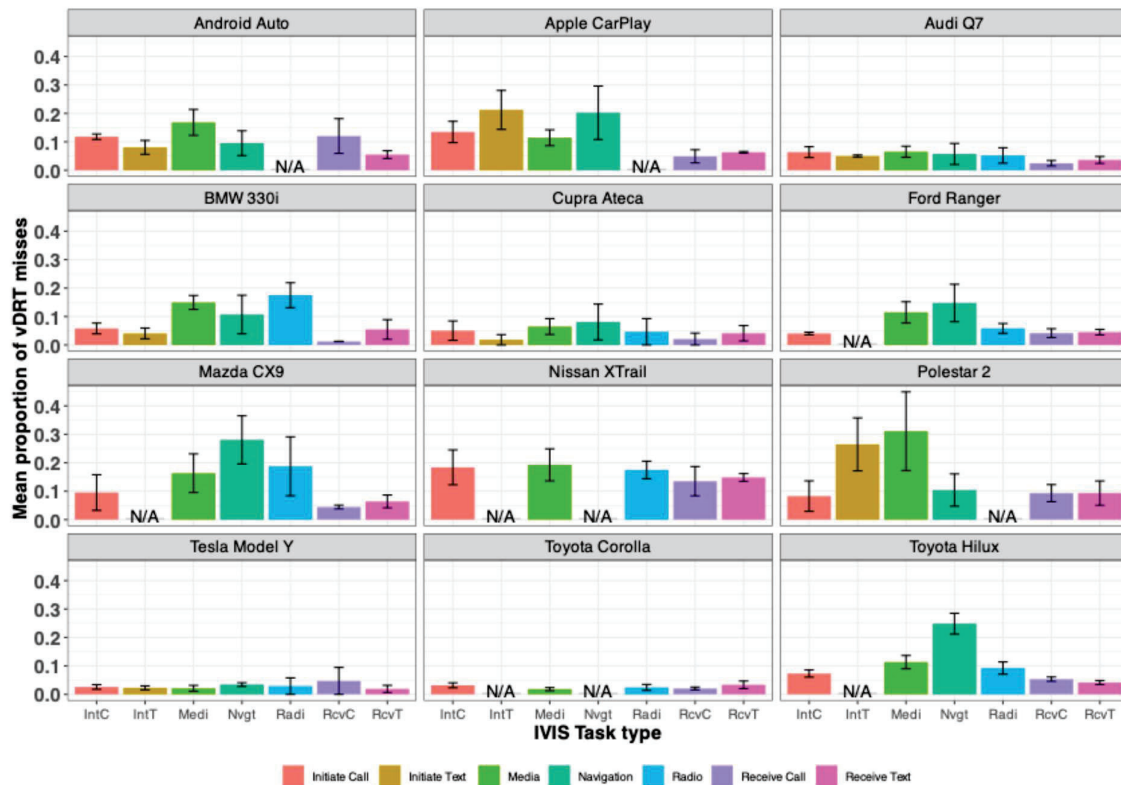


Figure 14. Mean proportion of misses on the visual Detection Response Task (DRT) for each IVIS task type, aggregated across all IVISs. The figure displays the pre-corrected miss rates, offering a direct representation of the visual attentional load associated with each task type before statistical modelling or correction procedures.

Response times to the visual DRT were largely uninformative (Figure 13). The observed RTs showed minimal variation between IVISs and across task types, indicating that response speed to the DRT did not meaningfully differentiate the visual demand of the systems. This pattern is consistent with the on-track DRT findings, where RTs similarly showed little sensitivity to differences in IVIS distractibility.

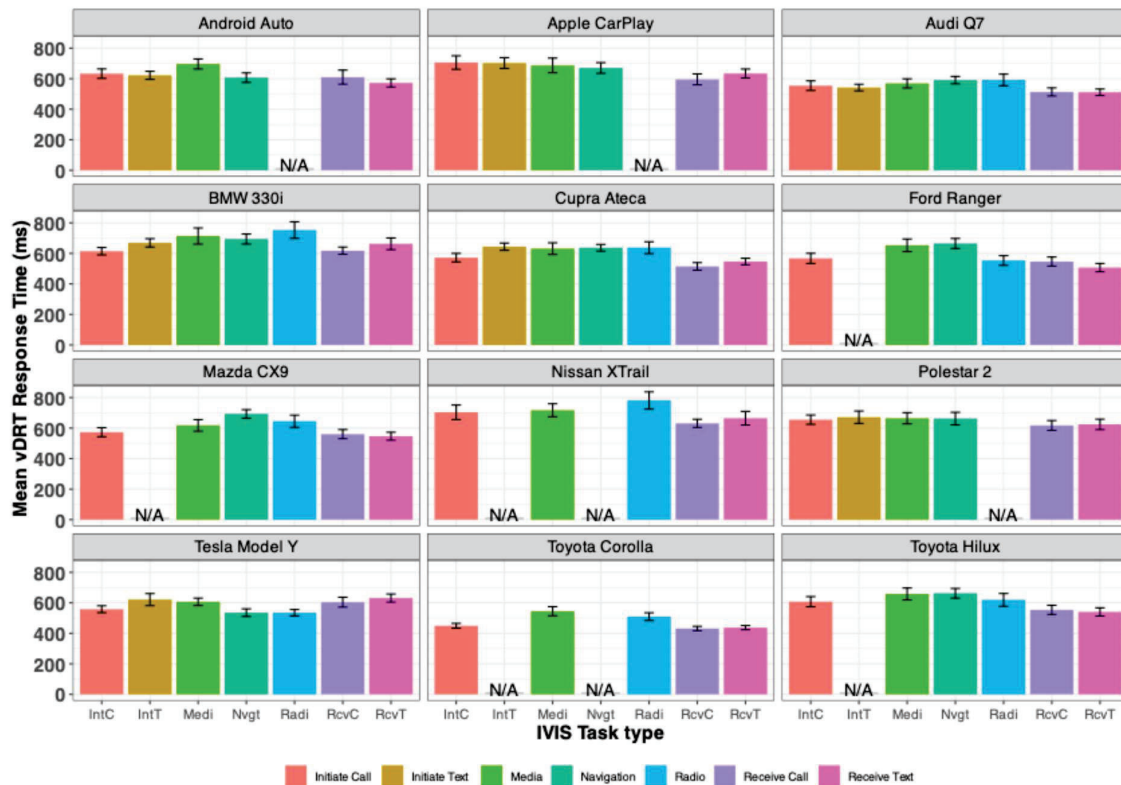


Figure 15. Pre-corrected mean response times to the Tracking Task visual DRT. These observed data reflect the raw variation in drivers' response times across IVIS systems and task types, prior to modelling.

Figure 14 presents the pre-corrected tracking error data. Tracking error served as a reasonable indicator of visual attention, with higher error reflecting greater visual demand imposed by an IVIS. The Ford Ranger, Mazda CX-9, Nissan X-Trail and BMW 330i native systems showed the highest levels of tracking error, indicating they required more visual attention to operate. In contrast, Apple CarPlay, the Toyota Corolla, the Tesla Model Y and the Polestar 2 produced the lowest tracking error values, suggesting comparatively lower visual demand. Consistent with patterns observed for the other distraction outcome measures, Navigation and Media tasks were the most visually demanding task types, whereas Receive Call and Receive Text tasks required the least visual attention.

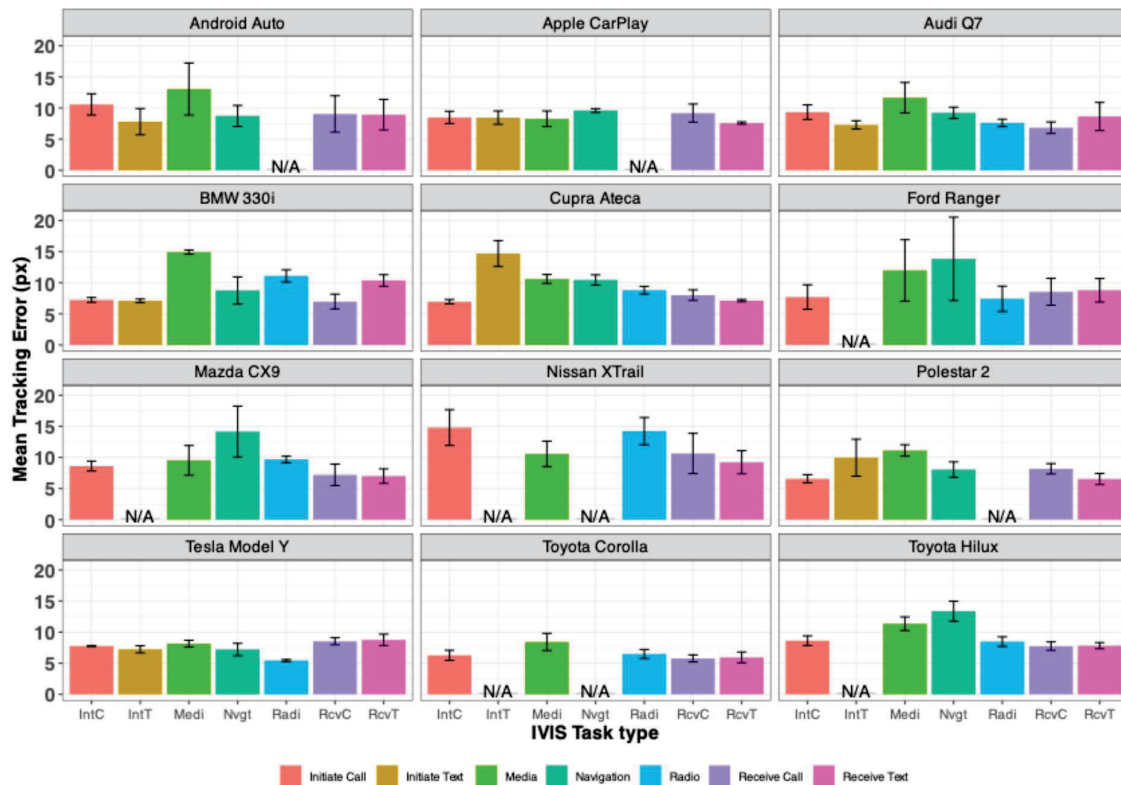


Figure 16. Pre-corrected mean tracking error from the Tracking Task for each IVIS and IVIS task type. These values represent the pre-correct data before applying statistical models or correction procedures, providing a direct depiction of the raw tracking performance associated with each system.

14.3 Modelling Off-Track Stationary Data

Figure 15 shows the pre-corrected VOT TSOT data (top panel) and the model-estimated marginal means (EMMs) for TSOT after statistical correction (bottom panel). The mixed-effects model was fitted to square-root-transformed block totals to stabilise variance. For interpretability, the fitted values were back-transformed by squaring the EMMs and dividing by ten (corresponding to the ten trials per block), yielding predicted mean eyes-off-road time per task in milliseconds. The back-transformed estimates indicate substantial variation in visual demand across systems. The Tesla Model Y and the Nissan X-Trail each require approximately 3,800 ms (3.8 s) of eyes-off-road time per task, whereas Android Auto, the Polestar 2 and the Toyota Corolla require far less visual attention, approximately 500–560 ms per task. The Cupra Ateca shows even higher predicted TSOT values (exceeding 7,500 ms per task); however, these estimates should be interpreted cautiously (see following paragraph). The Cupra results are presented here only for completeness and should not be considered representative of typical operational performance. Overall, these findings indicate that the least demanding systems require nearly an order of magnitude less eyes-off-road time

than the most demanding systems, underscoring large and practically meaningful differences in visual distraction across IVIS designs.

In addition, the Cupra Ateca produced a VOT TSOT or eyes-off-road time approximately two and a half standard deviations above the mean of all IVISs, placing it well outside the overall distribution. This extreme deviation was traced to repeated system errors during the VOT sessions. As these issues indicated that the observed VOT performance reflected technical malfunction rather than genuine interaction demand, the Cupra Ateca was treated as an outlier for the VOT analysis and removed from all subsequent modelling that depends on valid VOT values specifically, the Phase 2 VOT models (TSOT and step-count), the correlation analyses involving VOT, and the step-modelling and step-prediction models. The Cupra Ateca was retained in the Drive DRT and Tracking Task models, as no comparable technical anomalies were observed during those tests.

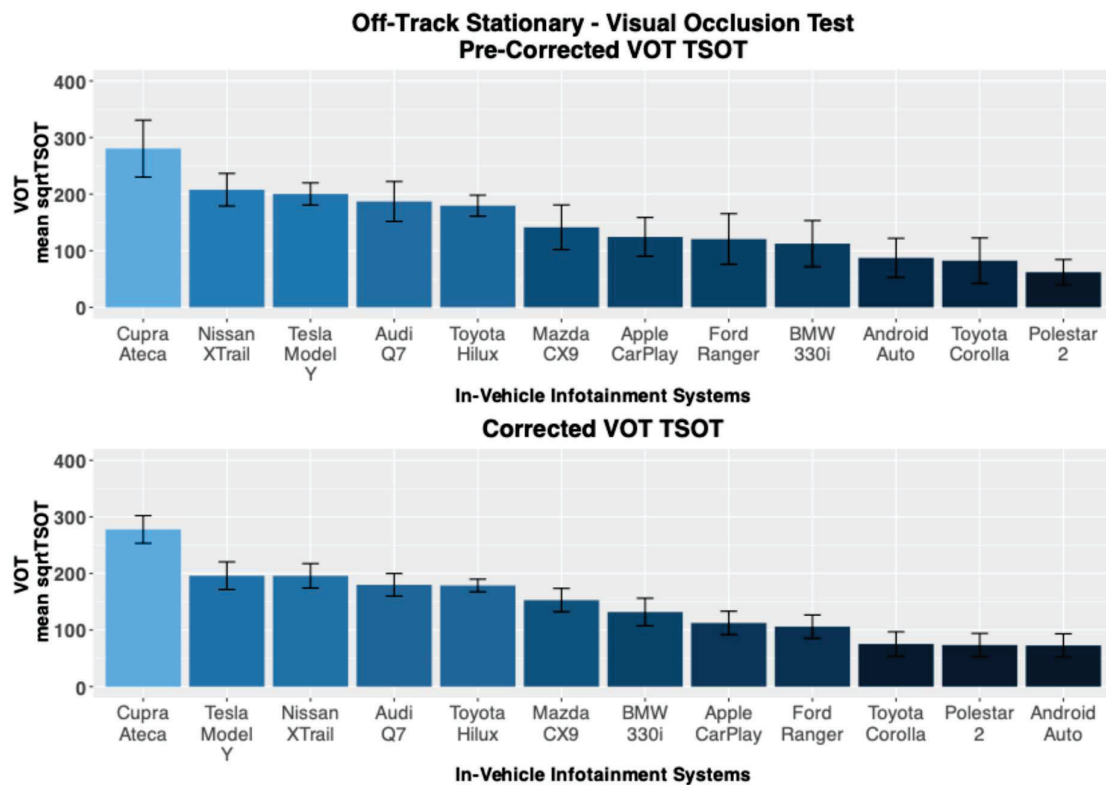


Figure 17. Model-estimated IVIS effects expressed on the square-root-transformed TSOT scale used for analysis. The square-root transformation was applied to stabilise variance in the block-level eyes-off-road time (TSOT). Estimates and standard errors are therefore displayed on the transformed scale; however, all interpretation in the text refers to the corresponding back-transformed values (in milliseconds) representing the average eyes-off-road time required to complete a typical IVIS task.

The Tracking Task Tracking Error model assessed how IVIS interaction demand influenced participants' ability to maintain a continuous-control tracking task, to simulate "driving", while stationary. Figure 18 shows the pre-corrected Tracking Error data (top panel) and the corresponding model corrected estimated marginal means from the mixed-effects model (bottom panel). These marginal means represent the average predicted deviation from the target trajectory for each IVIS, controlling for participant-level variability and task-type effects. Systems such as the Android Auto, Ford Ranger, and Toyota Corolla showed relatively small deviations (approximately 5–7 units), indicating that participants could maintain stable continuous-control performance while interacting with these systems. In contrast, several systems produced substantially greater tracking deviation. The Nissan X-Trail displayed the largest model-estimated deviation at approximately 14 units, more than double that of the lowest-demand systems, while the Cupra Ateca, Audi Q7, Tesla Model Y, Mazda CX-9, and Toyota Hilux showed elevated values in the 8 to 11 pixel range. Although modest in absolute magnitude, consistent with the simplicity of the task, these estimated effects illustrate a clear pattern: IVISs with more complex or effortful interaction demands produce systematically larger continuous-control errors.

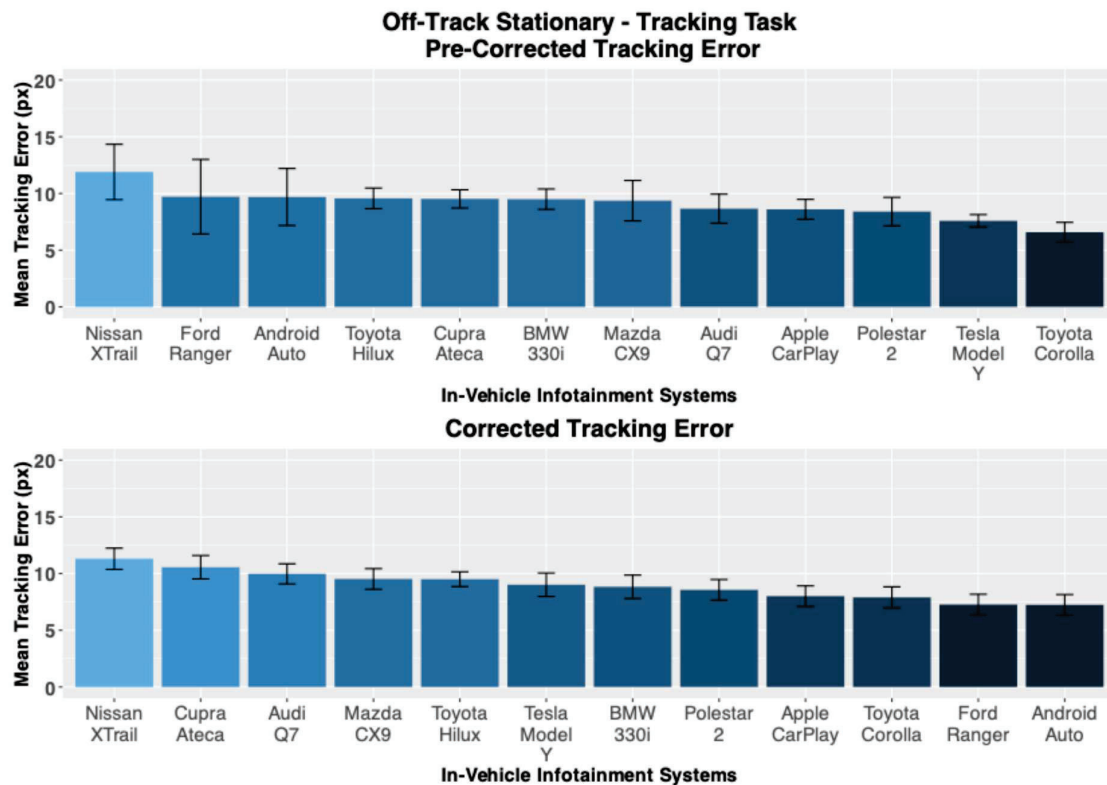


Figure 18. Tracking Task Tracking Error shown as pre-corrected raw data (top panel) and model corrected estimated marginal means with standard-error error bars (bottom panel). The estimated marginal means quantify the predicted deviation from the target path during IVIS interaction. Lower values indicate more stable continuous-control performance. Several systems (e.g., Ford Ranger, Polestar 2, Android Auto) produced consistently low error, whereas others (e.g., Nissan X-Trail, Cupra Ateca, Audi Q7, Tesla Model Y, Mazda CX-9) yielded substantially larger deviations. These patterns demonstrate that more demanding IVIS interfaces produce measurable disruptions to continuous-control performance even in a simplified laboratory task.

The Tracking Task tracking error showed small associations with the on-track DRT measures ($r = .29$ for reaction time; $r = .38$ for misses), indicating that it does not reproduce the same pattern of distraction-related impairment observed under real driving conditions. These modest associations suggest that while systems that impair driving performance may produce slightly greater continuous-control error in the Tracking Task, the effect is weak and unreliable. In addition, the Tracking Task DRT measures (miss rate and reaction time) were negatively correlated with the drive DRT measures ($r = -.37$ for misses; $r = -.27$ for reaction time), further demonstrating that the Tracking Task does not capture the same cognitive demands as real driving. For this reason, the Tracking Error and Tracking Task DRT outcomes are presented for completeness but are not used in subsequent analyses or treated as independent safety-relevant metrics.

By contrast, the step-count estimates derived from the Tracking Task were strongly correlated with the on-track step estimates (see Validating the Step Metric Across On-Track

and Off-Track Tests), showing that interaction demand generalises well across task contexts even when participants complete a simplified “driving-like” task. Accordingly, Tracking Error is retained only as supporting evidence in the step-prediction modelling. All Tracking Task model outputs are provided in the appendices and the associated correlations in Appendix 4 – Correlations of all Distraction Outcome Variables.

15 Stationary Off-Track Tests as valid substitutes of On-Track assessment

Phase 2 stationary testing demonstrated that the VOT and Tracking Task are sensitive to differences in IVIS distractibility. Building on this, the Phase 2 method was designed as a simpler, non-driving alternative to the on-track assessments used in Phase 1. We now investigate whether Phase 2 stationary tests provide valid approximations of on-track driver workload. Specifically, we examine the correspondence between model estimates of IVIS distractibility from Phase 1 and those derived from Phase 2 through correlation analyses.

