A User Experience-Based Toolset for Automotive Human-Machine Interface Technology Development

Innovation Report

Matthew John Pitts MEng MSc

Submitted in partial fulfilment of the degree of Doctor of Engineering (Int.)

WMG, University of Warwick

Submitted October 2017

Revised June 2019

Abstract

The development of new automotive Human-Machine Interface (HMI) technologies must consider the competing and often conflicting demands of commercial value, User Experience (UX) and safety. Technology innovation offers manufacturers the opportunity to gain commercial advantage in a competitive and crowded marketplace, leading to an increase in the features and functionality available to the driver. User response to technology influences the perception of the brand as a whole, so it is important that in-vehicle systems provide a high-quality user experience. However, introducing new technologies into the car can also increase accident risk. The demands of usability and UX must therefore be balanced against the requirement for driver safety.

Adopting a technology-focused business strategy carries a degree of risk, as most innovations fail before they reach the market. Obtaining clear and relevant information on the UX and safety of new technologies early in their development can help to inform and support robust product development (PD) decision making, improving product outcomes. In order to achieve this, manufacturers need processes and tools to evaluate new technologies, providing customer-focused data to drive development.

This work details the development of an Evaluation Toolset for automotive HMI technologies encompassing safety-related functional metrics and UX measures. The Toolset consists of four elements: an evaluation protocol, based on methods identified from the Human Factors, UX and Sensory Science literature; a fixed-base driving simulator providing a context-rich, configurable evaluation environment, supporting both hardware and software-based technologies; a standardised simulation scenario providing a repeatable basis for technology evaluations, allowing comparisons across multiple technologies and studies; and a technology scorecard that collates and presents evaluation data to support PD decision making processes.

The Evaluation Toolset was applied in three technology evaluation case studies, conducted in conjunction with the industrial partner, Jaguar Land Rover. All three were live technology development projects, representing hardware and software concepts with different technology readiness levels. Case study 1 evaluated a software-based voice messaging system with reference to industry guidelines, confirming its performance and identifying potential UX improvements. Case study 2 compared three touchscreen technologies, identifying user preference and highlighting specific usability issues that would not have been found though analytical means. Case study 3 evaluated autostereoscopic 3D displays, assessing the effectiveness of 3D information while highlighting design considerations for long-term use and defining a design space for 3D content.

Findings from the case studies, along with learning from visits to research facilities in the UK and USA, was used to validate and improve the toolset, with recommendations made for implementation into the PD workflow. The driving simulator received significant upgrades as part of a new interdisciplinary research collaboration with the Department of Psychology to support future research activity.

Findings from the case studies also directly supported their respective technology development projects; the main outcome from the research therefore is an Evaluation Toolset that has been demonstrated, through application to live technology development projects, to provide valid, relevant and detailed information on holistic evaluations of early-phase HMI technologies, thus adding value to the PD process.

Acknowledgements

I would like to offer my sincere thanks to my supervisors, Professor Mark Williams and Dr. Alex Attridge for their guidance, support and patience over the course of the project. I would also like to thank Professor Derrick Watson for supporting and encouraging our collaboration on the driving simulator.

Thanks also to my mentors and colleagues at Jaguar Land Rover who have supported this work: Lee Skrypchuk, Dion Thomas, Elvir Hasedžić, Simon Thomson, Chris Mitchell, Jean-Jacques Loeillet and Martin Dale.

Most of all I would like to thank my wife Angela and our beautiful boys Joshua and Daniel. Life has changed a lot since this journey began and I'm looking forward to taking the next step together with you.

Matthew John Pitts

October 2017

Contents

List of Figures

List of Tables

Declaration

I hereby confirm that the research contained within this document is my own work, with due acknowledgement of collaborating parties given in the text; and has not been previously submitted for the award of another degree.

Matthew John Pitts

October 2017

Abbreviations

HMI	Human-Machine Interface
NPD	New Product Development
EPSRC	European Physical Sciences Research Council
JLR	Jaguar Land Rover
UX	User Experience
SAE	Society of Automotive Engineers
R&D	Research and Development
PD	Product Development
TCDS	Technology Creation Delivery System
TAM	Technology Acceptance Model
ISO	International Standards Organisation
UCD	User-Centred Design
GUI	Graphical User Interface
OEM	Original Equipment Manufacturer
JAMA	Japan Automobile Manufacturers Association
NHTSA	National Highway Traffic Safety Administration
AAM	Alliance of Automobile Manufacturers
IVIS	In-Vehicle Information System
LCT	Lane Change Test
GOMS	Goals, Operators, Methods & Selection