We examined correlations among all linear mixed-effects model estimates for each outcome variable collected in Phases 1 and 2; however, only significant correlations are reported here (see Appendix 6 for the full set of correlations). Figure 16 shows a strong relationship between Drive DRT misses and the stationary VOT TSOT ($r = 0.70$). IVIS systems that impose higher cognitive workload during the on-track drive also tend to require greater visual attention during stationary testing. Similarly, Figure 17 shows a moderate correlation between Drive DRT reaction time and VOT TSOT ($r = 0.62$).

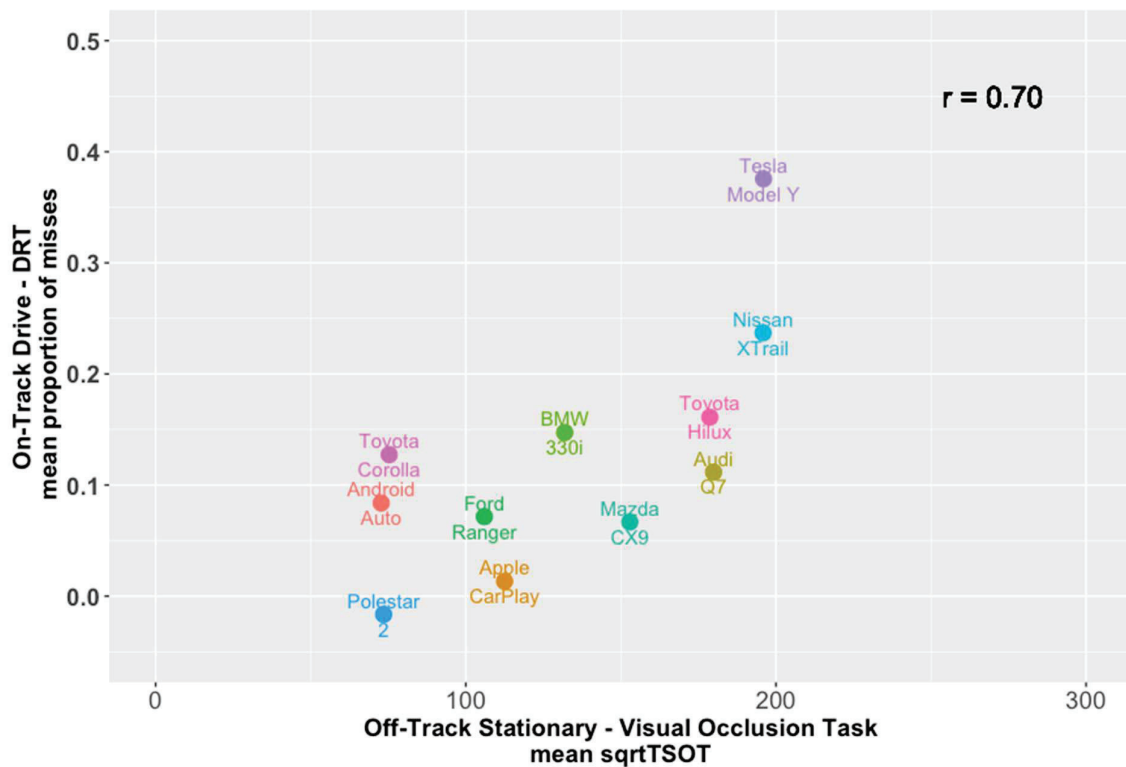


Figure 19. Correlation of model estimates for distractibility measures from On-track testing (Mean DRT Proportion of misses, Y-Axis) and Off-track testing (Mean VOT sqrtTSOT, X-Axis). Drive DRT miss and stationary VOT sqrtTSOT measures are highly correlated ($r = 0.70$); infotainment systems that induced higher levels of cognitive workload during on-track tests tended to demand higher levels of visual attention, or “eyes-off-road” time, during stationary tests. This result is consistent with our hypothesis that the stationary testing would complement the on-track tests, suggesting that a stationary test could be a realistic proxy for on-road assessment.

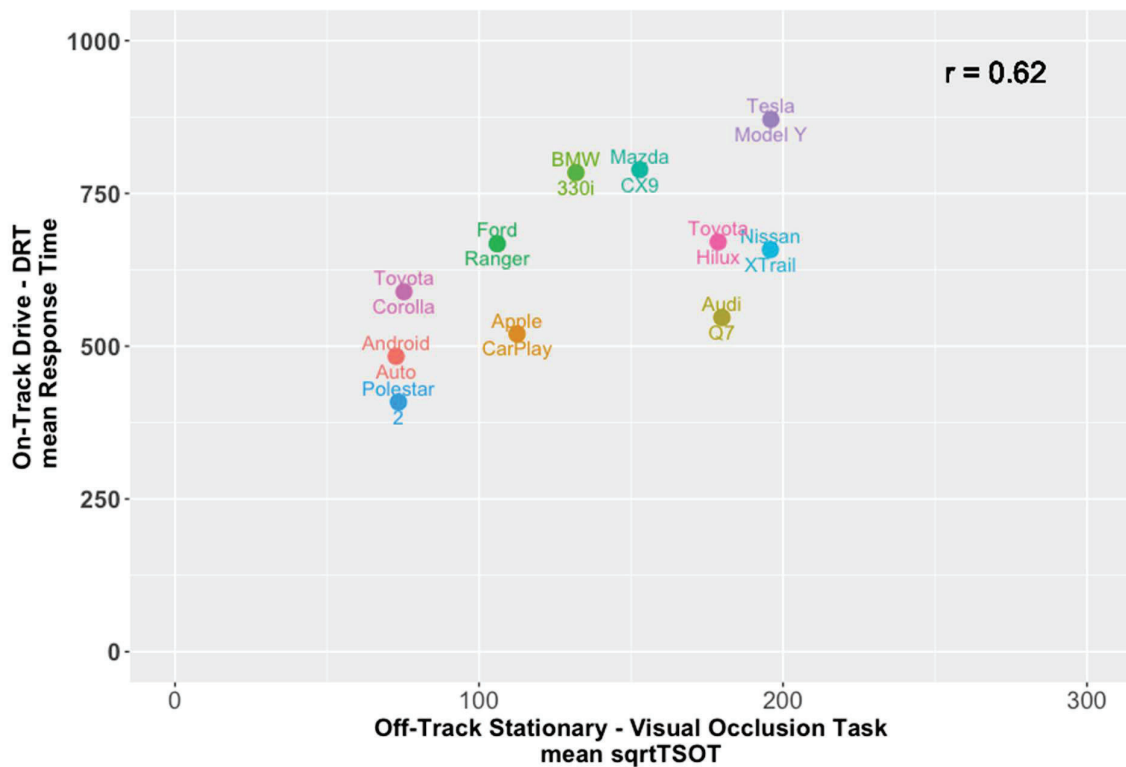


Figure 20. Correlation between Drive DRT reaction time and VOT TSOT. A moderate positive association ($r = 0.62$) indicates that systems requiring longer reaction times during the on-track DRT also tend to demand greater visual attention during the stationary VOT.

In addition, VOT and Tracking Task tracking error from Phase 2 were strongly correlated (see Figure 20). IVISs that required high levels of visual attention in the VOT also produced elevated tracking error during the Tracking Task. Because both outcome measures reflect the amount of time a participant's "eyes are off the road," this relationship is expected, but it nonetheless provides important confirmation that the two tasks index the same underlying construct. Together, these results provide strong evidence that both the VOT and the Tracking Task are sensitive to the visual attention demands imposed by different IVISs, and their potential to distract drivers.

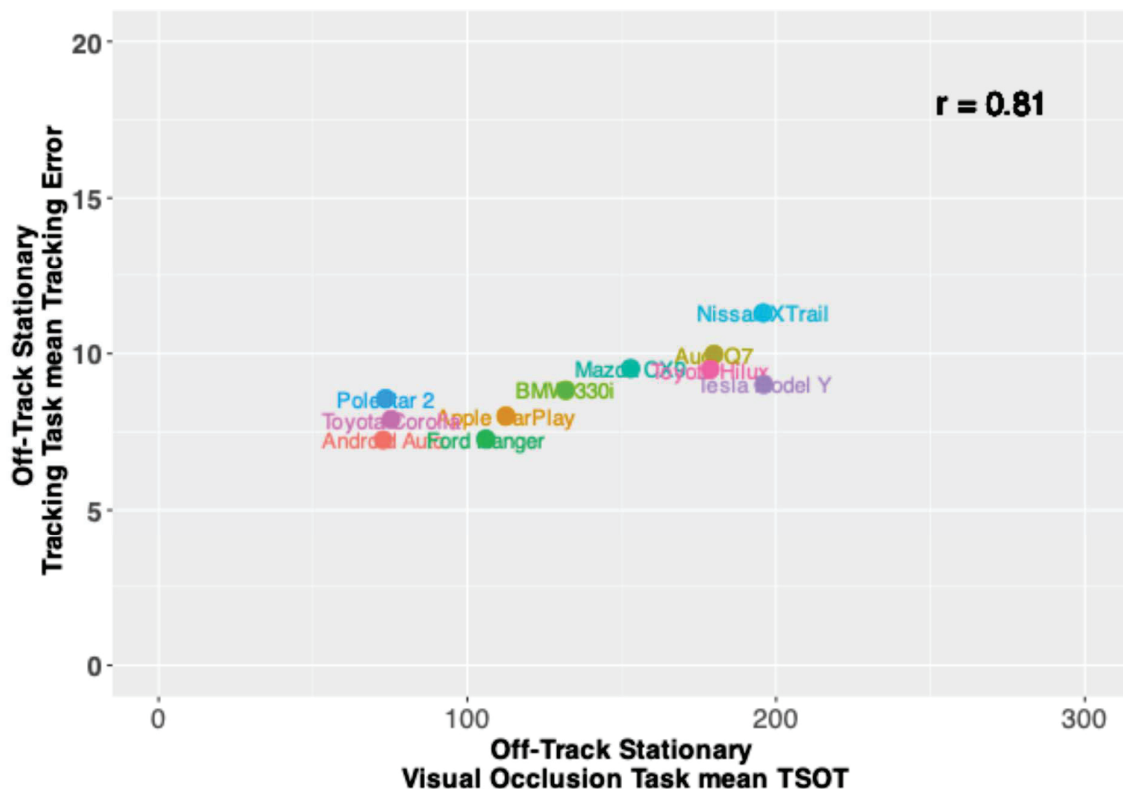


Figure 21. Correlation between VOT TSOT and Tracking Task tracking error from Phase 2. A strong positive association indicates that IVISs demanding greater visual attention in the VOT also elicited higher tracking error in the Tracking Task, showing that both measures capture similar aspects of visual distraction.

Correlations of model estimates from Phase 1 and Phase 2 provide evidence of the convergent validity of the stationary tests, indicating the extent to which they approximate the workload patterns observed in on-track performance. This establishes the link between the Phase 1 and Phase 2 test methods and achieves a primary aim of the project; stationary tasks can serve as reliable substitutes for on-track testing of IVIS distractibility. The stationary tests can be deployed with little financial outlay, in contrast to on-track testing that requires substantial costs for track hire, equipment and personnel. The stationary tests are more accessible and could be adopted across a broader spectrum of contexts and by technical and non-technical users.

16Phase 3: A practical cost-effective solution for assessing distractibility of an IVIS

While the detailed on-track and stationary tests used in Phases 1 and 2 yield rich data for researchers, OEMs, and independent vehicle-safety assessment bodies such as ANCAP, the resulting distraction metrics are not easily interpretable by consumers. There is a need for a

simple assessment method that produces clear, intuitive outputs so that new-car buyers can meaningfully compare the distraction potential of different IVIS systems. Consumers require straightforward information that communicates relative safety without the need for technical expertise or familiarity with experimental measures such as TSOT, DRT lapses or tracking error.

The purpose of Phase 3 was therefore to develop an easy-to-use metric that could translate the complexity of the earlier phases into a practical, accessible assessment tool for end users. To achieve this, we conducted exploratory modelling on the Phase 1 and 2 data to determine whether a readily obtained measure could reliably predict IVIS distractibility. This represents a novel contribution to the driver-distraction literature, offering a simplified and scalable method that has not previously been available to Australian consumers or industry.

During the project's planning phase, we initially considered examining the mode of IVIS interaction as a predictor of IVIS distractibility. However, as detailed in the 8.5 Modes of IVIS interaction section, the wide variation in how participants could interact with each system, combined with system-specific constraints and the large number of possible interaction configurations, made this approach impractical and of limited value. After extensive discussion and preliminary data analysis, we concluded that a more robust and generalisable metric was needed. As a result, we focused on quantifying the number of IVIS interactions or "steps" required to complete each function, as this measure provides a consistent basis for comparing the systems in our sample and can also be applied to new or emerging interfaces. Crucially, the step metric can be obtained by anyone, anywhere, at any time, and captures the core driver workload associated with completing an IVIS task without requiring specialised equipment or controlled testing environments. This makes it a uniquely scalable and future-proof indicator of distractibility: as new IVIS functions and modes of interaction appear, they can be readily assessed using the same straightforward approach. Accordingly, the number of interactions, or "steps," forms the foundation of the modelling and analyses outlined in the following sections and represents the key practical output of this research for consumers, industry, and safety organisations.

17 Method

17.1 Defining the number of IVIS interactions (steps)

We recorded the number of interactions required to complete each IVIS task for every participant across on-track (Phase 1) and off-track (Phase 2) tests. A step or IVIS interaction

was defined as any engagement with the system, including verbal commands (e.g., “Call Alex”), button presses on the steering wheel or other IVIS controls, touchscreen inputs, or verbal responses to system prompts (e.g., “Would you like to call Alex?”). Figure 11 illustrates how interactions were counted for typical Initiate Call and Media task types. Some tasks, such as making or receiving a phone call, typically required few interactions (< 2), whereas tasks like playing a song (Media), in which participants cycled through a playlist, involved substantially more interactions (> 10). Variability was also influenced by differences in an IVIS’s ability to recognise natural speech, as some systems required participants to repeat commands, adding extra steps. Importantly, all instances in which the system failed to recognise a command, requiring the participant to repeat the input, were counted as additional steps. These “failed” or repeated interactions were treated the same as successful commands because they still imposed visual, cognitive, and temporal demands on the driver. Step counts were obtained by reviewing video recordings from all test phases.

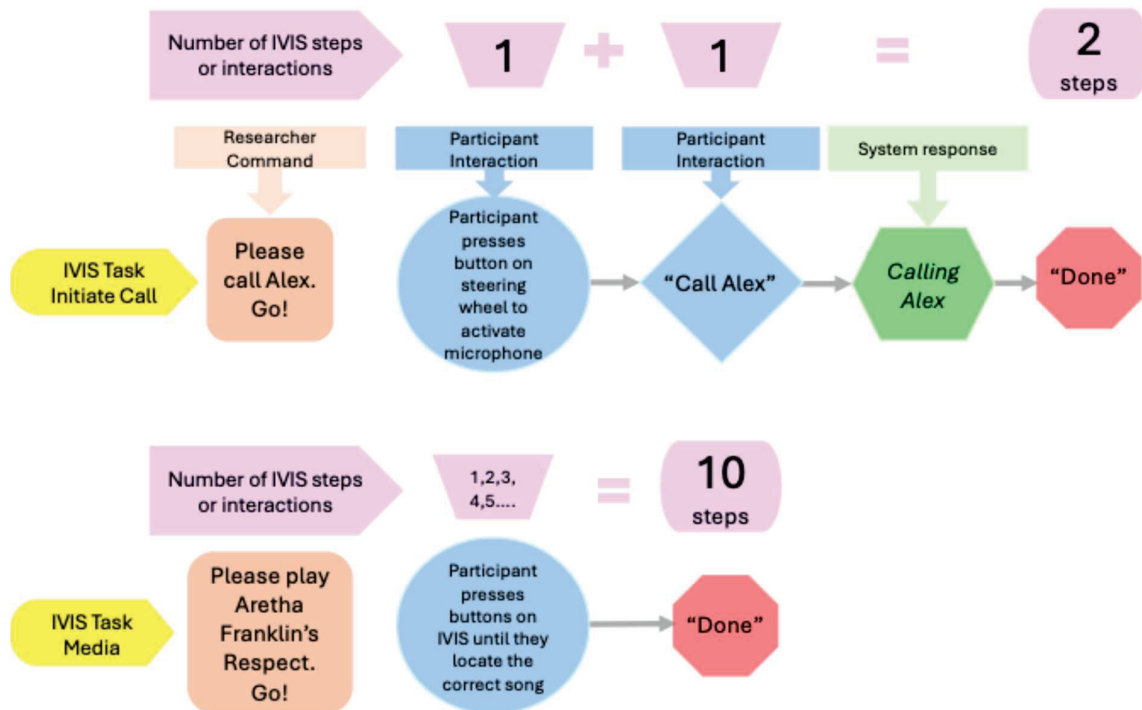


Figure 22. A schematic illustrating how we quantified the number of IVIS interactions, or “steps,” required to complete each IVIS task type. The top half of the schematic shows an example of an Initiate Call task. The sequence begins when the researcher asks the participant to call Alex (0 steps). The participant presses the microphone button on the steering wheel (Step 1) and ends when they instruct the system to call a contact (e.g., “Call Alex”; Step 2). For the Media task type, participants cycled through a playlist track by track, and each button press was counted as an interaction until the correct song was located.

17.2 Procedure

We conducted two complementary analyses to examine IVIS interaction demand (i.e., the number of interactions or steps) and its relationship to distraction outcomes. The first analysis focused solely on the stability of the step metric across test settings. We correlated IVIS-level fixed-effect estimates of mean step count derived from the Phase 1 on-track models with the corresponding IVIS-level fixed-effect estimates from the Phase 2 stationary models. This analysis served as a preliminary “sanity check,” assessing whether systems that required more interactions in on-track driving also required more interactions in stationary testing, and thereby supporting the consistency of the step metric across contexts. In the second analysis, we then modelled participant-level data, using the mean number of steps for each task type within each IVIS as a predictor of distraction outcomes (e.g., DRT misses, VOT TSOT). This approach allowed us to test the predictive relationship between interaction demand and distraction performance. Together, these analyses provided both a validation of the step metric as a stable indicator of IVIS interaction demand and a more granular predictive framework that can be used to assess distraction in future vehicles and IVIS systems. If this predictive framework proved successful, the number of interactions or “steps” could serve as a simple, practical metric of IVIS distractibility.

18 Results

18.1 Validating the Step Metric Across On-Track and Off-Track Tests

We evaluated whether on-track (Phase 1) and off-track (Phase 2) tests produced consistent estimates of IVIS interaction demand. Linear mixed-effects models of identical structure were fitted for each distraction outcome, and IVIS-level fixed-effect estimates for the dependent variable *number of interactions* were extracted. Figure 8 shows the correlations between these IVIS fixed-effect estimates: the upper panels plot the Phase 1 (on-track) estimates (x-axes) against the corresponding estimates from the VOT (top left) and Tracking Task (top right). The correlation coefficients ($r > 0.67$) indicate a substantial correspondence in IVIS interaction demand across test settings. Because the estimates are informed by the same participants interacting with the same IVISs across Phases 1 and 2, the strong correlations provide confidence that IVISs requiring more interactions on-track also require more interactions during stationary testing. Given the stability of the step metric across driving and stationary tests, we proceeded to Phase 3 modelling to assess whether the number of interactions could serve as a valid predictor of distraction-related outcomes.

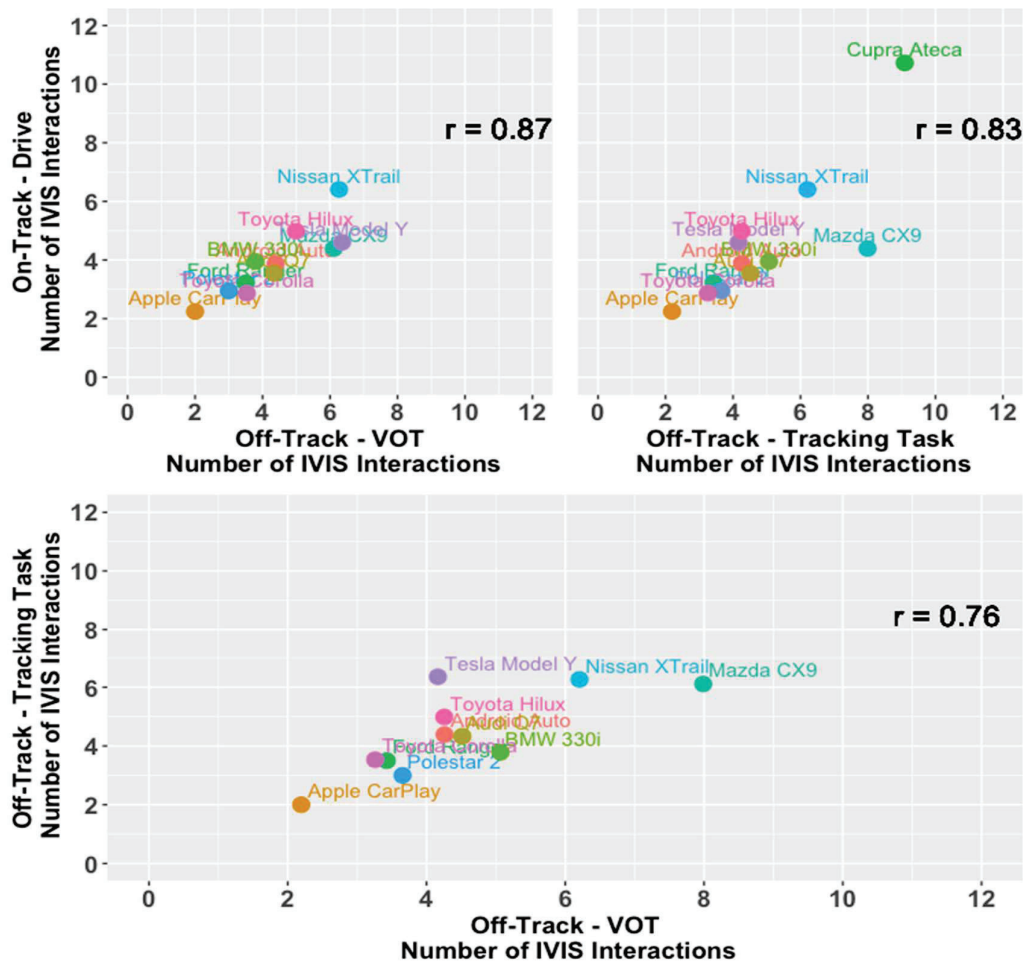


Figure 23. Consistency in IVIS interaction demand across on-track and stationary tests. Linear mixed-effects models were used to estimate the number of interactions required for each IVIS. The upper panels compare Phase 1 (on-track) estimates with those from the VOT (top left) and Tracking Task (top right). The substantial correlations ($r > 0.67$) indicate that IVISs requiring more interactions on-track also require more interactions during stationary testing. This consistency supports the use of the step-count metric in subsequent Phase 3 modelling as a predictor of distraction-related outcomes.

Note. The Cupra Ateca appears only in the Tracking Task $\times$ Drive panel because it was removed from all VOT analyses due to system failures that produced unreliable VOT interaction counts. See 13.7 Modelling Off-Track Stationary Data for details.

18.2 Evaluating Step Counts as a Predictor of Driver Distraction

In this section, we determine whether IVIS step counts reliably predict distraction-related performance measures. Essentially, we are asking if someone knows the number of steps, can they reasonably estimate how distracting a system is? To achieve this, we tested several candidate models representing different assumptions about participant variability and IVIS-specific effects. The model selection results presented below describe how we determined the most appropriate and statistically robust model for predicting distraction.

18.3 Model selection

We compared three mixed-effects models of increasing complexity, independently for the DRT and VOT datasets, to evaluate whether the number of IVIS interactions (steps) could reliably predict distraction outcomes (see Model Selection and Validation for the Step-Count Metric for full model specifications). Model fit was assessed using the Restricted Maximum Likelihood (REML) criterion, while predictive validity was determined via Leave-One-Out Cross-Validation (LOOCV). This validation technique—detailed in the following section—tests the model's ability to predict 'unseen' data, ensuring the results are generalisable to new vehicles and interfaces.

- **Model 1 (Random Intercept Only):** This model establishes a universal relationship between step count and distraction while accounting for baseline participant differences. It served as the baseline for both metrics (DRT REML = -474.3; VOT REML = 3389.1).
- **Model 2 (Interaction):** This model introduced system-specific slopes (nSteps x IVIS). For the DRT, this specification performed worse (REML = -425.1), indicating that system-specific slopes did not meaningfully improve the model. While it offered the lowest REML for the VOT (3305.2), this improvement required 11 system-dependent parameters, undermining the goal of a system-agnostic metric.
- **Model 3 (Random Intercept & Slope):** This specification allowed the "step effect" to vary across participants. For the DRT, this provided the best statistical fit (REML = -513.6). For the VOT, it provided a moderate improvement over the baseline (REML = 3366.1).

Despite the superior statistical fit of Model 3 for the DRT, **Model 1** was selected as the final specification for both outcome measures. This decision was driven by two key factors: predictive stability and metric standardization.

In the out-of-sample predictive validation for the VOT, Model 1 demonstrated the highest correlation between predicted and observed values ($r = 0.65$), outperforming both the Interaction model ($r = 0.65$) and the Random Slope model ($r = 0.63$). For the DRT, while the Random Slope model offered a higher correlation ($r = 0.75$), the performance of Model 1 remained robust ($r = 0.71$).

We have chosen Model 1 as the standard model to ensure a parsimonious and computationally stable framework. This approach prioritizes a consistent, system-agnostic

relationship between interaction steps and distraction that is easily interpretable for consumer safety ratings across both cognitive and visual modalities.

18.4 Model prediction

We examined the relationship between observed distraction outcomes and the model's out-of-sample predictions to evaluate the predictive performance of the final mixed-effects model (Model 1). As noted previously, these predictions were generated via Leave-One-Out Cross-Validation (LOOCV). This process estimates how a model performs when applied to a new, "unseen" IVIS by iteratively training the model on all but one data point and then testing its accuracy on the omitted observation.

The LOOCV scatterplots for the DRT (Figure 20), VOT (Figure 21 – Top), and Tracking Task (Figure 21 – Bottom) outcomes illustrate the model's predictive performance. Across all three tasks, a clear positive trend is evident, and the model predictions provide a strong approximation of the observed performance:

- **For the DRT (Figure 20):** Predicted miss rates align closely with the observed proportion of misses ($r = 0.71$), indicating that step count captures reliable variation in participants' likelihood of missing a DRT signal.
- **For the VOT (Figure 21 – Top):** Observed TSOT values and step-count model predictions are moderately-to-strongly correlated ($r = 0.65$), confirming that step count reflects broad performance trends in visual attention.
- **For the Tracking Task (Figure 21 – Bottom):** Observed and predicted tracking error values are strongly correlated ($r = 0.81$), indicating that interaction step count provides a highly reliable predictor of manual tracking performance.

These consistent predictive patterns across three independent distraction measures confirm that the number of steps required to complete an IVIS task provides a stable, generalisable, and practical indicator of distractibility across both cognitive and visual modalities.

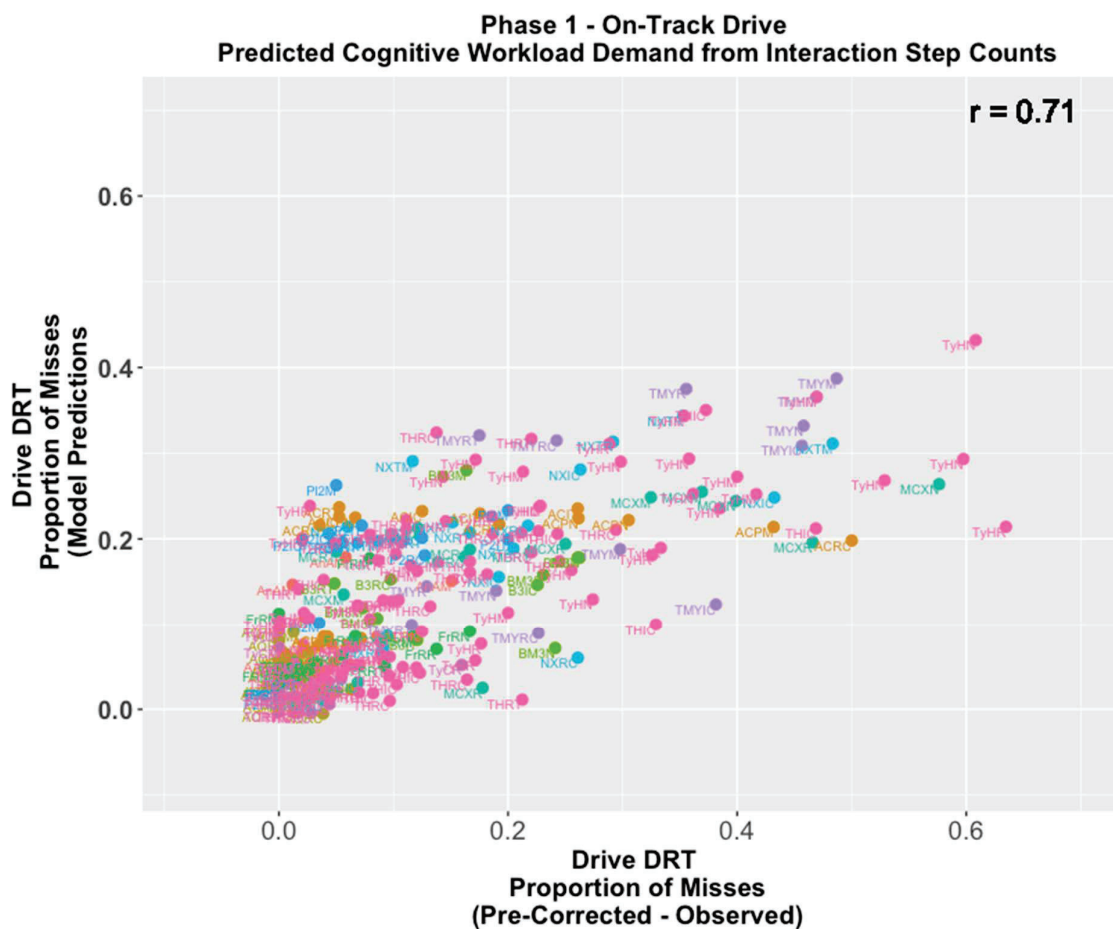


Figure 24. Interaction step counts were used to predict Drive DRT performance. The strong correlation between predicted and observed Drive DRT ($r = 0.71$) indicates that this simple metric provides a valid proxy for assessing IVIS-related distraction.

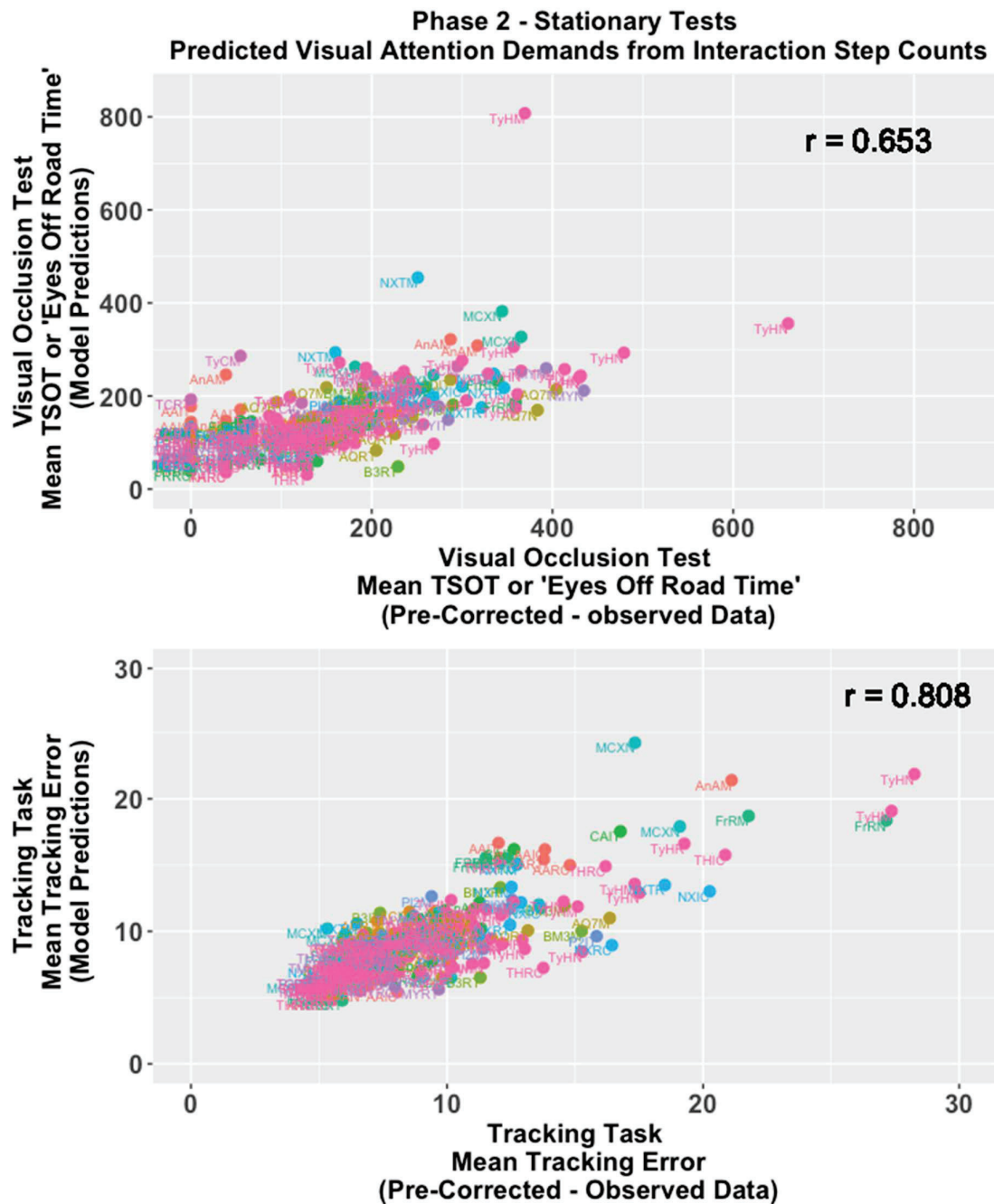


Figure 21. Interaction step counts were used to predict VOT TSOT (Top panel) and Tracking Task Tracking Error (Bottom panel) performance. The strong correlation between predicted and observed VOT ($r = 0.65$) and predicted and observed Tracking Task Tracking Error ($r = 0.81$) indicates that the simple step metric provides a valid proxy for assessing IVIS-related distraction.

19 Discussion – Phase 3

Phase 3 results show that the number of IVIS interactions, or “steps,” required to complete a task is a powerful predictor of driver distraction. Using data from on-track driving tests and stationary tasks, we fitted mixed-effects models that accounted for differences between participants and IVIS systems. The findings consistently showed that as the number of required interactions increases, so does the level of distraction, as measured by DRT misses, VOT TSOT, and Tracking Task tracking error. Crucially, this relationship held even when the model was evaluated under strict model validation conditions. This provides strong evidence that step count is a robust and generalisable indicator of IVIS distractibility.

A major advantage of the step metric is its ease and practicality. Steps can be counted directly by a user seated in a stationary vehicle without the need for specialised testing equipment or on-road procedures. This makes step count the simplest validated metric arising from the study, one that can be used by consumers, vehicle evaluators, manufacturers, and safety organisations alike. As illustrated in Figure 26, while on-track testing remains the benchmark for fidelity, the Step Count metric provides a unique utility by offering a highly scalable, low-resource alternative that maintains a validated link to driver behaviour.

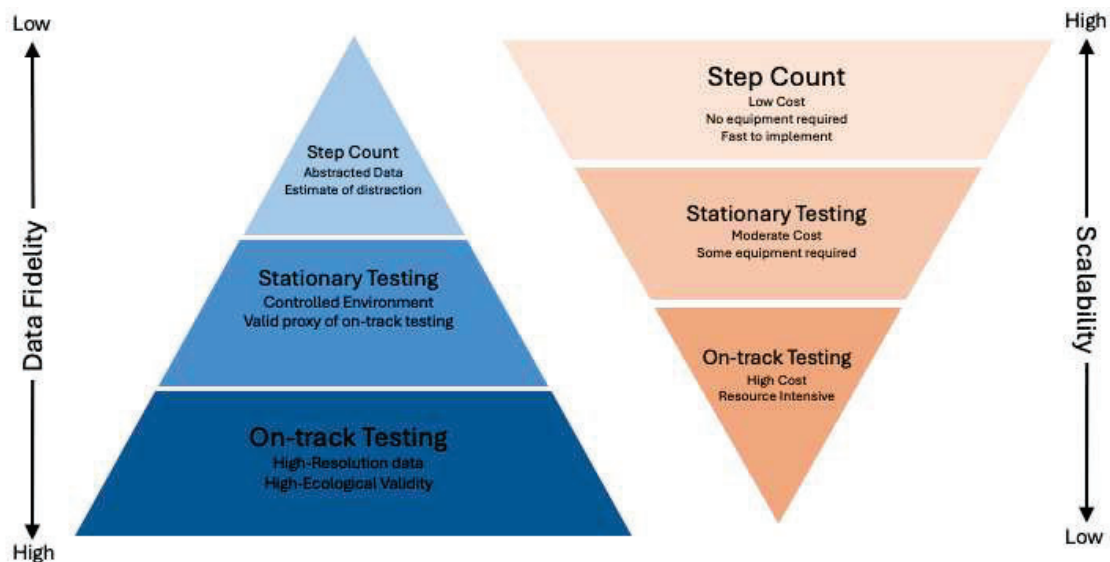


Figure 25. Conceptual trade-off between data fidelity and scalability across assessment modalities. The left pyramid illustrates data fidelity, increasing from step count (low fidelity) to stationary testing (moderate fidelity) and on-track testing (high ecological validity and resolution). The right inverted pyramid illustrates scalability and the pragmatic value to most stakeholders, with step count being most scalable (low cost, no equipment), stationary testing requiring moderate resources, and on-track testing being least scalable due to higher cost and resource intensity.

Phase 3 also reinforced the link between our stationary assessments and on-track performance. By modelling both data sources jointly, we demonstrated that stationary measures capture the same underlying behavioural patterns observed in driving. This strengthens the case for using stationary testing as a credible, lower-risk alternative to on-track assessment, enabling broader and more efficient evaluation of IVIS systems.

Overall, these outcomes provide strong support for our approach and establish the number of IVIS interactions as a practical, future-proof measure of distractibility. In the following section, we define benchmarks that distinguish between levels of IVIS distractibility and outline a safety rating scale framework that could be integrated into an ANCAP-style safety rating system.

20 Quantifying Interaction-Driven Distraction Costs to Inform an IVIS Safety Rating System

We have established that the number of IVIS interactions (“steps”) provides a robust, cross-modal predictor of distraction that discriminates between IVIS based on their level of distractibility. Next, we determine how much each additional interaction with an IVIS ‘costs’ in terms of distraction-related performance. This requires quantifying the functional relationship between step count and both visual attention (TSOT) and cognitive workload (DRT): for example, estimating how much TSOT increases per step, how quickly this growth accelerates for complex tasks, and how DRT performance deteriorates as tasks require more interactions. Using these model-derived cost functions, we have established threshold regions that categorise IVIS tasks and systems within a 5-star safety rating scheme, conceptually similar to ANCAP’s approach. These regions define the step-based boundaries of the scale ensuring that tasks with very few interactions map to high ratings, whereas tasks with many interactions, and therefore high distraction costs, map to lower ratings. This process results in a transparent, data-driven, and safety-oriented mapping from step count to star rating that is grounded in both visual and cognitive distraction outcomes.

20.1 Estimating the distraction cost of IVIS interactions

To calculate the cost of each additional IVIS interaction for the DRT, we use the fixed-effect slope from the mixed-effects model that predicts DRT performance from the number of IVIS interactions. This slope reflects the average increase in distraction for every extra interaction. In the DRT model, the slope is 0.011, meaning that each additional step increases the probability of missing a DRT stimulus by approximately 1 percentage point. This value represents the “cognitive cost” of a step i.e., how much more difficult it becomes for drivers to detect and respond to events as an IVIS task requires more interactions.

For the VOT, the model was fitted to the square root of TSOT, meaning that each additional step increases the predicted square-root TSOT by a fixed amount. To interpret these effects in meaningful units (milliseconds of eyes-off-road time), the predictions must be back-transformed. This back-transformation produces a non-linear increase in visual demand; although the model adds a constant amount on the transformed scale, the resulting TSOT values accelerate on the original scale. As a result, each extra step costs more eyes-off-road time than the previous step. For an average IVIS task, an additional interaction “costs” approximately 270 ms to 800 ms of eyes-off-road time, depending on where it falls within the

task sequence. This means the 4th step adds more visual demand than the 3rd, the 5th more than the 4th, and so on. This pattern reflects a core principle of task complexity; as an IVIS task requires more interactions, the driver must manage additional attentional and interactional demands, making later steps disproportionately demanding. Therefore, complex tasks do not simply take longer, they impose increasingly higher visual-attention costs as the number of required interactions increases. See Appendix 5 – The cost of IVIS interactions: step count Predictions of distraction outcomes for a full breakdown of these predictions.

Together, these estimates demonstrate that each IVIS interaction carries a quantifiable distraction cost, cognitive for the DRT and visual for the VOT, and that these costs accumulate as tasks require more steps. Although the two measures differ in scale and form, both show a consistent pattern: additional interactions reliably degrade drivers' capacity to maintain situational awareness. Importantly, these step-based costs provide a unified framework for comparing IVISs with very different interface designs. By focusing on TSOT and DRT rather than abstract tracking error scores, we convert diverse behavioural outcomes into a common currency of interaction burden, eyes-off-road time and missed signals, that is directly interpretable by stakeholders. This ensures the 5-star rating is anchored in tangible safety risks that are intuitively understood. By grounding the assessment in model-derived estimates rather than subjective judgement, the step metric offers a transparent, evidence-based foundation for evaluating task complexity and supports its use as the core input to an IVIS 5-star distractibility rating system.

20.2 Deriving IVIS Interaction Thresholds and Their Mapping to IVIS Star Ratings

We derived a star-rating framework that parallels the categorical structure used by ANCAP and Euro NCAP, and it is anchored in the empirically modelled relationship between IVIS task complexity (steps) and distraction (See Table 3 below). Our mixed-effects model analyses showed that each IVIS interaction produces a predictable, monotonic increase in visual demand (TSOT) and cognitive workload (DRT), providing a statistically defined “cost per step” that applies across vehicles, tasks, and participants. On this foundation, three considerations motivated the choice of “two-step” bands:

1. *Statistical meaning of a step*: the fixed-effect slopes quantify the average change in distraction per additional interaction, and grouping systems by step counts preserves the modelled ordering of distraction burden, without forcing large-scale distraction outcomes (such as TSOT in milliseconds or DRT miss

rates) to be compressed into discrete star categories; a process that would require additional, arbitrary scaling decisions.