HTA	Hierarchical Task Analysis
CPA	Critical Path Analysis
SDLP	Standard Deviation of Lane Position
SYS	System Usability Scale
NASA-TLX	NASA Task Load Index
LCVTP	Low Carbon Vehicle Technology Project
TMETC	Tata Motors European Technical Centre
API	Application Programming Interface
ECU	Electronic Control Unit
HLDF	High Level Display Front
HID	Human Interface Device
SSQ	Simulator Sickness Questionnaire
BSREC	Biomedical Science Research Ethics Committee
TGT	Total Glance Time
ANOVA	Analysis of Variance
HSD	Honestly Significant Difference
CPD	Convergence Plane Distance
MIT	Massachusetts Institute of Technology
UMTRI	University of Michigan Transportation Research Institute
NADS	National Advanced Driving Simulator
TRL	Technology Readiness Level
DRT	Detection-Response Task

1. Introduction

The development of new automotive Human-Machine Interface (HMI) technologies offers manufacturers the opportunity to gain commercial advantage in a competitive and crowded marketplace, leading to an increase in the features and functionality available to the driver. User response to technology influences the perception of the brand as a whole, so it is important that in-vehicle systems provide a high-quality user experience. However, introducing new technologies into the car can also increase accident risk. The demands of usability and user experience must therefore be balanced against the requirement for driver safety.

Adopting a technology-focused business strategy carries a degree of risk, as most new products fail before they reach the market. One of the key factors for the success for new products is a robust New Product Development (NPD) process requiring clear information on product performance and customer perception. For the development of new in-vehicle technologies, it is critical that manufacturers are able to obtain clear and relevant information on user experience and safety to support development decisions.

This document describes the work conducted under an Engineering Doctorate project to create a toolset for the evaluation of early-stage automotive technology concepts. The project was funded by the European Physical Sciences Research Council (EPSRC) and supported by Jaguar Land Rover (JLR).

1.1. Research Question

The project focuses on the use of driving simulators to provide a context-rich environment in which novel technologies can be evaluated. The evaluation process must take into account the issues of user experience and driver safety. The research question adopted therefore is as follows:

> *"How can simulation tools be used to support advanced product development processes to deliver HMI solutions that optimise the user experience while minimizing safety risk?"*

1.2. Objectives and Research Design

The objectives for the project are:

1. Determine the key objective performance indicators for driver behaviour
2. Develop a context-rich environment to enable evaluation of novel HMI technologies

3. Identify and evaluate methods for capturing the User Experience of in-car technology
4. Create and test a protocol for the evaluation of technologies from the perspectives of User Experience and driver safety
5. Determine how the evaluation protocol can be integrated into the up-front NPD process for in-vehicle technology

The research design is illustrated in Figure 1. The research consists of three phases, described below:

- Phase 1 focuses on development and validation of evaluation tools, encompassing hardware and software tools along with data gathering and analysis methods. This includes development of the evaluation protocol and selection of appropriate methods, as outlined in chapter 10 of submission 2.
- Phase 2 consists of a series of technology case studies conducted on concepts from the JLR advanced product creation processes, representing technologies at different technology readiness levels and requiring different degrees of maturation and differing outputs.
- Phase 3 applies the learning from the case studies conducted within phase 2 to revise the toolset, which will then be validated and integrated into the NPD process.

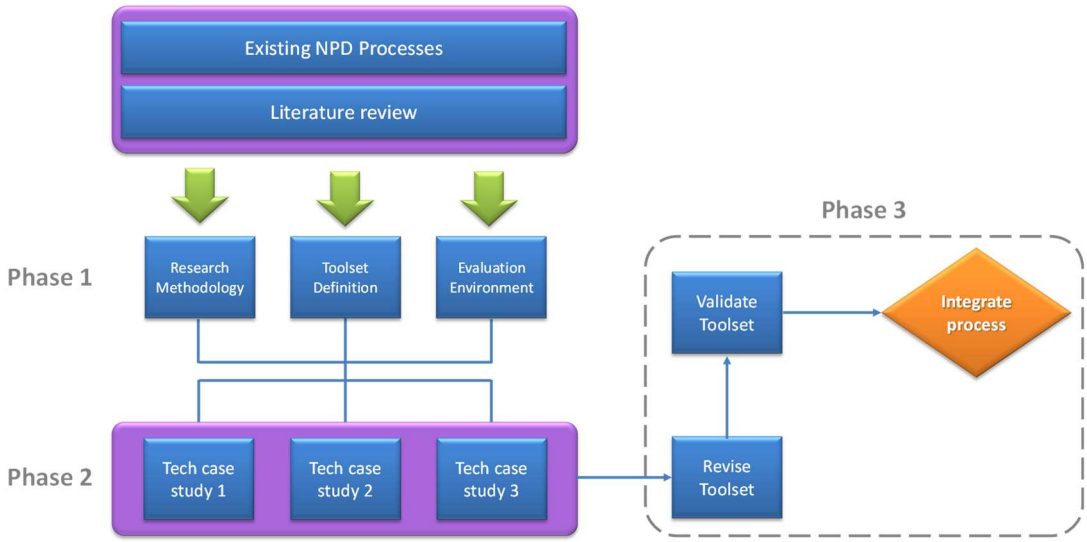


Figure 1 - Research design

1.3. Structure of Portfolio

The structure of the portfolio reflects the research design shown in Figure 1 above. Note that submissions 1, 5 and 6 were submitted to the portfolio at an earlier date and hence do not feature the revised project title.