2. *Empirical clustering*: the mean step counts of real-world IVIS tasks fall within narrow integer ranges (approximately 2–10 steps), and a two-step interval corresponds to a behaviourally meaningful difference in predicted distraction; large enough to matter, but fine-grained enough to discriminate between systems.
3. *Alignment with observed limits*: mapping these two-step bands to a 0–5 star structure captures the full empirical range, with 1–2 steps corresponding to very low modelled distraction (5 stars) and ≥ 11 steps corresponding to extremely high distraction (0 stars). This structure ensures that each star reflects a statistically coherent change in eyes-off-road time and workload and that the thresholds are transparent, repeatable, and directly tied to the modelled cost of an interaction. Therefore, while OEMs or system designers need only count interactions (with the assessment procedure detailed in Section 19.3 Procedure for Assessing IVIS Distractibility Using the Step Metric), the underlying step bands are calibrated to the empirically derived relationship between task complexity and distraction, providing a technically defensible foundation for the rating system.

Stars	Number of IVIS Interactions Required	Description
★★★★★	1–2 steps	Very low driver distraction
★★★★☆	3–4 steps	Low distraction; efficient interface
★★★☆☆	5–6 steps	Moderate distraction; acceptable but improvable
★★☆☆☆	7–8 steps	High distraction; redesign recommended
★☆☆☆☆	9–10 steps	Very high distraction; unsafe for driving
☆☆☆☆☆ (0 stars)	≥ 11 steps (or system fails usability/safety criteria)	Extreme distraction; unacceptable for use while driving

Table 3. Step-based IVIS star-rating thresholds. Each star category corresponds to a two-step interval in the mean number of interactions required to complete IVIS tasks. These thresholds are anchored in the modelled relationship between task complexity and distraction, with low-step systems rated as low distraction (5 stars) and high-step or failing systems classified as 0-star.

To demonstrate both the practicality and the statistical validity of our star-rating system, we applied the classification method across all IVISs in this study and summarised the results in Table 4. The table presents the model-estimated VOT and DRT step counts for

each system and shows how these values are combined to determine the final star rating. For each IVIS, the modelled VOT and DRT step estimates were averaged to produce a **Mean Combined Step** count, which was then rounded to the nearest whole number before applying the predefined step-band thresholds. This approach ensures that the final rating is driven directly by the modelled task complexity and avoids the need to generate separate VOT and DRT star scores. The rounded mean provides a stable and representative index of overall interaction demand because VOT and DRT step estimates were generally consistent across systems. The resulting star ratings range from 5 stars (low distraction, minimal interactions) to 1 star (high distraction), with a zero-star category reserved for systems whose step demands exceed acceptable thresholds. Across all IVISs, systems requiring fewer modelled interactions achieved higher star ratings, while more complex systems were downgraded due to their greater visual and cognitive demands. The final star rating therefore provides a transparent, statistically grounded, and easily interpretable summary of overall IVIS usability and driver-distraction potential.

IVIS	VOT Steps (Model)	DRT Steps (Model)	Mean Combined Steps (Model)	Rounded Mean Step Count	IVIS Safety Rating
Apple CarPlay	2.00	2.24	2.12	2	★★★★★
Polestar 2	3.00	2.95	2.97	3	★★★★☆
Ford Ranger	3.50	3.22	3.36	3	★★★★☆
Toyota Corolla	3.53	2.87	3.20	3	★★★★☆
BMW 330i	3.79	3.95	3.87	4	★★★★☆
Audi Q7	4.33	3.55	3.94	4	★★★★☆
Android Auto	4.39	3.88	4.14	4	★★★★☆
Toyota Hilux	4.99	4.98	4.98	5	★★★★☆
Mazda CX9	6.11	4.39	5.25	5	★★★★☆
Tesla Model Y	6.36	4.60	5.48	5	★★★★☆
Nissan X-Trail	6.27	6.41	6.34	6	★★★★☆
Cupra Ateca*	-	10.72*	11*	11*	☆☆☆☆☆*

Table 4. Model-estimated IVIS interaction demands and resulting star ratings. VOT Steps and DRT Steps indicate the mean number of interactions (“steps”) required to complete tasks in the Visual Occlusion Test and Drive DRT tests. For each IVIS, these modelled values were averaged to produce a **Mean Combined Steps** count, which was then rounded to the nearest whole number and mapped onto the predefined step-band thresholds shown in Table 3 to determine the **IVIS Safety Rating**. This single rating represents the final classification of each IVIS. The star-rating system is derived directly from the modelled relationship between task complexity and distraction. To align with the step-based threshold definitions, rounding is applied only after averaging the modelled VOT and DRT step estimates (with .5 values rounded upward). This procedure provides a stable and statistically coherent basis for assigning the final star rating.

***Note.** The Cupra Ateca is included in the rating table to illustrate how the step-based star-rating framework classifies a clearly poor IVIS. Its rating is based solely on Drive DRT step counts, as system failures rendered its VOT data unusable for modelling. The resulting rating is therefore indicative rather than fully model-validated but remains consistent with the system’s observed interaction complexity.

Using the model-predicted TSOT values, we linked each star-rating category to concrete estimates of eyes-off-road time and then applied these thresholds across all IVISs in our study. Five-star systems, represented only by Apple CarPlay, required 1–2 steps, corresponding to approximately 0.93–1.25 seconds of eyes-off-road time per task. These systems impose very low visual demand. Four-star systems, including the Polestar 2, Ford Ranger, Toyota Corolla, BMW 330i, Audi Q7, and Android Auto, required 3–4 steps, producing around 1.62–2.03 seconds of eyes-off-road time per task, still within a comparatively low-risk range. A number of systems fell into the three-star category, such as the Toyota Hilux, Mazda CX-9, Tesla Model Y, and Nissan X-Trail, requiring 5–6 steps (and in some cases slightly higher), which translates to 2.50–3.01 seconds of eyes-off-road time per task. These systems show a clear increase in visual-attention demand compared with the four- and five-star performers. In contrast, the Cupra Ateca fell squarely into the zero-star category, with 11 steps, resulting in 4.83–5.53 seconds of eyes-off-road time per task. Systems exceeding 11 steps would constitute a zero-star rating, reflecting more than 6 seconds of eyes-off-road time per task, although no system in this dataset reached this extreme. Together, these values demonstrate how the star-rating system provides an intuitive, metric-based interpretation of IVIS distractibility: small increases in interaction frequency produce meaningful and progressively larger increases in the visual-attention cost experienced by drivers.

20.3 Procedure for Assessing IVIS Distractibility Using the Step Metric

The step metric provides a simple, practical, and statistically validated method for assessing the distraction potential of IVIS tasks. Following establishment of the relationship between steps and distraction through modelling (Section 17.1), an IVIS can be evaluated by counting the number of interactions required across all task types, calculating the mean step count for the system as a whole, and comparing this value to the star-rating thresholds in Table 3. Step-based IVIS star-rating thresholds. Each star category corresponds to a two-step interval in the mean number of interactions required to complete IVIS tasks. These thresholds are anchored in the modelled relationship between task complexity and distraction, with low-step systems rated as low distraction (5 stars) and high-step or failing systems classified as 0-star.. The following procedure outlines the steps required to conduct a step-based distractibility assessment.

20.3.1 Identify the IVIS task

The assessor first identifies the full set of IVIS task types to be evaluated for a system.

Typical examples include:

- Initiating a phone call
- Entering a navigation destination
- Selecting or changing a media item
- Sending a text message
- Adjusting climate

To ensure comparability across repetitions and across different IVISs, each task type requires a clearly defined and consistently applied start point and end point, so that step counts reflect the same portion of the interaction process every time. Table 5 provides suggested start and end points for some example task type to support consistent task definition across assessors and IVISs.

Task Type	Start Point	End Point
Initiate Call	When the driver first issues the voice command or presses the call-control UI	When the call begins ringing or call screen opens
Voice Command Call	When the wake word is spoken (“Hey BMW...”)	When the system confirms the call target (e.g., “Calling Alex...”)
Navigation Entry	When the driver first interacts with the navigation interface	When the destination is accepted or route guidance begins
Media Selection	When the driver first touches the media interface or issues a media voice command	When the selected track, station, or playlist begins playback
Change Source/Mode	When the driver initiates the change (e.g., touches “Audio Source”)	When the new source is fully selected and begins outputting audio
Adjust Settings	When the driver first interacts with the relevant settings UI	When the chosen setting change (e.g., volume, temperature) is enacted

Table 5. Examples of defined start and end points for IVIS task types to ensure standardised step-count measurement across trials and systems.

20.3.2 Perform multiple repetitions of each task type

To obtain stable estimates, assessors should complete **approximately 10 trials of each task type**. This accounts for variability in system response, user repetition, and occasional anomalies.

Step-counting accuracy is greatly improved when supported by:

- Video recording (recommended – GoPro or Smart phone), or
- A second observer dedicated to logging interactions

A complete step-count assessment for one IVIS (covering all task types) typically requires 45–50 minutes, based on Phase 2 stationary testing.

20.3.3 Count all driver–IVIS interactions

A step, as outlined in Section *Defining the number of IVIS interactions (steps)*, is any **driver-initiated** input, including touchscreen taps, physical button presses, rotary-controller movements, steering-wheel controls, and voice interactions (wake words, commands, confirmations).

- Do **not** count passive waiting (screen transitions, loading), unless the driver must perform an additional action.
- **Repeated or re-issued commands:** for example, when the system misrecognises or fails to respond, a step **must be counted**. Step counts must reflect **what the driver actually does**, not designer-intended workflows.

This ensures the final rating captures real-world interaction demands, including variability driven by speech recognition accuracy, interface responsiveness, or unexpected menu behaviour.

20.3.4 Document any system anomalies

IVIS performance can vary across vehicles and conditions therefore assessors should note any behaviours that change the number of interactions required. Examples include:

- Slow or lagging screen transitions
- Touchscreen inputs not registering
- Unexpected menu states
- Required confirmations not described in the intended workflow
- Poor speech recognition (a notable anomaly in our testing, where repeated voice-command attempts increased the actual number of steps required)

These issues directly affect the number of interactions a driver must make and therefore influence the overall safety rating.

20.3.5 Calculate the mean step count for each task type

For each task type:

- Average the number of steps across the ~10 trials
- This yields the *task-level* mean step count

These means reflect the observed task complexity independent of outliers or single-trial issues.

20.3.6 Calculate the overall mean step count for the IVIS

The overall IVIS distractibility score is not based on any single task but on the combined behaviour of all task types within the system. To generate this system-level value, calculate the mean of the task-level mean step counts across all IVIS task types. This produces the Mean Combined Step Count, representing the typical interaction burden imposed by the system across its core functions. This method prevents any unusually easy or difficult task from disproportionately influencing the result and ensures that the final rating reflects the IVIS's overall usability profile.

20.3.7 Determine the IVIS star rating

To convert the system-level mean into a rating:

1. Round the Mean Combined Step Count to the nearest whole number (.5 rounds up).
2. Assign the star rating based on the step-band thresholds:
 - **1–2 steps** → ★★★★★★
 - **3–4 steps** → ★★★★★☆
 - **5–6 steps** → ★★★★★☆
 - **7–8 steps** → ★★★★★☆
 - **9–10 steps** → ★★★★★☆
 - **≥ 11 steps** → ★★★★★☆ (unacceptable)

This rounded system-level mean step count yields the final, consumer-facing IVIS distractibility rating.

21 Conclusion, recommendations, and future steps

This project delivered a comprehensive, multi-method evaluation framework for assessing the distractibility of in-vehicle infotainment systems (IVISs) in a selection of vehicles currently available in the Australian market. By combining on-track testing, stationary vehicle testing, and a step-based interaction metric, we assessed IVIS performance across environments ranging from controlled, repeatable settings to real-world driving conditions. The on-track protocol provided high-validity measurements of distraction under realistic driving demands and offers insights valuable to automotive manufacturers, ANCAP and Euro NCAP, and human factors researchers; while resource-intensive, its findings can directly inform vehicle design and future safety assessments. The stationary vehicle methods offered a more cost-effective alternative that retained strong predictive value and can be adopted by research organisations to evaluate emerging interfaces. The step-based interaction metric, requiring no specialist equipment or technical expertise, provides a scalable, accessible tool that fleet managers and individual drivers can use to gauge IVIS complexity based solely on interaction demand. Most importantly, the distraction measures derived from these methods enable the development of an ANCAP-style IVIS distractibility rating, giving consumers a clear and evidence-based way to understand and compare the safety of IVISs in vehicles they may consider purchasing. Together, these complementary methods establish a robust, future-proof framework for evaluating the distractibility of both current and next-generation IVIS technologies, supporting safer vehicle design, regulatory decision-making, and informed consumer choice.

21.1 Applications of Key Findings

The findings of this project support several key recommendations to guide the development of an evidence-based IVIS distractibility rating for the Australian market. First, we recommend the multi-method assessment framework established in this research, using stationary vehicle tests and the step-based interaction metric as practical and scalable tools for regular IVIS assessment, while reserving on-track testing for validation activities and the evaluation of priority systems. The step metric is transparent and easily interpretable and could be incorporated into consumer-facing communication to enable purchasers to compare IVISs based on distraction potential.

The project's modelling outputs also provide a strong foundation for developing an ANCAP-style 5-star IVIS distractibility rating, aligned with existing safety-rating processes and consistent with international methodological directions. Manufacturers should be

encouraged to apply these tools during early IVIS design to identify distraction risks and support safer interface development.

While the proposed assessment methods emerging from this project are applicable to a contemporary matrix of interaction modes, future interface technologies may introduce novel effects on driver attention. For this reason, the framework will need to remain adaptable, with the capacity to incorporate new IVIS modalities and, when necessary, return to on-track validation to ensure continued alignment between interaction metrics and real-world safety outcomes. This need for adaptability underpins the importance of considering how the protocol may evolve over time, as outlined in the next section.

21.2 Future Adaptability of the IVIS Distraction Assessment Protocol

The test protocol we have validated could be used in future applications, structured so that each phase helps address a different aspect of how IVIS interactions affect driver distraction. On-track driving tests could remain valuable because they allow assessors to periodically check that interaction-based metrics, particularly step counts, still correspond to meaningful safety outcomes as IVIS technologies evolve. As new modalities emerge (e.g., AI-driven voice systems, augmented-reality displays, haptic surfaces), these real-world evaluations may help ensure that any rating approach stays anchored to observable driving performance rather than drifting toward purely theoretical indicators.

Stationary tests may also retain an important role by offering diagnostic insight into why certain systems impose visual or cognitive load. Their controlled conditions make them well suited to exploring emerging interface designs, identifying subtle usability issues, and examining whether the mechanisms linking interaction complexity to distraction remain consistent across new technologies and interface styles.

The step-count metric could continue to function as a practical consumer-facing measure because it is simple, scalable, and comparable across vehicles, model years, and interface types. It may also offer an efficient way for organisations such as ANCAP to assess a wide range of systems, including those updated via software, while giving consumers and fleet managers a straightforward indicator of usability-related distraction risk.

Taken together, these phases could provide an adaptable assessment framework: on-track tests helping to maintain real-world relevance, stationary tests supporting rapid evaluation and deeper scientific understanding, and step counts offering an accessible, system-wide metric for consumers and industry. These possibilities suggest that the overall approach has the flexibility to remain credible, future-proof, and broadly useful without prescribing how assessments must be conducted.

21.3 Requirements for model re-validation

While the assessment framework offers flexibility in how each phase could be used in future applications, it is equally important to recognise the limitations of relying solely on the lower-cost approaches developed in Phases 2 and 3. These methods provide efficient and scalable indicators of visual and cognitive demand, but they remain indirect measures of the safety implications of IVIS interactions. Phase 1 on-track testing continues to offer the highest-fidelity evidence because it captures how interaction burden translates into observable driving performance under realistic conditions.

For this reason, occasional high-fidelity validation may be required when new IVIS technologies enter the market, particularly those that differ substantially from the systems included in the original model-development sample. For example, the increasing prevalence of audio-only, multimodal AI-assistant, or voice-first interfaces, technologies not widely represented during earlier field studies (Strayer et al., 2017), could introduce patterns of demand that fall outside the assumptions embedded in the current statistical models and star-rating thresholds. As IVIS ecosystems evolve, the relationship between interaction complexity, cognitive load, and safety outcomes may shift in ways that stationary tests or step-count metrics alone cannot capture. Accordingly, periodic on-road benchmarking could help detect any such “model drift” and ensure that the rating framework remains aligned with real-world driver behaviour. This approach preserves the efficiency benefits of Phases 2 and 3 while maintaining the scientific and safety credibility afforded by high-fidelity validation.

21.4 Implications of Task Type Consistency across IVISs for the Star-Rating Procedure

Variability in the tasks supported by different IVISs is an inherent aspect of current vehicle design. To accommodate this diversity in a fair and transparent way, we outline a suggested rationale for defining core tasks based on interaction domains rather than dataset availability. This approach is described in Appendix 13 - Rationale for Establishing Core Task Domains. Core tasks were defined according to the major functional domains of IVIS interaction, phone operations, messaging, media control, and navigation, rather than by the specific tasks available in the current dataset. These domains reflect the types of interactions most performed by drivers and those known to impose visual and cognitive load (Strayer et al., 2017; Strayer et al., 2019). Individual tasks were selected within each domain to represent a typical interaction sequence (e.g., initiating a call, receiving a text, selecting media, or entering a navigation destination). Although some vehicles in the study did not support every task, the definition of core tasks remains anchored in these functional domains to ensure the star-rating system is meaningful, stable across vehicles, and aligned with real-world IVIS usage. Task

availability in the sample influenced how missing domains were handled, but it did not determine which tasks were considered core.

21.5 Step-Count Temporal Validity

It is important to recognise that the step-count metric is calibrated on the temporal characteristics of the IVIS tasks examined in this study. The relationship between interaction frequency and distraction outcomes was estimated using brief, continuous IVIS interactions completed within a single driving episode where participants were instructed to complete tasks as quickly and accurately as possible. Consequently, our step-cost estimates are most valid when applied to tasks of comparable duration and structure, and where there is an incentive for fast task completion. For tasks that extend over longer periods, contain intermittent pauses, or unfold across multiple driving epochs, the assumed linear relationship between step count and distraction may not hold, as drivers may self-pace or interleave interactions with driving demands in ways that were not captured in the present modelling.

Accordingly, the step-count metric should be interpreted as an index of interaction complexity (i.e., more interactions, more complexity) for short-duration IVIS tasks rather than a general measure of distraction risk across an exhaustive range of scenarios. When evaluating longer or multi-stage interactions, assessors may need to decompose the task into shorter, functionally coherent subcomponents or treat the resulting step counts as indicative rather than definitive. Future research is required to develop calibrated step-cost functions for extended or self-paced interactions and to examine how temporal structure influences the relationship between interaction demand and driver performance.

21.6 Ensuring future impact

To build on the foundations established in this study, several avenues for future development could be considered. The step-based thresholds used to define the 5-star rating system may benefit from ongoing refinement as new IVIS designs and interaction modalities become available, ensuring that rating boundaries remain aligned with contemporary in-vehicle technologies. Expanding the catalogue of assessed IVIS tasks, particularly those involving emerging or next-generation HMI concepts, would help maintain the relevance and applicability of the framework as interfaces continue to evolve.

Strengthening international alignment with organisations such as ANCAP, Euro NCAP and IIHS could support methodological harmonisation and encourage movement toward a more globally consistent approach to IVIS safety assessment. Euro NCAP, for example, is developing protocols for the safe use of “general controls” and has announced 2026 rating

updates that reward vehicles retaining physical controls for key driving functions as a way of managing distraction by design (Euro NCAP, 2023, 2024). In parallel, IIHS is increasingly focusing on technology-related distraction and driver attention, including research on smartphone- and voice-related distraction and planned *Top Safety Pick+* criteria that incorporate driver-attention and impairment-detection capabilities (IIHS, 2021, 2022, 2024). Positioning the present framework in relation to these international developments would enhance its credibility and potential for uptake.

Recent developments in production vehicles indicate that IVIS design is continuing to diversify in ways that reinforce the need for an adaptable and evidence-based assessment framework. New vehicles contain increasingly expanding touchscreen functionality such as larger display formats and the migration of additional controls, including climate and window functions, into digital menus, while others are shifting toward more integrated ecosystems that blend native IVIS platforms with external systems such as Apple CarPlay and Android Auto. Even within a single manufacturer, markedly different interface philosophies are emerging across model lines (e.g., varying IVIS architectures within Mazda's new 6e and CX-series vehicles). These trends highlight the pace and variability of interface design across the industry and underscore the importance of maintaining a rating protocol that can accommodate different interaction strategies, evolving feature sets, and multiple system configurations.

As in-vehicle monitoring technologies continue to mature, this framework may also help guide the integration of Driver Monitoring Systems (DMS) and emerging Occupant Monitoring Systems (OMS). To date, the applications described in this report focus on evaluating IVIS design using research-grade behavioural and physiological measures. However, production vehicles are rapidly adopting DMS for distraction and drowsiness detection, and OMS technologies, capable of monitoring cognitive and attentional states across all vehicle occupants, are beginning to enter the market (e.g., optical and deep-learning-based systems; Bergasa et al., 2024). As these systems become more sophisticated, and as Euro NCAP moves toward requiring cognitive distraction detection in the 2029–2032 timeframe, there will be increasing opportunities for the present framework to inform real-time vehicle behaviour. Specifically, a validated, evidence-based definition of interaction-related cognitive load could be used to shape how DMS/OMS systems modulate IVIS functionality dynamically (e.g., restricting high-demand interactions, escalating alerts, or adapting interface complexity when load exceeds safe thresholds). This represents a natural extension of the framework from evaluating design to supporting adaptive safety functions in next-generation vehicles.

Together, the methods, analyses, and outcomes of this project establish a rigorous, future-proof foundation for assessing the distraction potential of in-vehicle infotainment

systems across the Australian market. By integrating scientifically validated distraction measures with practical, scalable assessment tools, we provide a pathway toward an ANCAP-style IVIS distractibility rating that can inform consumers, guide industry practice, and support regulatory decision-making. As vehicle technologies continue to evolve, the framework developed here ensures that Australia is positioned to evaluate both current and emerging HMIs in a consistent, transparent, and safety-oriented manner. This project lays the groundwork for a national approach to in-vehicle safety – one that empowers drivers, encourages safer design, and contributes to reducing distraction-related risk on Australian roads.

22 References

- Australian Institute of Health and Welfare. (2023). *Australia's hospitals at a glance 2023*. AIHW. <https://www.aihw.gov.au/>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). *Fitting linear mixed-effects models using lme4*. *Journal of Statistical Software*, 67(1), 1–48.
- Beanland, V., Fitzharris, M., Young, K. L., & Lenné, M. G. (2013). Driver inattention and driver distraction in serious casualty crashes: Data from the Australian National Crash In-depth Study. *Accident Analysis & Prevention*, 54, 99-107.
- Bergasa, L. M., Nuevo, J., Sotelo, M. A., Barea, R., & Lopez, M. E. (2006). Real-time system for monitoring driver vigilance. *IEEE Transactions on intelligent transportation systems*, 7(1), 63-77.
- Bureau of Infrastructure and Transport Research Economics. (2023). *Road deaths Australia: Monthly bulletin 2023*. Australian Government. <https://www.bitre.gov.au/>
- Choudhary, P., & Velaga, N. R. (2017). Mobile phone use during driving: Effects on speed and effectiveness of driver compensatory behaviour. *Accident Analysis & Prevention*, 106, 370-378.
- Crundall, E., Large, D. R., & Burnett, G. (2016). A driving simulator study to explore the effects of text size on the visual demand of in-vehicle displays. *Displays*, 43, 23-29.
- Dingus, T. A., Guo, F., Lee, S., Antin, J. F., Perez, M., Buchanan-King, M., & Hankey, J. (2016). Driver crash risk factors and prevalence evaluation using naturalistic driving data. *Proceedings of the National Academy of Sciences*, 113(10), 2636-2641.
- Euro NCAP. (2023). *Future protocols and roadmap 2026: Safe driving and general controls*. European New Car Assessment Programme. <https://www.euroncap.com/en/for-engineers/protocols/2026-protocols/>
- Euro NCAP. (2024). *2026 rating update: Safe use of general controls*. European New Car Assessment Programme. <https://www.euroncap.com>
- Imberger, K., Poulter, C., Regan, M. A., Cunningham, M. L., & Paine, M. (2020). Considerations for the development of a driver distraction safety rating system for new vehicles. *Journal of road safety*, 31(2), 23-34.
- Insurance Institute for Highway Safety. (2021). *Effects of smartphone-related distraction on driving performance*. IIHS. <https://www.iihs.org>
- Insurance Institute for Highway Safety. (2022). *Voice assistant and in-vehicle technology distraction research*. IIHS. <https://www.iihs.org>
- Insurance Institute for Highway Safety. (2024). *Top Safety Pick+ roadmap update: Driver-attention and impairment-detection requirements*. IIHS. <https://www.iihs.org>
- International Organization for Standardization. (2016). Road vehicles. Transport information and control systems. Detection-response task (DRT) for assessing

attentional effects of cognitive load in driving. <https://www.iso.org/standard/59887.html>. (ISO Standard No. 17488:2016)

Liang, Y. & Lee, J. (2010). Combining cognitive and visual distraction: Less than the sum of its parts. *Accident Analysis and Prevention*, 42, 881-890.

McEvoy, S. P., Stevenson, M. R., & Woodward, M. (2006). The impact of driver distraction on road safety: results from a representative survey in two Australian states. *Injury prevention*, 12(4), 242-247.

Metz, B., Landau, A., & Just, M. (2014). Frequency of secondary tasks in driving—Results from naturalistic driving data. *Safety science*, 68, 195-203.

Regan, M. A., Lee, J. D., & Young, K. (2008). *Driver distraction: Theory, effects, and mitigation*. CRC press.

Strayer, D. L., Watson, J. M., & Drews, F. A. (2011). Cognitive distraction while multitasking in the automobile. In *Psychology of learning and motivation* (Vol. 54, pp. 29-58). Academic Press.

Strayer, D. L., & Johnston, W. A. (2001). Driven to distraction: Dual-task studies of simulated driving and conversing on a cellular telephone. *Psychological Science*, 12(6), 462-466.

Strayer, D.L., Cooper, J.M., Turrill, J., Coleman, J., Medeiros-Ward, N., & Biondi, F. (2013). Measuring cognitive distraction in the automobile. *AAA Foundation for Traffic Safety*.

Strayer, D. L., & Cooper, J. M. (2017). *The cognitive control hypothesis of driver distraction*. *Human Factors*, 59(1), 5–37. <https://doi.org/10.1177/0018720817690639>

Strayer, D.L., Turrill, J., Cooper, J.M., Coleman, J.R., Medeiros-Ward, N., & Biondi, F. (2015). Assessing cognitive distraction in the automobile. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 57(8), 1300–1324

Strayer, D. L., Cooper, J. M., Goethe, R. M., McCarty, M. M., Getty, D., & Biondi, F. (2017). Visual and cognitive demands of using in-vehicle infotainment systems.

Strayer, D.L., Cooper, J.M., McCarty, M.M., Getty, D.J., Wheatley, C.L., Motzkus, C.J., Goethe, R.M., Biondi, F. and Horrey, W.J., (2019). Visual and cognitive demands of carplay, android auto, and five native infotainment systems. *Human factors*, 61(8), 1371-1386.

Steinhauser, R., Lancsar, E., Bourke, S., Munira, L., Breunig, R., Gruen, R., Dobes, L., Bulfone, L., Glass, K., Gordon, C., & Cox, J. (2022). Social cost of road crashes: Report for the Bureau of Infrastructure and Transport Research Economics. The Australian National University. <https://www.bitre.gov.au/publications/2022/social-cost-road-crashes>

Thorpe, A., Innes, R., Townsend, J., Heath, R., Nesbitt, K., & Eidels, A. (2020). Assessing cross-modal interference in the detection response task. *Journal of Mathematical Psychology*, 98, 102390.

Young, K., Regan, M., & Hammer, M. (2007). Driver distraction: A review of the literature. *Distracted driving*, 2007, 379-405.

23 Appendix 1 - Example “Run Sheet”

Participant Running Sheet

Participant ID: Px01_C01_D

Date: __/__/2023

INITIATE TEXT	
Time started:	NOTES
Text my number to say you're going to be late	
Message James to let them know you can't talk right now	
Ask Sarah what she is doing on the weekend	
I need you to message Sarah and ask "What's for dinner"	
Tell Anna you are going to be late	
Message James to ask "How's the weather?"	
Send a message to your own phone saying "Hi there".	
Ask Vicky if she can pick you up later	
Send David Thorpe a message "Can't wait to see you"	
Send Michael a message saying "I am on my way"	
Time finished:	

RADIO	
Time started:	NOTES
Tune the radio to 91.1 FM	
Change the radio to 1116 AM	
Go back to radio station 94.9 FM	
Tune the radio to 774 AM	
Tune the radio to 105.1 FM	
Go back to radio station 98.5 FM	
Tune the radio to 98.5 FM	
Go back to radio station 1116 AM	
Select 94.9 FM	
Tune the radio to 94.7 FM	
Time finished:	

INITIATE CALL	
Time started:	NOTES
Enter your own number and call it	
Dial your own number	
James is waiting for your call	
Call my phone	
You need to talk to Sarah on the phone	
Give Rachael a call	
Call David Thorpe's phone	
You need to talk to David Thorpe on the phone	
Call the last person you called again	
Michael is waiting for your call	
Time finished:	

Participant Running Sheet

Participant ID: Px01_C01_D

Date: __/__/2023

NAVIGATION	
Time started:	NOTES
Fill up at the closest petrol station.	
Drive to the Tropicana Motel, Torquay.	
Use the navigation system to take us to Barwon Heads Victoria	
Go to the Torquay Visitor Information Centre	
Please take us to Pennyroyal Raspberry Farm & Cafe	
You need to pick up a friend at Avalon Airport.	
Navigate to Anglesea Bike Park	
Drive to Last One Inn, Anglesea.	
Drive to surf coast Navigate to Surf Coast Strawberry Fields	
Navigate to Jan Juc beach Victoria	
Time finished:	

MEDIA	
Time started:	NOTES
Change the song to Michael Jackson "Beat it"	
Play Yellow Submarine by The Beatles	
Play Silverchair "Straight Lines"	
Play "Rock n Roll" by Led Zeppelin	
Play "Break on Through" by The Doors	
Please play Stevie Wonder "Superstition"	
Play "Under the bridge" by The Red Hot Chilli Peppers	
Change the song to Elvis Presley "Hound Dog"	
Play Queen "Bohemian Rhapsody"	
Change the music to The Animals "House of the Rising Sun"	
Time finished:	

RECEIVE TEXT	
Time started:	NOTES
Send message: "Hello, how are you?"	
Send message: "Can you please buy some coffee?"	
Send message: "How is the weather?"	
Send message: "I am on my way"	
Send message: "What is for dinner?"	
Send message: "We need bread and milk"	
Send message: "What time should I meet you?"	
Send message: "The game starts at 7pm"	
Send message: "Great! I am already going there."	
Send message: "I have arrived."	
Time finished:	

Participant Running Sheet

Participant ID: Px01_C01_D

Date: __/__/2023

RECEIVE CALL	
Time started:	NOTES
Call car phone	
Call car phone	
Call car phone	
Call car phone	
Call car phone	
Call car phone	
Call car phone	
Call car phone	
Call car phone	
Call car phone	
Time finished:	

END OF DRIVE

24 Appendix 2 - Full set of IVIS trials

The full set of IVIS task type trials are shown below. The list includes trials created for “Initiate Call”, “Initiate Text”, “Media”, “Navigation” and “Radio”. We did not create trial lists for ‘Receive Call’ or ‘Receive Text’ because participants were required to simply respond to the IVIS (e.g., press the phone call/microphone button on the steering wheel or press a button on the IVIS to receive a text message) rather than complete a complex interaction with the IVIS.

Initiate Call	
1	Enter your own number and call it
2	Dial your own number
3	Call David Thorpe's phone
4	Call James Freeman
5	Give Tom Randell a call
6	Call the last person you called again
7	Call my phone
8	You need to talk to Sarah Tavern on the phone
9	Michael Walsh is waiting for your call
10	Call Anna Hay
11	Vicky is waiting for your call
12	Give Rachael a call
13	James Freeman is waiting for your call
14	You need to talk to David Thorpe on the phone
15	Call your own phone

Initiate Text	
1	Text my number to say you're going to be late
2	Message James Freeman to let them know you can't talk right now
3	Send a message to your own phone saying "Hi there".
4	I need you to message Vicky and say "where are you?"
5	Please message the researcher and say "how are you?"
6	Send David Thorpe a message "Can't wait to see you"
7	I need you to message Sarah Tavern and ask "What's for dinner"
8	Tell Anna Hay you are going to be late
9	Send Michael a message saying "I am on my way"
10	Message Michael Walsh to say "I will be there in 1 hour"
11	Message Rachael to say You will bring her coffee
12	Message James Freeman to ask "How's the weather?"
13	Ask Sarah Tavern what she is doing on the weekend
14	Ask Vicky if she can pick you up later
15	Please message David Thorpe and say "how are you?"

Media	
1	Change the song to Michael Jackson "Beat it"
2	Play Yellow Submarine by The Beatles
3	Play "Under the bridge" by The Red Hot Chilli Peppers
4	Change the music to the Rolling Stones "Wild Horses"
5	Change to Angel of Harlem by U2
6	Play Queen "Bohemian Rhapsody"
7	Play "Rock n Roll" by Led Zeppelin
8	Play "Break on Through" by The Doors
9	Change the music to The Animals "House of the Rising Sun"
10	Play The Who "My Generation"
11	Change the music to Aretha Franklin "Respect"
12	Please play Stevie Wonder "Superstition"
13	Play Silverchair "Straight Lines"
14	Change the song to Elvis Presley "Hound Dog"
15	Play Bruce Springsteen "Thunder Road"

Navigation	
1	Fill up at the closest petrol station.
2	Drive to the Tropicana Motel, Torquay.
3	Go to the Torquay Visitor Information Centre
4	You need to pick up a friend at Avalon Airport.
5	Drive to Last One Inn, Anglesea.
6	Navigate to Jan Juc beach Victoria
7	Navigate to the Great Ocean Road Chocolaterie
8	Find route to 2 Grossmans Rd, Torquay VIC
9	Use the navigation system to find 221 Swanston St, Geelong VIC
10	Find the fastest route to 36 Bell St, Torquay VIC
11	Use the GPS to get to Anglesea 3230 VIC
12	Navigate to the Great Ocean Road, Anglesea
13	Take us to Bells Beach VIC
14	Use the navigation system to get to the Australian National Surfing Museum in Torquay
15	Drive to Coles Torquay
16	Please take us to Bunnings Torquay
17	Drive to surf coast Navigate to Surf Coast Strawberry Fields
18	Navigate to Anglesea Bike Park
19	Please take us to Pennyroyal Raspberry Farm & Cafe
20	Use the navigation system to take us to Barwon Heads Victoria

Radio	
1	Tune the radio to 91.1 FM
2	Change the radio to 1116 AM
3	Tune the radio to 98.5 FM
4	Tune the radio to 693 AM
5	Change the radio to 93.9 FM
6	Select 94.9 FM
7	Tune the radio to 774 AM
8	Tune the radio to 105.1 FM
9	Tune the radio to 94.7 FM
10	Tune the radio to 621 AM
11	Go back to radio station 91.1FM
12	Go back to radio station 98.5 FM
13	Go back to radio station 94.9 FM
14	Go back to radio station 1116 AM
15	Go back to radio station 693 AM

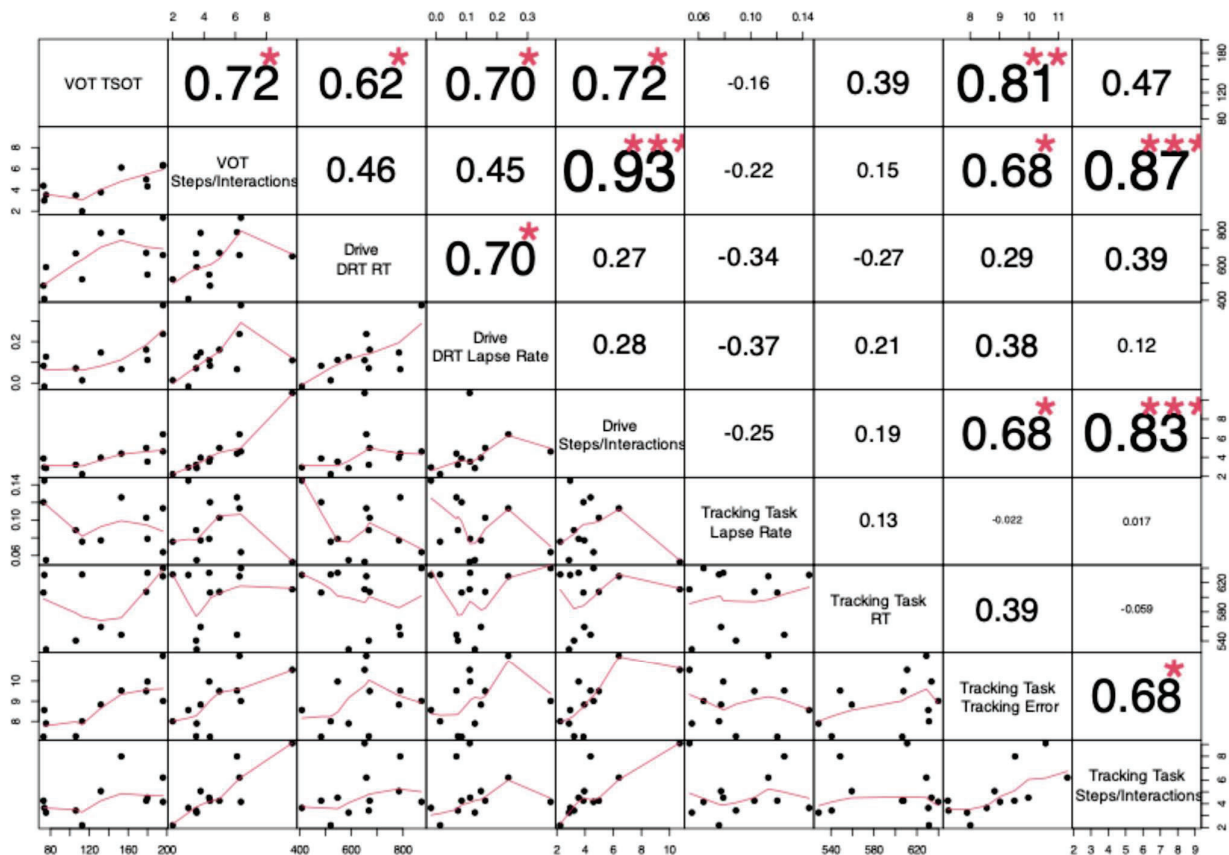
25 Appendix 3 – Analysis of “Task Order”

A linear mixed-effects model (fit using maximum likelihood) was used to evaluate whether the order in which participants completed the three on-track drives influenced DRT miss rate (pEr). The model included fixed effects for drive order, IVIS, and the IVIS $\times$ task type interaction, with a random intercept for participant. The fixed effect for “**task order**” was **not significant** (Estimate = 0.0170, SE = 0.0120, $t = 1.416$, $p = 0.158$), indicating that neither practice nor fatigue across drive blocks meaningfully affected DRT performance. All substantive IVIS effects remained unchanged when drive order was included. As a result, drive order was not retained as a predictor in the primary analyses.

26 Appendix 4 – Correlations of all Distraction Outcome Variables

The full correlation matrix examining relationships among all indirect and direct distraction measures derived from Phases 1 and 2. Variables include the VOT total eyes-off-road time (TSOT), VOT interaction count, Drive DRT response time, Drive DRT lapse rate, Drive interaction count, Tracking Task response time, Tracking Task lapse rate, Tracking Task tracking error, and Tracking Task interaction count. The upper diagonal displays Pearson correlation coefficients, where stars denote statistical significance ($p < .05$, $p < .01$, $p < .001$). The lower diagonal shows scatterplots with locally weighted smoothing lines, illustrating the form, strength, and direction of each relationship. As shown in the top rows of the matrix, IVIS interaction counts correlate strongly with distraction outcomes across multiple modalities (e.g., $r = .93$ for Drive Steps $\times$ DRT Lapse Rate; $r = .87$ for Tracking Steps $\times$ Tracking Error), demonstrating the consistency with which higher interaction demand is associated with degraded driver performance. These patterns support the use of interaction step counts as a unifying predictor within Phase 3 modelling.

Note. The Cupra Ateca does not appear in correlations involving VOT measures because it was removed from all VOT analyses due to repeated system failures that produced incomplete or invalid VOT task performance data. However, the Cupra appears in correlations for the Drive and the stationary Tracking Task, which was unaffected by these failures, allowing reliable estimates to be obtained for those tasks.



27 Appendix 5 – The cost of IVIS interactions: step count Predictions of distraction outcomes

Below illustrates how we determined the cost to cognitive workload (DRT) and visual attention (VOT) each additional IVIS interaction or “step”. The output from each model is shown below.

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: pEr ~ nSteps + (nSteps | participant)
## Data: ddrtSumPart
##
## REML criterion at convergence: -513.6
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -3.0468 -0.5133 -0.1450  0.4225  4.6284
##
## Random effects:
## Groups      Name                Variance Std.Dev. Corr
## participant (Intercept) 2.760e-03 0.052535
##              nSteps      9.811e-05 0.009905 0.88
## Residual              7.294e-03 0.085404
## Number of obs: 283, groups: participant, 19
##
## Fixed effects:
##              Estimate Std. Error      df t value Pr(>|t|)
## (Intercept)  0.091110   0.014323  18.030249   6.361 5.38e-06 ***
## nSteps       0.010858   0.002631  16.924262   4.128 0.00071 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##      (Intr)
## nSteps 0.443
```

For the drive DRT miss model, the fixed-effect intercept is **0.091** and the slope for the number of IVIS interactions (**nSteps**) is **0.011**. These values allow us to estimate the visual-attention cost associated with each additional interaction or “step.” The fitted model is:

$$pEr = 0.091 + 0.011 \cdot nSteps$$

Table 4 presents these translated values. The final column (“Increase in Eyes-off-Road time for an average IVIS task”) shows how each additional IVIS interaction increases the amount of cognitive workload imposed on the driver. Each additional IVIS interaction increases the

predicted DRT miss rate by **1.1 percentage points**, so workload rises in a strictly linear, stepwise fashion as tasks require more interactions.

Number of Interactions or “Steps”	Predicted pEr	Miss rate (%)	Δ pEr from previous step	Δ miss rate (percentage points)
0	0.091	9.11	—	—
1	0.102	10.20	+0.0109	+1.09
2	0.113	11.28	+0.0109	+1.09
3	0.124	12.37	+0.0109	+1.09
4	0.135	13.46	+0.0109	+1.09
5	0.145	14.54	+0.0109	+1.09
6	0.156	15.63	+0.0109	+1.09
7	0.167	16.71	+0.0109	+1.09
8	0.178	17.80	+0.0109	+1.09
9	0.189	18.89	+0.0109	+1.09
10	0.200	19.97	+0.0109	+1.09
11	0.211	21.06	+0.0109	+1.09
12	0.221	22.14	+0.0109	+1.09

Table 6. Predicted DRT miss rate (last column) by number of IVIS interactions (steps). The mixed-effects model shows a strictly linear increase in cognitive workload, with each additional step increasing the predicted DRT miss rate by approximately one percentage point. The model indicates that as tasks require more interactions, drivers miss more DRT signals at a consistent rate of roughly one percentage point per step, showing how increasing task complexity steadily elevates cognitive demand.

For the VOT, the model takes data at the block level for each participant, IVIS task type and IVIS/vehicle.

```
Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: sqrtTSOT ~ nSteps + (nSteps | participant)
## Data: votdfCC
##
## REML criterion at convergence: 3366.1
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.2770 -0.7904  0.0099  0.5819  4.3104
##
## Random effects:
## Groups      Name                Variance Std.Dev. Corr
## participant (Intercept) 1335.56   36.545
##              nSteps      44.41    6.664   -0.18
## Residual              5569.65   74.630
## Number of obs: 290, groups: participant, 19
##
## Fixed effects:
##              Estimate Std. Error    df t value Pr(>|t|)
## (Intercept)   80.895     10.613 17.792   7.622 5.23e-07 ***
## nSteps        15.417      1.878 16.215   8.208 3.61e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##              (Intr)
## nSteps -0.364
```

For the VOT model, the fixed-effect intercept is **80.89** and the slope for the number of IVIS interactions (**nSteps**) is **15.42**. These values allow us to estimate the visual-attention cost associated with each additional interaction or “step.” The fitted model is:

$$\sqrt{TSOT} = 80.89 + 15.42 \cdot nSteps$$

Back-transforming to the original scale gives:

$$TSOT = (80.89 + 15.42 \cdot nSteps)^2$$

This value represents the **Total Shutter Open Time** for a *block* of trials. Because each VOT block consists of **10 trials**, we divide the block-level TSOT estimate by 10 to obtain the average visual-attention cost per trial. Table 3 presents these translated values. The final column (“Increase in Eyes-off-Road time for an average IVIS task”) shows how each additional IVIS interaction increases the amount of time a driver’s eyes are diverted from the road. These increases compound as the number of steps rise, illustrating the non-linear escalation in visual demand as tasks become more complex.

Number of Interactions or “Steps”	sqrtTSOT (pred)	Eyes-off-Road time (ms)	TSOT (s)	Increase in Eyes-off-Road time (ms)	Increase in “Eyes-off-Road” time for an average IVIS task (ms)
0	80.89	6,543	6.54	—	—
1	96.31	9,276	9.28	+2,733	273
2	111.73	12,479	12.48	+3,203	320
3	127.15	16,179	16.18	+3,700	370
4	142.57	20,327	20.33	+4,148	415
5	157.99	24,961	24.96	+4,634	463
6	173.41	30,072	30.07	+5,111	511
7	188.83	35,682	35.68	+5,610	561
8	204.25	41,742	41.74	+6,060	606
9	219.67	48,252	48.25	+6,510	651
10	235.09	55,267	55.27	+7,015	702
11	250.51	62,755	62.76	+7,488	749
12	265.93	70,717	70.72	+7,962	796

Table 7. Model predictions for the visual-attention “cost” associated with each additional IVIS interaction or “step”. The final column (“Increase in Eyes-Off-Road time for an average IVIS task”) translates these model estimates into interpretable units, showing how each extra interaction incrementally increases the amount of time a driver’s eyes are diverted from the road. These increases accumulate in a non-linear manner as tasks require more steps, reflecting the compounding attentional demands associated with more complex IVIS interactions.

28 Appendix 6 – Linear mixed effects model for Drive DRT Misses

28.1 Overall IVIS Effects (Collapsed Across Tasks)

A linear mixed-effects model was used to examine how missed DRT stimuli (pEr; proportion of missed responses) varied across IVIS systems and IVIS task types. The model included fixed effects for IVIS, task type, and their interaction, along with a random intercept for participant to account for repeated measures. The model converged successfully (REML = – 318.8) and residual diagnostics indicated no concerning deviations from normality or homoscedasticity. The random-effects structure demonstrated meaningful between-participant variability in DRT miss propensity.

28.2 Main Findings: IVIS Effects

Across IVIS systems, most coefficients were positive, indicating higher DRT miss rates during the Initiate Call task, which serves as the model's reference task type. Because the model is specified without a global intercept ($0 + \text{IVIS} + \text{IVIS}:\text{taskType}$), each IVIS coefficient reflects the estimated probability of missing a DRT stimulus for that system during the Initiate Call task. Positive values therefore indicate that the system imposed greater cognitive load relative to the model's implicit baseline.

Several systems showed statistically significant increases in miss rate, demonstrating that engaging with these interfaces impaired drivers' ability to detect DRT stimuli:

- Tesla Model Y (Estimate = 0.376, $p < .001$)
- Nissan X-Trail (0.237, $p < .001$)
- Toyota Hilux (0.161, $p < .001$)
- BMW 330i (0.147, $p = .040$)
- Toyota Corolla (0.127, $p = .024$)
- Audi Q7 (0.112, $p = .048$)

Other systems showed small, non-significant increases (e.g., Android Auto, CarPlay, Ford Ranger, Mazda CX-9, Cupra Ateca), while Polestar 2 displayed a negligible negative coefficient.

Overall, the pattern confirms that DRT Miss Rate is sensitive to IVIS-induced cognitive workload, with several systems producing substantial and reliable increases in missed stimuli.

28.3 Model Specification

Model type: Linear Mixed-Effects Model

Outcome variable: pEr - proportion of missed DRT signals

Fixed effects: IVIS, Task Type, IVIS $\times$ Task Type

Random effects: Random intercept for Participant

Estimation: REML (lmerModLmerTest)

Observations: 290

Participants: 19

28.4 Model Fit Statistics

Statistic	Value
REML criterion	−318.8
Random intercept SD (participant)	0.080
Residual SD	0.087
Scaled residual range	−2.78 to 4.37

28.5 Fixed Effects: IVIS Main Effects

Reference Task Type = Initiate Call

These coefficients represent the expected miss rate (pEr) during the Initiate Call task for each IVIS. The DRT consistently detects reductions in attentional capacity during IVIS interactions, with several systems producing large and reliable increases in missed stimuli.

IVIS System	Estimate	SE	df	t	p-value
Android Auto	0.0838	0.0561	207.1	1.49	.137
Apple CarPlay	0.0134	0.0561	207.1	0.24	.811
Audi Q7	0.1116	0.0561	207.1	1.99	.048 *
BMW 330i	0.1473	0.0711	203.3	2.07	.040 *
Cupra Ateca	0.1100	0.0677	214.8	1.63	.106
Ford Ranger	0.0716	0.0561	207.1	1.28	.204
Mazda CX-9	0.0667	0.0561	207.1	1.19	.236
Nissan X-Trail	0.2370	0.0561	207.1	4.22	3.62e-05 ***
Polestar 2	−0.0164	0.0561	207.1	−0.29	.770
Tesla Model Y	0.3756	0.0677	214.8	5.55	8.37e-08 ***
Toyota Corolla	0.1273	0.0561	207.1	2.27	.024 *
Toyota Hilux	0.1611	0.0276	64.5	5.83	1.94e-07 ***

28.6 Fixed Effects: IVIS $\times$ Task-Type Interactions

These interaction coefficients indicate how each task type differs from Initiate Call *within* an IVIS system; therefore, comparisons can ONLY be made between IVIS task types WITHIN an IVIS. We do not perform post-hoc IVIS-to-IVIS comparisons because the goal is to demonstrate method sensitivity, not to statistically rank systems. Pairwise contrasts would introduce hundreds of unnecessary comparisons, inflate Type I/II error, and risk misinterpretation regarding system performance that is not relevant to the final star-rating framework. To maintain clarity and follow the tracking-error appendix format, interactions are grouped by task type.

A. Initiate Text

IVIS	Est	SE	df	t	p
Android Auto	0.0179	0.0708	198.4	0.25	.801
Apple CarPlay	0.0807	0.0708	198.4	1.14	.256
Audi Q7	-0.0240	0.0708	198.4	-0.34	.735
BMW 330i	0.0712	0.0868	198.4	0.82	.413
Cupra Ateca	0.1152	0.0868	198.4	1.33	.186
Polestar 2	0.0749	0.0708	198.4	1.06	.291
Tesla Model Y	-0.0449	0.108	198.9	-0.42	.677

B. Media

IVIS	Est	SE	df	t	p
Android Auto	0.0269	0.0708	198.4	0.38	.704
Apple CarPlay	0.1177	0.0708	198.4	1.66	.098
Audi Q7	-0.0172	0.0708	198.4	-0.24	.808
BMW 330i	-0.0012	0.0868	198.4	-0.01	.989
Cupra Ateca	0.0348	0.0868	198.4	0.40	.689
Ford Ranger	0.0308	0.0708	198.4	0.43	.664
Mazda CX-9	0.1457	0.0708	198.4	2.06	.041 *
Nissan X-Trail	-0.0668	0.0708	198.4	-0.94	.346
Polestar 2	0.0737	0.0708	198.4	1.04	.299
Tesla Model Y	-0.0266	0.0868	198.4	-0.31	.759
Toyota Corolla	-0.0384	0.0708	198.4	-0.54	.588
Toyota Hilux	0.0168	0.0289	198.4	0.58	.563

C. Navigation

IVIS	Est	SE	df	t	p
Android Auto	-0.0032	0.0708	198.4	-0.05	.964
Apple CarPlay	0.1577	0.0708	198.4	2.23	.027 *
Audi Q7	-0.0147	0.0708	198.4	-0.21	.836
BMW 330i	0.1322	0.0868	198.4	1.52	.129

IVIS	Est	SE	df	t	p
Cupra Ateca	0.0435	0.0868	198.4	0.50	.616
Ford Ranger	0.0527	0.0708	198.4	0.74	.458
Mazda CX-9	0.2522	0.0708	198.4	3.56	0.00046 ***
Polestar 2	0.0306	0.0708	198.4	0.43	.667
Tesla Model Y	-0.0952	0.0868	198.4	-1.10	.274
Toyota Hilux	0.0966	0.0289	198.4	3.34	.001 ***

D. Radio

IVIS	Est	SE	df	t	p
Audi Q7	-0.0134	0.0708	198.4	-0.19	.850
BMW 330i	0.0387	0.0868	198.4	0.45	.656
Cupra Ateca	0.0907	0.0868	198.4	1.05	.297
Ford Ranger	0.0265	0.0708	198.4	0.37	.709
Mazda CX-9	0.1935	0.0708	198.4	2.73	.0069 **
Nissan X-Trail	-0.0127	0.0708	198.4	-0.18	.858
Tesla Model Y	-0.1767	0.0868	198.4	-2.04	.043 *
Toyota Corolla	0.0155	0.0708	198.4	0.22	.827
Toyota Hilux	-0.0027	0.0289	198.4	-0.09	.927

E. Receive Call

IVIS	Est	SE	df	t	p
Android Auto	-0.0020	0.0708	198.4	-0.03	.978
Apple CarPlay	0.1178	0.0708	198.4	1.66	.098 .
Audi Q7	-0.0048	0.0708	198.4	-0.07	.947
BMW 330i	-0.0541	0.0868	198.4	-0.62	.533
Cupra Ateca	-0.0638	0.0868	198.4	-0.74	.463
Ford Ranger	-0.0192	0.0708	198.4	-0.27	.787
Mazda CX-9	0.0180	0.0708	198.4	0.25	.800
Nissan X-Trail	-0.1164	0.0708	198.4	-1.64	.102
Polestar 2	0.0370	0.0708	198.4	0.52	.602
Tesla Model Y	-0.1845	0.0868	198.4	-2.13	.0347 *
Toyota Corolla	-0.0122	0.0708	198.4	-0.17	.863
Toyota Hilux	-0.0566	0.0289	198.4	-1.96	.0518 .

F. Receive Text

IVIS	Est	SE	df	t	p
Android Auto	-0.0347	0.0708	198.4	-0.49	.625
Apple CarPlay	0.0236	0.0708	198.4	0.33	.740
Audi Q7	0.0003	0.0708	198.4	0.00	.997
BMW 330i	-0.0883	0.0868	198.4	-1.02	.310

IVIS	Est	SE	df	t	p
Cupra Ateca	−0.0149	0.0868	198.4	−0.17	.864
Ford Ranger	0.0033	0.0708	198.4	0.05	.963
Mazda CX-9	−0.0066	0.0708	198.4	−0.09	.925
Nissan X-Trail	−0.1595	0.0708	198.4	−2.25	.025 *
Polestar 2	0.0432	0.0708	198.4	0.61	.543
Tesla Model Y	−0.2735	0.0868	198.4	−3.15	.0019 **
Toyota Corolla	−0.0344	0.0708	198.4	−0.49	.628
Toyota Hilux	−0.0732	0.0289	198.4	−2.53	.0121 *

N.B. A small number of IVIS $\times$ task-type interaction terms are negative. These negative coefficients do not indicate that a system reduces distraction or improves driver performance. Instead, they arise because the *reference task* in the model (the baseline against which all other tasks are compared) was not the easiest or lowest-demand task. For several IVIS, the reference task was more demanding than the other tasks (e.g., “Initiate Call” is not always the simplest). As a result, when a system performs slightly better on another task type, the interaction term becomes negative purely due to how the contrast is coded, not because the task reduces workload.

This issue disappears entirely if the data are re-coded so that the lowest-demand task is used as the reference category. In that coding scheme, all interaction coefficients would be zero or positive, aligning with the theoretical expectation that more complex interactions impose equal or greater cognitive load relative to the simplest task.

29 Appendix 7 – Linear Mixed-Effects Model: Drive DRT Response Time

29.1 Overview of Model

A linear mixed-effects model was used to examine how mean DRT reaction time (mRT) varied across IVIS systems and IVIS task types during on-track driving. Fixed effects included IVIS, task type, and their interaction, with a random intercept for participant to account for repeated measures. The model converged successfully (REML = 2719.1), and residuals showed no concerning deviations from normality or homoscedasticity. The random-effects structure demonstrated substantial between-participant variability in baseline reaction speed.

Because longer RTs reflect impaired stimulus detection and slower cognitive responsiveness, increases in mRT are interpreted as evidence of higher attentional workload.

29.2 Main Findings: IVIS Effects

All IVIS systems produced large, positive, and highly significant increases in mean DRT reaction time relative to the reference task condition (Initiate Call). Because the model uses 0 + IVIS + IVIS:taskType, each IVIS coefficient represents the mean RT for that system during the Initiate Call task.

Several systems demonstrated pronounced RT slowing, indicating substantial cognitive load:

- Mazda CX-9 (Estimate = 788.8 ms, $p < .001$)
- Tesla Model Y (871.0 ms, $p < .001$)
- BMW 330i (784.1 ms, $p < .001$)
- Toyota Hilux (670.6 ms, $p < .001$)
- Ford Ranger (667.6 ms, $p < .001$)
- Nissan X-Trail (658.0 ms, $p < .001$)
- Cupra Ateca (651.5 ms, $p < .001$)

Systems with relatively shorter RTs (but still substantial relative to baseline) included:

- Polestar 2 (408.7 ms, $p < .001$)
- Android Auto (483.2 ms, $p < .001$)
- Apple CarPlay (520.1 ms, $p < .001$)
-

The consistent pattern of large, significant increases confirms that DRT reaction time is highly sensitive to IVIS-induced cognitive workload.

29.3 Model Specification

Model type: Linear Mixed-Effects Model

Outcome variable: mRT — mean DRT reaction time (ms)

Fixed effects: IVIS, Task Type, IVIS $\times$ Task Type

Random effects: Random intercept for Participant

Estimation: REML (`lmerModLmerTest`)

Observations: 290

Participants: 19

29.4 Model Fit Statistics

Statistic	Value
REML criterion	2719.1
Random intercept SD (participant)	178.05 ms
Residual SD	92.83 ms
Scaled residual range	−2.56 to 2.60

29.5 Fixed Effects: IVIS Main Effects

Reference Task Type = Initiate Call

These coefficients represent the predicted reaction time (ms) during the Initiate Call task for each IVIS.

IVIS System	Estimate (ms)	SE	df	t	p
Android Auto	483.2	70.28	111.5	6.88	p < .001
Apple CarPlay	520.1	70.28	111.5	7.40	p < .001
Audi Q7	546.9	70.28	111.5	7.78	p < .001
BMW 330i	784.1	86.54	156.8	9.06	p < .001
Cupra Ateca	651.5	81.27	151.3	8.02	p < .001
Ford Ranger	667.6	70.28	111.5	9.50	p < .001
Mazda CX-9	788.8	70.28	111.5	11.22	p < .001
Nissan X-Trail	658.0	70.28	111.5	9.36	p < .001
Polestar 2	408.7	70.28	111.5	5.82	p < .001
Tesla Model Y	871.0	81.27	151.3	10.72	p < .001
Toyota Corolla	589.2	70.28	111.5	8.38	p < .001
Toyota Hilux	670.6	46.44	28.3	14.44	p < .001

29.6 Fixed Effects: IVIS × Task-Type Interactions

These interaction coefficients indicate how each task type differs from Initiate Call *within* an IVIS system; therefore, comparisons can ONLY be made between IVIS task types WITHIN an IVIS. We do not perform post-hoc IVIS-to-IVIS comparisons because the goal is to demonstrate method sensitivity, not to statistically rank systems. Pairwise contrasts would introduce hundreds of unnecessary comparisons, inflate Type I/II error, and risk misinterpretation regarding system performance that is not relevant to the final star-rating framework. Interactions are grouped by task type.

A. Initiate Text

IVIS	Estimate (ms)	SE	df	t	p
Android Auto	83.9	75.80	198.2	1.11	.270
Apple CarPlay	85.2	75.80	198.2	1.12	.262
Audi Q7	-21.0	75.80	198.2	-0.28	.782
BMW 330i	29.4	92.83	198.2	0.32	.751
Cupra Ateca	51.5	92.83	198.2	0.55	.580
Polestar 2	-20.1	75.80	198.2	-0.27	.792
Tesla Model Y	-286.0	115.23	198.4	-2.48	p = .014

B. Media

IVIS	Est (ms)	SE	df	t	p
Android Auto	166.9	75.80	198.2	2.20	p = .029
Apple CarPlay	-26.8	75.80	198.2	-0.35	.724
Audi Q7	44.3	75.80	198.2	0.59	.559
BMW 330i	127.0	92.83	198.2	1.37	.173
Cupra Ateca	-52.4	92.83	198.2	-0.56	.573
Ford Ranger	-41.7	75.80	198.2	-0.55	.583
Mazda CX-9	-34.6	75.80	198.2	-0.46	.649
Nissan X-Trail	-6.5	75.80	198.2	-0.09	.931
Polestar 2	58.8	75.80	198.2	0.78	.439
Tesla Model Y	-145.5	92.83	198.2	-1.57	.119
Toyota Corolla	142.6	75.80	198.2	1.88	.061
Toyota Hilux	61.6	30.94	198.2	1.99	p = .048

C. Navigation

IVIS	Est (ms)	SE	df	t	p
Android Auto	125.9	75.80	198.2	1.66	.098
Apple CarPlay	114.5	75.80	198.2	1.51	.132
Audi Q7	10.5	75.80	198.2	0.14	.890
BMW 330i	152.3	92.83	198.2	1.64	.103
Cupra Ateca	-16.0	92.83	198.2	-0.17	.863
Ford Ranger	16.8	75.80	198.2	0.22	.824

IVIS	Est (ms)	SE	df	t	p
Mazda CX-9	52.3	75.80	198.2	0.69	.491
Nissan X-Trail	-35.1	75.80	198.2	-0.46	.643
Polestar 2	22.5	75.80	198.2	0.30	.767
Tesla Model Y	-22.7	92.83	198.2	-0.24	.807
Toyota Hilux	62.1	30.94	198.2	2.01	p = .046

D. Radio

IVIS	Est (ms)	SE	df	t	p
Audi Q7	-28.8	75.80	198.2	-0.38	.704
BMW 330i	92.9	92.83	198.2	1.00	.318
Cupra Ateca	-15.2	92.83	198.2	-0.16	.870
Ford Ranger	0.85	75.80	198.2	0.01	.991
Mazda CX-9	45.1	75.80	198.2	0.60	.552
Nissan X-Trail	94.1	75.80	198.2	1.24	.216
Tesla Model Y	-45.9	92.83	198.2	-0.49	.622
Toyota Corolla	118.0	75.80	198.2	1.56	.121
Toyota Hilux	18.0	30.94	198.2	0.58	.561

E. Receive Call

IVIS	Est (ms)	SE	df	t	p
Android Auto	21.3	75.80	198.2	0.28	.779
Apple CarPlay	33.6	75.80	198.2	0.44	.658
Audi Q7	-9.4	75.80	198.2	-0.13	.901
BMW 330i	-174.4	92.83	198.2	-1.88	.062
Cupra Ateca	-210.8	92.83	198.2	-2.27	p = .024
Ford Ranger	-166.9	75.80	198.2	-2.20	p = .029
Mazda CX-9	-108.8	75.80	198.2	-1.44	.153
Nissan X-Trail	-35.1	75.80	198.2	-0.46	.643
Polestar 2	19.6	75.80	198.2	0.26	.796
Tesla Model Y	-134.7	92.83	198.2	-1.45	.148
Toyota Corolla	-67.8	75.80	198.2	-0.89	.372
Toyota Hilux	-123.9	30.94	198.2	-4.00	p < .001

F. Receive Text

IVIS	Est (ms)	SE	df	t	p
Android Auto	73.3	75.80	198.2	0.97	.335
Apple CarPlay	60.2	75.80	198.2	0.80	.428
Audi Q7	0.45	75.80	198.2	0.01	.995
BMW 330i	-80.2	92.83	198.2	-0.86	.389
Cupra Ateca	-171.5	92.83	198.2	-1.85	.066 .
Ford Ranger	-138.8	75.80	198.2	-1.83	.069 .
Mazda CX-9	-45.4	75.80	198.2	-0.60	.550
Nissan X-Trail	-72.6	75.80	198.2	-0.96	.339
Polestar 2	109.5	75.80	198.2	1.44	.150
Tesla Model Y	-185.1	92.83	198.2	-1.99	p = .048
Toyota Corolla	-90.4	75.80	198.2	-1.19	.235
Toyota Hilux	-93.0	30.94	198.2	-3.01	p = .003

30 Appendix 8 - Linear Mixed-Effects Model — Visual Occlusion Test – Total Shutter Open Time (TSOT)

30.1 Overview of Model

A linear mixed-effects model was used to examine how visual attention shifting, quantified as square-root-transformed Total Shifting of Attention Time (sqrtTSOT), varied across IVIS systems and VOT task types. Fixed effects included vehicle (car), task type, and their interaction, with a random intercept for participant to account for repeated measures in the stationary VOT. The model converged successfully (REML = 2499.1), and residual diagnostics showed no concerning deviations from normality or homoscedasticity.

TSOT captures the total time spent shifting gaze between the “roadway” and the IVIS display, higher sqrtTSOT values indicate greater visual demand and more prolonged gaze shifts away from the primary task of driving.

30.2 Main Findings: IVIS Effects

The model was specified as $0 + \text{car} + \text{car}:\text{taskType}$, so each car main effect represents the expected sqrtTSOT for that vehicle during the Initiate Call task (the reference task type). Several vehicles showed substantial and statistically reliable increases in sqrtTSOT during Initiate Call, indicating higher visual workload:

- Nissan X-Trail (Estimate = 265.1, $p < .001$)
- Tesla Model Y (202.2, $p < .001$)
- Audi Q7 (179.8, $p < .001$)
- Cupra Ateca (147.1, $p = .002$)
- Toyota Hilux (144.3, $p < .001$)
- Mazda CX-9 (118.6, $p = .002$)

Apple CarPlay showed a smaller, borderline effect (73.6, $p = .053$). Other systems (Android Auto, BMW 330i, Ford Ranger, Polestar 2, Toyota Corolla) had relatively modest and non-significant main effects for the Initiate Call task.

Taken together, these results demonstrate that, even for a relatively simple VOT task like Initiate Call, some IVIS configurations (particularly Nissan X-Trail, Tesla Model Y, Audi Q7, Cupra Ateca, Mazda CX-9, and Toyota Hilux) impose substantial visual attention shifting demands.

30.3 Model Specification

Model type: Linear Mixed-Effects Model

Outcome variable: sqrtTSOT — square-root-transformed Total Shifting of Attention Time

Fixed effects: car, task Type, car $\times$ task Type

Random effects: Random intercept for Participant

Estimation: REML (lmerModLmerTest)

Observations: 290

Participants: 19

30.4 Model Fit Statistics

Statistic	Value
REML criterion	2499.1
Random intercept SD (participant)	41.17
Residual SD	60.23
Scaled residual range	−2.24 to 3.91

30.5 Fixed Effects: IVIS Main Effects

Reference Task Type = Initiate Call

These coefficients represent the predicted sqrtTSOT for each IVIS during the Initiate Call task.

Vehicle (IVIS)	Estimate	SE	df	t	p
Android Auto	44.77	37.75	214.7	1.19	.237
Apple CarPlay	73.60	37.75	214.7	1.95	.053
Audi Q7	179.84	37.75	214.7	4.76	p < .001
BMW 330i	19.33	46.15	215.9	0.42	.676
Cupra Ateca	147.12	46.15	215.9	3.19	p = .002
Ford Ranger	20.59	37.75	214.7	0.55	.586
Mazda CX-9	118.60	37.75	214.7	3.14	p = .002
Nissan X-Trail	265.06	37.75	214.7	7.02	p < .001
Polestar 2	24.30	37.75	214.7	0.64	.521
Tesla Model Y	202.18	46.15	215.9	4.38	p < .001
Toyota Corolla	51.90	37.75	214.7	1.38	.171
Toyota Hilux	144.26	16.74	101.4	8.62	p < .001

30.6 Fixed Effects: Vehicle $\times$ Task-Type Interactions

These interaction coefficients indicate how each task type differs from Initiate Call *within* an IVIS system; therefore, comparisons can ONLY be made between IVIS task types WITHIN an IVIS. We do not perform post-hoc IVIS-to-IVIS comparisons because the goal is to demonstrate method sensitivity, not to statistically rank systems. Pairwise contrasts would introduce hundreds of unnecessary comparisons, inflate Type I/II error, and risk misinterpretation regarding system performance that is not relevant to the final star-rating framework. Results are grouped by task type.

A. Initiate Text

Vehicle	Estimate	SE	df	t	p
Android Auto	-41.20	49.17	199.4	-0.84	.403
Apple CarPlay	35.45	49.17	199.4	0.72	.472
Audi Q7	-9.78	49.17	199.4	-0.20	.843
BMW 330i	92.87	60.23	199.4	1.54	.125
Cupra Ateca	301.35	60.23	199.4	5.00	p < .001
Polestar 2	9.45	49.17	199.4	0.19	.848
Tesla Model Y	72.70	60.23	199.4	1.21	.229

B. Media

Vehicle	Estimate	SE	df	t	p
Android Auto	154.87	49.17	199.4	3.15	p = .002
Apple CarPlay	30.96	49.17	199.4	0.63	.530
Audi Q7	8.39	49.17	199.4	0.17	.865
BMW 330i	174.71	60.23	199.4	2.90	p = .004
Cupra Ateca	120.76	60.23	199.4	2.01	p = .046
Ford Ranger	118.41	49.17	199.4	2.41	p = .017
Mazda CX-9	65.14	49.17	199.4	1.33	.187
Nissan X-Trail	-28.30	49.17	199.4	-0.58	.566
Polestar 2	126.36	49.17	199.4	2.57	p = .011
Tesla Model Y	-30.91	60.23	199.4	-0.51	.608
Toyota Corolla	94.01	49.17	199.4	1.91	.057
Toyota Hilux	71.86	19.54	199.4	3.68	p < .001

C. Navigation

Vehicle	Estimate	SE	df	t	p
Android Auto	-33.64	49.17	199.4	-0.68	.495
Apple CarPlay	86.83	49.17	199.4	1.77	.079 .
Audi Q7	157.21	49.17	199.4	3.20	p = .002
BMW 330i	188.22	60.23	199.4	3.13	p = .002
Cupra Ateca	302.87	60.23	199.4	5.03	p < .001
Ford Ranger	223.56	49.17	199.4	4.55	p < .001
Mazda CX-9	201.21	49.17	199.4	4.09	p < .001
Polestar 2	58.88	49.17	199.4	1.20	.233
Tesla Model Y	207.37	60.23	199.4	3.44	p < .001
Toyota Hilux	212.77	22.56	203.3	9.43	p < .001

D. Radio

Vehicle	Estimate	SE	df	t	p
Audi Q7	-92.49	49.17	199.4	-1.88	.061 .
BMW 330i	136.90	60.23	199.4	2.27	p = .024
Cupra Ateca	159.46	60.23	199.4	2.65	p = .009
Ford Ranger	-17.01	49.17	199.4	-0.35	.730
Mazda CX-9	55.34	49.17	199.4	1.13	.262
Nissan X-Trail	27.36	49.17	199.4	0.56	.579
Tesla Model Y	30.59	60.23	199.4	0.51	.612
Toyota Corolla	76.61	49.17	199.4	1.56	.121
Toyota Hilux	-4.06	19.54	199.4	-0.21	.836

E. Receive Call

Vehicle	Estimate	SE	df	t	p
Android Auto	-12.39	49.17	199.4	-0.25	.801
Apple CarPlay	13.00	49.17	199.4	0.26	.792
Audi Q7	-91.43	49.17	199.4	-1.86	.064 .
BMW 330i	19.37	60.23	199.4	0.32	.748
Cupra Ateca	-13.18	60.23	199.4	-0.22	.827
Ford Ranger	54.09	49.17	199.4	1.10	.273
Mazda CX-9	-62.49	49.17	199.4	-1.27	.205
Nissan X-Trail	-177.20	49.17	199.4	-3.60	p < .001
Polestar 2	-12.91	49.17	199.4	-0.26	.793
Tesla Model Y	-159.78	60.23	199.4	-2.65	p = .009
Toyota Corolla	5.47	49.17	199.4	0.11	.912
Toyota Hilux	-69.31	19.54	199.4	-3.55	p < .001

F. Receive Text

Vehicle	Estimate	SE	df	t	p
Android Auto	100.05	49.17	199.4	2.04	p = .043
Apple CarPlay	67.19	49.17	199.4	1.37	.173
Audi Q7	28.45	49.17	199.4	0.58	.564
BMW 330i	175.80	60.23	199.4	2.92	p = .004
Cupra Ateca	42.72	60.23	199.4	0.71	.479
Ford Ranger	133.38	49.17	199.4	2.71	p = .007
Mazda CX-9	-53.48	49.17	199.4	-1.09	.278
Nissan X-Trail	-167.92	49.17	199.4	-3.42	p < .001
Polestar 2	113.46	49.17	199.4	2.31	p = .022
Tesla Model Y	-163.23	60.23	199.4	-2.71	p = .007
Toyota Corolla	-59.08	49.17	199.4	-1.20	.231
Toyota Hilux	-4.81	19.54	199.4	-0.25	.806

31 Appendix 9 - Linear Mixed-Effects Model: Tracking Task Tracking Error

31.1 Overview of Model

A linear mixed-effects model examined how mean tracking-task error (in pixels) varied across IVIS systems and IVIS task types. The model included fixed effects for IVIS, task type, and their interaction, with a random intercept for participant to account for repeated measures. The model fit reasonably well (REML = 1072.3) and residuals showed no concerning deviations.

31.2 Main Findings: IVIS Effects

Across systems, mean tracking error increased substantially relative to the reference level, with all IVIS showing large, statistically significant positive coefficients (all $p < .001$). This indicates that performing any IVIS task, regardless of system, increased tracking error and, therefore, visual inattention, relative to baseline. The main IVIS effects are substantial and consistently positive, showing that the tracking task is sensitive to visual attention demands imposed by IVIS interactions.

Systems associated with the largest increases in tracking error included:

- **Nissan X-Trail** (Estimate = 14.21, $p < 2e-16$)
- **Audi Q7** (10.65, $p < 5e-12$)
- **Tesla Model Y** (9.20, $p = 2.5e-7$)
- **Mazda CX-9** (8.77, $p = 6e-9$)

Systems with comparatively **lower tracking error increases** were:

- **Ford Ranger** (5.25, $p < .001$)
- **BMW 330i** (6.58, $p < .001$)
- **Polestar 2** (6.73, $p < .001$)

31.3 Model Specification

Model: Linear Mixed-Effects Model

Outcome variable: mean Tracking Error (continuous tracking task error in pixels)

Fixed effects: IVIS, Task Type, IVIS $\times$ Task Type

Random effects: Random intercept for Participant

Estimation method: REML (lmerModLmerTest)

Number of observations: 290

Participants: 19

31.4 Model Fit Statistics

Statistic	Value
REML criterion at convergence	1072.3
Random intercept SD (participant)	2.675
Residual SD	2.117
Scaled residual range	-2.86 to 3.39

31.5 Fixed Effects: IVIS Main Effects

Reference task type = Initiate Call

These estimates represent the expected tracking error (in task units) for each IVIS system during the Initiate Call task.

IVIS System	Estimate	SE	df	t	p-value
Android Auto	8.130	1.4368	181.878	5.659	< .001 ***
Apple CarPlay	7.895	1.4368	181.878	5.495	< .001 ***
Audi Q7	10.654	1.4368	181.878	7.415	< .001 ***
BMW 330i	6.580	1.7246	204.560	3.816	< .001 ***
Cupra Ateca	7.998	1.7246	204.560	4.637	< .001 ***
Ford Ranger	5.250	1.4368	181.878	3.654	< .001 ***
Mazda CX-9	8.774	1.4368	181.878	6.107	< .001 ***
Nissan X-Trail	14.209	1.4368	181.878	9.889	< .001 ***
Polestar 2	6.732	1.4368	181.878	4.685	< .001 ***
Tesla Model Y	9.201	1.7246	204.560	5.335	< .001 ***
Toyota Corolla	7.589	1.4368	181.878	5.282	< .001 ***
Toyota Hilux	8.630	0.7827	41.576	11.026	< .001 ***

31.6 Fixed Effects: IVIS $\times$ Task-Type Interactions

These interaction coefficients indicate how each task type differs from Initiate Call *within* an IVIS system; therefore, comparisons can ONLY be made between IVIS task types WITHIN an IVIS. We do not perform post-hoc IVIS-to-IVIS comparisons because the goal is to demonstrate method sensitivity, not to statistically rank systems. Pairwise contrasts would introduce hundreds of unnecessary comparisons, inflate Type I/II error, and risk misinterpretation regarding system performance that is not relevant to the final star-rating framework.

To maximise clarity, the interactions are formatted by task type.

A. Initiate Text

IVIS	Estimate	SE	df	t	p
Android Auto	-2.771	1.729	198.43	-1.60	.110
Apple CarPlay	-0.035	1.729	198.43	-0.02	.984
Audi Q7	-2.044	1.729	198.43	-1.18	.238
BMW 330i	-0.147	2.117	198.43	-0.07	.945
Cupra Ateca	7.733	2.117	198.43	3.65	< .001 ***
Polestar 2	3.391	1.729	198.43	1.96	.051
Tesla Model Y	-0.533	2.117	198.43	-0.25	.801

B. Media

IVIS	Estimate	SE	df	t	p
Android Auto	2.468	1.729	198.43	1.43	.155
Apple CarPlay	-0.213	1.729	198.43	-0.12	.902
Audi Q7	2.324	1.729	198.43	1.34	.180
BMW 330i	7.678	2.117	198.43	3.63	< .001 ***
Cupra Ateca	3.655	2.117	198.43	1.73	.086
Ford Ranger	4.288	1.729	198.43	2.48	.014 *
Mazda CX-9	0.920	1.729	198.43	0.53	.595
Nissan X-Trail	-4.228	1.729	198.43	-2.45	.015 *
Polestar 2	4.549	1.729	198.43	2.63	.009 **
Tesla Model Y	0.376	2.117	198.43	0.18	.859
Toyota Corolla	2.152	1.729	198.43	1.25	.215
Toyota Hilux	2.725	0.687	198.43	3.97	< .001 ***

C. Navigation

IVIS	Estimate	SE	df	t	p
Android Auto	-1.847	1.729	198.43	-1.07	.287
Apple CarPlay	1.108	1.729	198.43	0.64	.522
Audi Q7	-0.111	1.729	198.43	-0.06	.949
BMW 330i	1.515	2.117	198.43	0.72	.475
Cupra Ateca	3.485	2.117	198.43	1.65	.101
Ford Ranger	6.141	1.729	198.43	3.55	< .001 ***
Mazda CX-9	5.536	1.729	198.43	3.20	.002 **
Polestar 2	1.480	1.729	198.43	0.86	.393
Tesla Model Y	-0.566	2.117	198.43	-0.27	.789
Toyota Corolla	2.152	1.729	198.43	1.25	.215
Toyota Hilux	4.339	0.796	200.17	5.45	< .001 ***

D. Radio

IVIS	Estimate	SE	df	t	p
Audi Q7	-1.731	1.729	198.43	-1.00	.318
BMW 330i	3.854	2.117	198.43	1.82	.070
Cupra Ateca	1.837	2.117	198.43	0.87	.387
Ford Ranger	-0.281	1.729	198.43	-0.16	.871
Mazda CX-9	1.065	1.729	198.43	0.62	.539
Nissan X-Trail	-0.581	1.729	198.43	-0.34	.737
Tesla Model Y	-2.338	2.117	198.43	-1.11	.271
Toyota Corolla	0.220	1.729	198.43	0.13	.899
Toyota Hilux	-0.155	0.687	198.43	-0.23	.822

E. Receive Call

IVIS	Estimate	SE	df	t	p
Android Auto	-1.532	1.729	198.43	-0.89	.376
Apple CarPlay	0.700	1.729	198.43	0.41	.686
Audi Q7	-2.502	1.729	198.43	-1.45	.149
BMW 330i	-0.273	2.117	198.43	-0.13	.898
Cupra Ateca	1.046	2.117	198.43	0.49	.622
Ford Ranger	0.849	1.729	198.43	0.49	.624
Mazda CX-9	-1.417	1.729	198.43	-0.82	.413
Nissan X-Trail	-4.157	1.729	198.43	-2.41	.017 *
Polestar 2	1.605	1.729	198.43	0.93	.354
Tesla Model Y	0.759	2.117	198.43	0.36	.720
Toyota Corolla	-0.492	1.729	198.43	-0.29	.776
Toyota Hilux	-0.875	0.687	198.43	-1.27	.204

F. Receive Text

IVIS	Estimate	SE	df	t	p
Android Auto	-1.650	1.729	198.43	-0.96	.341
Apple CarPlay	-0.899	1.729	198.43	-0.52	.603
Audi Q7	-0.700	1.729	198.43	-0.41	.686
BMW 330i	3.121	2.117	198.43	1.47	.142
Cupra Ateca	0.171	2.117	198.43	0.08	.936
Ford Ranger	1.094	1.729	198.43	0.63	.528
Mazda CX-9	-1.605	1.729	198.43	-0.93	.354
Nissan X-Trail	-5.578	1.729	198.43	-3.23	.001 **
Polestar 2	-0.036	1.729	198.43	-0.02	.983
Tesla Model Y	0.987	2.117	198.43	0.47	.641
Toyota Corolla	-0.334	1.729	198.43	-0.19	.847
Toyota Hilux	-0.802	0.687	198.43	-1.17	.244

32 Appendix 10 – Linear Mixed-Effects Model – Tracking Task – visual DRT Response Times

32.1 Overview of Model

A linear mixed-effects model was used to examine how mean vDRT reaction time (mRT, in milliseconds) during the continuous tracking task varied across IVIS systems and task types. The model included fixed effects for IVIS, task type, and their interaction, with a random intercept for participant to account for repeated observations. Reaction times were analysed on the raw millisecond scale. The model converged successfully, showed no abnormal residual patterns, and provided a good-quality fit to the data (REML = 2464.2).

32.2 Main Findings: IVIS Effects

All IVIS systems produced substantially elevated vDRT reaction times relative to undistracted driving expectations, reflecting the attentional demands imposed by simultaneously performing the tracking task and IVIS interactions. During the reference task (Initiate Call), every IVIS system showed significantly slower responses compared with the model's implicit baseline (all $p < .001$), demonstrating strong between-system sensitivity. Task-type effects varied by system: Media and Navigation tasks produced notable increases in reaction times for several systems, particularly the Ford Ranger, Mazda CX-9, Toyota Corolla, and Toyota Hilux. Radio interactions yielded slower responses for the BMW 330i. Telephony tasks showed mixed patterns, with Apple CarPlay producing faster responses for Receive Call while the Toyota Hilux exhibited slower responses for Receive Call and Receive Text. These findings confirm that tracking-task DRT reaction time is sensitive to both system-level differences and the additional workload imposed by specific IVIS tasks.

32.3 Model Specification

Model: Linear Mixed-Effects Model

Outcome variable: Mean vDRT Reaction Time (mRT, ms)

Fixed effects: IVIS, Task Type, IVIS $\times$ Task Type

Random effects: Random intercept for participant

Estimation method: REML (lmerModLmerTest)

Number of observations: 290

Participants: 19

32.4 Model Fit Statistics

Statistic	Value
REML criterion at convergence	2464.2
Random intercept SD (participant)	78.67
Residual SD	52.38
Scaled residual range	−2.09 to 3.86

32.5 Fixed Effects: IVIS Main Effects

Reference task type = Initiate Call. These coefficients represent the expected vDRT reaction time (ms) for each IVIS during the Initiate Call task.

IVIS System	Estimate	SE	df	t	p
Android Auto	606.354	36.907	153.830	16.429	< .001 ***
Apple CarPlay	631.191	36.907	153.830	17.102	< .001 ***
Audi Q7	633.468	36.907	153.830	17.164	< .001 ***
BMW 330i	559.383	43.851	188.618	12.757	< .001 ***
Cupra Ateca	610.845	43.851	188.618	13.930	< .001 ***
Ford Ranger	540.661	36.907	153.830	14.649	< .001 ***
Mazda CX-9	548.609	36.907	153.830	14.865	< .001 ***
Nissan X-Trail	628.776	36.907	153.830	17.037	< .001 ***
Polestar 2	630.609	36.907	153.830	17.087	< .001 ***
Tesla Model Y	639.809	43.851	188.618	14.591	< .001 ***
Toyota Corolla	528.568	36.907	153.830	14.322	< .001 ***
Toyota Hilux	607.533	21.681	33.121	28.021	< .001 ***

32.6 Fixed Effects: IVIS × Task-Type Interactions

These interaction coefficients indicate how each task type differs from Initiate Call *within* an IVIS system; therefore, comparisons can ONLY be made between IVIS task types WITHIN an IVIS. We do not perform post-hoc IVIS-to-IVIS comparisons because the goal is to demonstrate method sensitivity, not to statistically rank systems. Pairwise contrasts would introduce hundreds of unnecessary comparisons, inflate Type I/II error, and risk misinterpretation regarding system performance that is not relevant to the final star-rating framework.

A. Initiate Text

IVIS	Estimate	SE	df	t	p
Android Auto	−10.992	42.766	197.677	−0.257	.797
Apple CarPlay	−3.580	42.766	197.677	−0.084	.933
Audi Q7	−12.417	42.766	197.677	−0.290	.772
BMW 330i	54.454	52.378	197.677	1.040	.300
Cupra Ateca	71.846	52.378	197.677	1.372	.172
Polestar 2	16.184	42.766	197.677	0.378	.706
Tesla Model Y	63.192	52.378	197.677	1.206	.229

B. Media

IVIS	Estimate	SE	df	t	p
Android Auto	63.411	42.766	197.677	1.483	.140
Apple CarPlay	-18.979	42.766	197.677	-0.444	.658
Audi Q7	15.117	42.766	197.677	0.353	.724
BMW 330i	99.518	52.378	197.677	1.900	.059
Cupra Ateca	59.229	52.378	197.677	1.131	.260
Ford Ranger	85.688	42.766	197.677	2.004	.046 *
Mazda CX-9	44.492	42.766	197.677	1.040	.299
Nissan X-Trail	13.549	42.766	197.677	0.317	.752
Polestar 2	9.412	42.766	197.677	0.220	.826
Tesla Model Y	48.590	52.378	197.677	0.928	.355
Toyota Corolla	95.774	42.766	197.677	2.239	.026 *
Toyota Hilux	50.503	16.994	197.677	2.972	.003 **

C. Navigation

IVIS	Estimate	SE	df	t	p
Android Auto	-26.213	42.766	197.677	-0.613	.541
Apple CarPlay	-35.529	42.766	197.677	-0.831	.407
Audi Q7	36.365	42.766	197.677	0.850	.396
BMW 330i	79.720	52.378	197.677	1.522	.130
Cupra Ateca	64.780	52.378	197.677	1.237	.218
Ford Ranger	97.514	42.766	197.677	2.280	.024 *
Mazda CX-9	119.773	42.766	197.677	2.801	.006 **
Polestar 2	7.374	42.766	197.677	0.172	.863
Tesla Model Y	-22.481	52.378	197.677	-0.429	.668
Toyota Hilux	43.132	19.710	199.027	2.188	.030 *

D. Radio

IVIS	Estimate	SE	df	t	p
Audi Q7	37.777	42.766	197.677	0.883	.378
BMW 330i	138.647	52.378	197.677	2.647	.009 **
Cupra Ateca	65.294	52.378	197.677	1.247	.214
Ford Ranger	-14.108	42.766	197.677	-0.330	.742
Mazda CX-9	71.734	42.766	197.677	1.677	.095
Nissan X-Trail	77.643	42.766	197.677	1.816	.071
Tesla Model Y	-22.907	52.378	197.677	-0.437	.662
Toyota Corolla	60.122	42.766	197.677	1.406	.161
Toyota Hilux	11.699	16.994	197.677	0.688	.492

E. Receive Call

IVIS	Estimate	SE	df	t	p
Android Auto	-23.017	42.766	197.677	-0.538	.591
Apple CarPlay	-110.408	42.766	197.677	-2.582	.011 *
Audi Q7	-40.784	42.766	197.677	-0.954	.341
BMW 330i	3.486	52.378	197.677	0.067	.947
Cupra Ateca	-57.445	52.378	197.677	-1.097	.274
Ford Ranger	-20.746	42.766	197.677	-0.485	.628
Mazda CX-9	-12.635	42.766	197.677	-0.295	.768
Nissan X-Trail	-72.262	42.766	197.677	-1.690	.093
Polestar 2	-37.623	42.766	197.677	-0.880	.380
Tesla Model Y	46.350	52.378	197.677	0.885	.377
Toyota Corolla	-17.492	42.766	197.677	-0.409	.683
Toyota Hilux	-53.551	16.994	197.677	-3.151	.0019 **

F. Receive Text

IVIS	Estimate	SE	df	t	p
Android Auto	-60.871	42.766	197.677	-1.423	.156
Apple CarPlay	-71.845	42.766	197.677	-1.680	.095
Audi Q7	-43.189	42.766	197.677	-1.010	.314
BMW 330i	48.641	52.378	197.677	0.929	.354
Cupra Ateca	-25.390	52.378	197.677	-0.485	.628
Ford Ranger	-60.771	42.766	197.677	-1.421	.157
Mazda CX-9	-26.450	42.766	197.677	-0.618	.537
Nissan X-Trail	-38.661	42.766	197.677	-0.904	.367
Polestar 2	-30.768	42.766	197.677	-0.719	.473
Tesla Model Y	73.106	52.378	197.677	1.396	.164
Toyota Corolla	-11.983	42.766	197.677	-0.280	.780
Toyota Hilux	-67.230	16.994	197.677	-3.956	.000106 ***

33 Appendix 11 – Linear Mixed-Effects Model – Tracking Task vDRT Misses

33.1 Overview of Model

A linear mixed-effects model examined how DRT miss probability during the tracking task varied across IVIS systems and task types. The model included fixed effects for IVIS, task type, and their interaction, with a random intercept for participant to account for repeated measures. Model fit was satisfactory (REML = -445.2), and residual diagnostics showed no concerning deviations.

33.2 Main Findings: IVIS Effects (Initiate Call Reference Task)

Most IVIS systems produced higher DRT miss rates than the reference system during the Initiate Call task, demonstrating that the tracking-task DRT miss measure is sensitive to between-system differences in attentional demand. Significant increases in miss rate were observed for:

- Android Auto (Estimate = 0.132, $p = .0016$)
- Audi Q7 (0.093, $p = .0258$)
- Mazda CX-9 (0.082, $p = .0490$)
- Nissan X-Trail (0.130, $p = .0019$)
- Toyota Hilux (0.073, $p < .001$)

Several systems produced positive but nonsignificant increases, indicating directionally higher miss probability but without reliable statistical separation from the reference condition.

33.3 Model Specification

Model: Linear Mixed-Effects Model

Outcome variable: DRT Miss Probability (pEr)

Fixed effects: IVIS, Task Type, IVIS $\times$ Task Type

Random effects: Intercept for Participant

Estimation method: REML (lmerModLmerTest)

Observations: 290

Participants: 19

33.4 Model Fit Statistics

Statistic	Value
REML criterion at convergence	-445.2
Random intercept SD (participant)	0.0435
Residual SD	0.0662
Scaled residual range	-3.39 to 3.04

33.5 Fixed Effects: IVIS Main Effects

Reference task type = Initiate Call

These coefficients represent the estimated DRT miss probability for each IVIS during the Initiate Call task.

IVIS System	Estimate	SE	df	t	p
Android Auto	0.1321	0.0414	214.83	3.19	.0016
Apple CarPlay	0.0811	0.0414	214.83	1.96	.0513
Audi Q7	0.0929	0.0414	214.83	2.25	.0258
BMW 330i	0.0500	0.0506	215.89	0.99	.325
Cupra Ateca	0.0572	0.0506	215.89	1.13	.260
Ford Ranger	0.0545	0.0414	214.83	1.32	.189
Mazda CX-9	0.0819	0.0414	214.83	1.98	.0490
Nissan X-Trail	0.1303	0.0414	214.83	3.15	.0019
Polestar 2	0.0693	0.0414	214.83	1.68	.0954
Tesla Model Y	0.0613	0.0506	215.89	1.21	.227
Toyota Corolla	0.0609	0.0414	214.83	1.47	.143
Toyota Hilux	0.0732	0.0182	99.32	4.03	<.001

33.6 Fixed Effects: IVIS × Task-Type Interactions

Each coefficient represents the change in DRT miss probability relative to the Initiate Call task within the same IVIS; therefore, comparisons can ONLY be made between IVIS task types WITHIN an IVIS. We do not perform post-hoc IVIS-to-IVIS comparisons because the goal is to demonstrate method sensitivity, not to statistically rank systems. Pairwise contrasts would introduce hundreds of unnecessary comparisons, inflate Type I/II error, and risk misinterpretation regarding system performance that is not relevant to the final star-rating framework.

A. Initiate Text

IVIS	Estimate	SE	df	t	p
Android Auto	−0.0375	0.0541	197.03	−0.694	.489
Apple CarPlay	0.0776	0.0541	197.03	1.435	.153
Audi Q7	−0.0138	0.0541	197.03	−0.255	.799
BMW 330i	−0.0175	0.0662	197.03	−0.264	.792
Cupra Ateca	−0.0321	0.0662	197.03	−0.484	.629
Polestar 2	0.1820	0.0541	197.03	3.366	.0009
Tesla Model Y	−0.0031	0.0662	197.03	−0.047	.963

B. Media

IVIS	Estimate	SE	df	t	p
Android Auto	0.0504	0.0541	197.03	0.932	.353
Apple CarPlay	-0.0203	0.0541	197.03	-0.375	.708
Audi Q7	0.0013	0.0541	197.03	0.024	.981
BMW 330i	0.0915	0.0662	197.03	1.382	.169
Cupra Ateca	0.0146	0.0662	197.03	0.220	.826
Ford Ranger	0.0743	0.0541	197.03	1.374	.171
Mazda CX-9	0.0680	0.0541	197.03	1.258	.210
Nissan X-Trail	0.0088	0.0541	197.03	0.163	.870
Polestar 2	0.2283	0.0541	197.03	4.223	3.69e-05
Tesla Model Y	-0.0047	0.0662	197.03	-0.071	.943
Toyota Corolla	-0.0135	0.0541	197.03	-0.249	.803
Toyota Hilux	0.0402	0.0215	197.03	1.869	.063

C. Navigation

IVIS	Est.	SE	df	t	p
Android Auto	-0.0229	0.0541	197.03	-0.423	.673
Apple CarPlay	0.0675	0.0541	197.03	1.249	.213
Audi Q7	-0.0065	0.0541	197.03	-0.121	.904
BMW 330i	0.0490	0.0662	197.03	0.739	.461
Cupra Ateca	0.0305	0.0662	197.03	0.460	.646
Ford Ranger	0.1072	0.0541	197.03	1.982	.0489
Mazda CX-9	0.1853	0.0541	197.03	3.426	.0007
Nissan X-Trail	0.0214	0.0541	197.03	0.395	.693
Polestar 2	0.0214	0.0541	197.03	0.395	.693
Tesla Model Y	0.0083	0.0662	197.03	0.126	.900
Toyota Hilux	0.1693	0.0248	201.69	6.826	1.0e-10

D. Radio

IVIS	Est.	SE	df	t	p
Audi Q7	-0.0117	0.0541	197.03	-0.217	.828
BMW 330i	0.1170	0.0662	197.03	1.767	.079
Cupra Ateca	-0.0039	0.0662	197.03	-0.059	.953
Ford Ranger	0.0179	0.0541	197.03	0.331	.742
Mazda CX-9	0.0922	0.0541	197.03	1.706	.090
Nissan X-Trail	-0.0091	0.0541	197.03	-0.168	.867
Tesla Model Y	0.0030	0.0662	197.03	0.045	.964
Toyota Corolla	-0.0074	0.0541	197.03	-0.138	.891
Toyota Hilux	0.0192	0.0215	197.03	0.891	.374

E. Receive Call

IVIS	Est.	SE	df	t	p
Android Auto	0.0024	0.0541	197.03	0.044	.965
Apple CarPlay	-0.0856	0.0541	197.03	-1.584	.115
Audi Q7	-0.0392	0.0541	197.03	-0.724	.470
BMW 330i	-0.0459	0.0662	197.03	-0.693	.489
Cupra Ateca	-0.0294	0.0662	197.03	-0.444	.658
Ford Ranger	0.0013	0.0541	197.03	0.023	.981
Mazda CX-9	-0.0510	0.0541	197.03	-0.944	.346
Nissan X-Trail	-0.0486	0.0541	197.03	-0.898	.370
Polestar 2	0.0105	0.0541	197.03	0.193	.847
Tesla Model Y	0.0214	0.0662	197.03	0.323	.747
Toyota Corolla	-0.0110	0.0541	197.03	-0.203	.839
Toyota Hilux	-0.0197	0.0215	197.03	-0.916	.361

F. Receive Text

IVIS	Est.	SE	df	t	p
Android Auto	-0.0631	0.0541	197.03	-1.166	.245
Apple CarPlay	-0.0714	0.0541	197.03	-1.320	.188
Audi Q7	-0.0277	0.0541	197.03	-0.513	.609
BMW 330i	-0.0035	0.0662	197.03	-0.053	.958
Cupra Ateca	-0.0092	0.0662	197.03	-0.139	.890
Ford Ranger	0.0046	0.0541	197.03	0.085	.932
Mazda CX-9	-0.0313	0.0541	197.03	-0.579	.563
Nissan X-Trail	-0.0355	0.0541	197.03	-0.657	.512
Polestar 2	0.0104	0.0541	197.03	0.193	.847
Tesla Model Y	-0.0068	0.0662	197.03	-0.103	.918
Toyota Corolla	0.0019	0.0541	197.03	0.034	.973
Toyota Hilux	-0.0316	0.0215	197.03	-1.471	.143

34 Appendix 12 - Vehicle Telemetry Model Output

34.1 Linear Mixed-Effects Model: Mean Lateral Acceleration (g)

Model specification

Linear mixed-effects model fitted by REML; t-tests use Satterthwaite's method.

Formula: $\text{meanLatAccel} \sim 0 + \text{IVIS} + \text{IVIS}:\text{taskType} + (1 \mid \text{participant})$

This parameterisation omits the global intercept, allowing each IVIS to have its own predicted baseline value for the reference task. Interaction terms represent deviations from these task-specific baselines.

Model diagnostics

Scaled residuals ranged from -3.17 to 2.77 , with a symmetric distribution and no evidence of heteroscedasticity. Random-effects variance at the participant level was small (0.005 g SD), indicating relatively consistent lane-keeping across drivers. Residual variability was also low (0.007 g SD), consistent with the low-demand driving environment.

Data characteristics

Mean lateral acceleration values ranged from approximately 0.064 – 0.078 g across systems for the reference task. These values fall well within typical lane-keeping variation on a flat proving-ground circuit. The small numerical range explains the small fixed-effect estimates and contributes to high statistical significance despite negligible practical differences.

Random effects

- Participant intercept variance = 2.70×10^{-5}
- Residual variance = 5.41×10^{-5}

Fixed effects summary

Although many IVIS coefficients reached statistical significance, these differences were extremely small in magnitude and reflect only baseline differences between vehicles in the reference task. The $\text{IVIS} \times \text{task-type}$ interactions likewise did not materially alter lateral-acceleration values. These coefficients are presented here for transparency but are not interpreted as indicators of distraction.

34.2 Linear Mixed-Effects Model: Mean Speed Delta (m/s change at 100 Hz)

Model specification

Linear mixed-effects model fitted by REML; t-tests use Satterthwaite's method.

Formula: $\text{mSpeedDelta_ms} \sim 0 + \text{IVIS} + \text{IVIS}:\text{taskType} + (1 \mid \text{participant})$

As above, the model estimates IVIS-specific baseline values and task-type deviations relative to a reference task.

Telemetry characteristics

Speed delta was calculated at a 100 Hz sampling frequency, producing extremely small numerical values (residual SD = 0.00015 m/s). Because variance at both the participant and residual levels is minimal, even trivial fixed-effect differences may yield low p-values.

Model diagnostics

Residuals were narrowly distributed and centred near zero, with no convergence issues. Participant-level variance was estimated at zero, indicating negligible between-driver variability in this measure.

Random effects

- Participant intercept variance = 0
- Residual variance = 2.236×10^{-8}

Fixed effects summary

IVIS main effects ranged from -0.00019 to $+0.00005$ m/s, values far below any threshold of practical significance in driving behaviour. Although one IVIS coefficient and one interaction term reached statistical significance, these reflect numerical precision rather than meaningful effects. Speed-delta coefficients are included here for completeness and transparency only.

35 Model Selection and Validation for the Step-Count Metric

35.1.1.1 Overview

We evaluated three mixed-effects model specifications independently for both the DRT and VOT datasets to determine the most appropriate statistical framework for the step-count metric. Model fit was assessed using Restricted Maximum Likelihood (REML), and predictive validity was determined via Leave-One-Out Cross-Validation (LOOCV). While more complex specifications provided a slightly better statistical fit for certain outcomes, **Model 1 (Random Intercept Only)** was selected as the universal specification for this report.

35.1.1.2 Rationale for Selection

The decision to utilize Model 1 across all metrics was based on the following:

1. Superior Predictive Stability (VOT): In the out-of-sample predictive validation for the VOT, Model 1 demonstrated the highest correlation between predicted and observed values ($r = 0.653$), outperforming both the Interaction model ($r = 0.646$) and the Random Slope model ($r = 0.633$). This suggests that the additional parameters in the complex models introduced over-fitting rather than improved generalisability.
2. Metric Standardisation: By applying the same model specification to both DRT and VOT, the "Step Count" remains a consistent and comparable measure across different modalities of distraction. While Model 3 offered a slightly higher correlation for the DRT ($r = 0.748$ vs $r = 0.710$), the marginal gain did not justify using divergent methodologies.
3. Computational Robustness: Model 1 consistently reached numerical convergence across all cross-validation folds. In contrast, the Random Intercept and Slope specification (Model 3) frequently failed to converge, indicating that the model was over-parameterized for the current dataset.
4. Parsimony and Consumer Interpretation: Model 1 establishes a direct, system-agnostic relationship between interaction steps and distraction. This simplicity is vital for a consumer-facing rating, where a single, stable coefficient can be used to communicate the risk associated with each additional task step.

Model	Specification	Model Formula	DRT (REML)	DRT (r)	VOT (REML)	VOT (r)
Model 1	Random Intercept Only	Outcome $\sim$ nSteps + (1 Participant)	-474.3	0.710	3389.1	0.653*
Model 2	Interaction	Outcome $\sim$ nSteps + nSteps:IVIS + (1 Participant)	-425.1	0.725	3305.2	0.646
Model 3	Random Intercept & Slope	Outcome $\sim$ nSteps + (nSteps Participant)	-513.6	0.748	3366.1	0.633

*Model 1 was selected for all distraction metrics to ensure a standardized, system-agnostic consumer-facing rating.

36 Appendix 13 - Rationale for Establishing Core Task Domains

The purpose of the IVIS star-rating system is to provide a stable, meaningful, and generalisable measure of a vehicle's interaction demand. To achieve this, the rating cannot depend solely on which tasks happened to be available in each vehicle in the current sample. Instead, the assessment must be anchored in a set of **theoretically defined core task domains** that represent the typical ways drivers interact with modern infotainment systems.

Core tasks must therefore reflect:

- Major functional categories of IVIS use,
- Common real-world usage patterns,
- Known sources of visual and cognitive distraction, and
- Interface elements that manufacturers are expected to support in contemporary vehicles.

This ensures the rating system reflects **actual driver workload**, not artefacts of feature availability.

36.1 The Four Core Task Domains

Based on industry standards, human factors research, and driver-behaviour literature, four domains consistently emerge as the primary categories of in-vehicle interaction:

1. Phone Operations

Includes actions such as:

- Initiating a call
- Accepting an incoming call

These tasks require contact management, menu navigation, and confirmation actions, engaging both visual attention and cognitive decision-making. Phone use has been a central focus of distraction research for nearly two decades and remains one of the most frequent IVIS activities.

2. Messaging / Notifications

Includes:

- Receiving a text
- Accessing message playback
- (If supported) replying via voice or canned responses

Messaging represents a distinct cognitive task involving linguistic processing, working memory, and comprehension. These characteristics make it a potent distractor, even when voice interfaces are used. Its inclusion ensures the rating system assesses systems that manage interruptions and information retrieval.

3. Media Control

Includes:

- Selecting audio sources and,
- Browsing playlists, radio stations, or selecting podcasts

Media interactions are near-universal and impose varying degrees of visual search and menu navigation, making them important for assessing interface design and step complexity.

4. Navigation / Destination Entry

Includes:

- Entering an address,
- Searching for a point of interest, and
- Confirming a destination

In our study, navigation tasks were among the most visually and cognitively demanding IVIS functions. They require multi-step sequences, spatial reasoning, working memory, and visual verification. They also show large performance differences across systems and provide strong discrimination in the rating system.

36.2 Why Core Tasks Are Defined Theoretically (Not Empirically)

Although some vehicles in the current study did not support all tasks (e.g., messaging or navigation functions), feature availability varies by model year, trim, OEM design philosophy, and integration with CarPlay/Android Auto.

Using only tasks present in the current sample would:

- Make the rating system unstable across datasets,
- Produce different task sets for different vehicle cohorts,
- Undermine comparability over time, and
- Risk rewarding vehicles simply for lacking difficult tasks.

A theoretically grounded task set avoids these problems by ensuring:

- Continuity across years and vehicle generations
- Fairness across systems with different feature mixes
- Predictability and transparency for consumers and regulators
- Alignment with established HMI practice and safety guidelines

36.3 What Happens When a Vehicle Lacks a Core Task?

A vehicle may not support a particular feature (e.g., no native navigation, no messaging). From a distraction perspective, the absence of a task does not imply lower interaction burden; it simply means the driver could use alternative means (e.g., smartphone, CarPlay/Android Auto, or external navigation).

We suggest three options to solve this issue:

- Option 1. Assess the domain using the best available method (recommended if using CarPlay/AA)
 - Example: If a vehicle lacks native navigation but supports CarPlay, use the CarPlay navigation task.
- Option 2. Exclude the task from the step average but note the missing domain
 - This avoids penalising or rewarding absent features entirely.
- Option 3. Apply a penalty or minimum score for missing demanding tasks
 - This mirrors regulatory approaches (e.g., Euro NCAP penalises absence of safety-relevant functions)