Submission 1: Definition – this document provides the background to the project, including literature review in the areas of automotive HMI and driver safety, along with an overview of JLR's HMI development processes and an outline of the project methodology.

Submission 2: Toolset Definition – this document provides additional literature review in the areas of NPD and User Experience (UX), along with an updated methodology. This document also provides the basis for the creation of the Evaluation Toolset.

Submission 3: Evaluation Environment – this document details the development and testing of the driving simulator facility at the core of the Evaluation Toolset.

Submission 4: Case Study 1 – Voice Dictation – this is the first of three technology case studies that utilise the Evaluation Toolset, and focuses on a novel voice interaction interface concept.

Submission 5: Case Study 2 – Touchscreen Technology – the second case study involving the evaluation of three touchscreen technology concepts.

Submission 6: Case Study 3 – 3D Displays – the third and final case study focuses on the use of autostereoscopic 3D displays in the instrument cluster.

Innovation Report (this document) – provides an overall summary of the project, discusses the findings from the case studies and their application, and details the work from phase 3 of the project.

International Placement report – details a trip to the USA in 2015, visiting research establishments and presenting at SAE World Congress.

Personal Profile – details how the competencies required for the EngD have been satisfied.

1.4. Structure of Innovation Report

The Innovation Report includes an overview of the work from the submissions, along with a discussion of outcomes, application and impact. The document is structured as follows:

Chapter 1 gives an introduction to the Innovation Report and the project as a whole. **Chapters 2 to 4** present a summary of literature in the areas of Technology in the Automotive Industry, Customer Acceptance of Technology and Automotive Human-Machine Interface. Figure 2

illustrates how these topics relate to the project. **Chapter 5** provides the rationale for the development of the Evaluation Toolset, while **chapter 6** outlines the research methodology.

Chapter 7 details the definition of the Evaluation Toolset, with **chapters 8 and 9** describing the development of the evaluation protocol and driving simulator respectively.

Chapters 10 to 12 present summaries of the technology case studies, with the outcomes discussed in **chapter 13**.

Chapter 14 outlines how the Toolset could be applied to the NPD process, with **chapter 15** detailing the impact from the project. Finally, **chapter 16** presents the conclusions, limitations and further work.

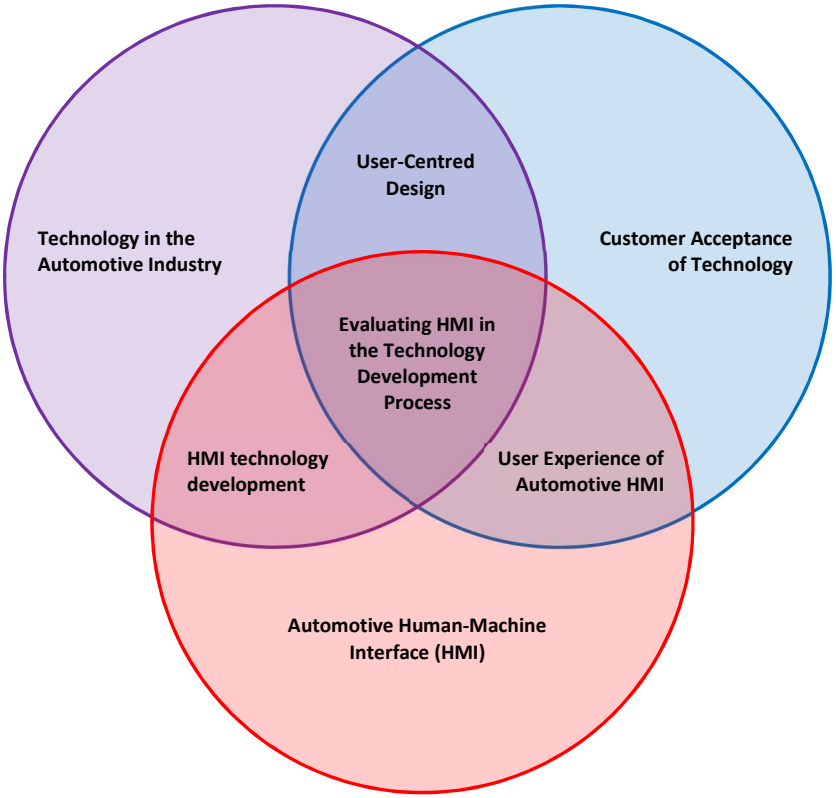


Figure 2 – Literature review diagram

2. Technology in the Automotive Industry

The effective application of technology is a critical factor in the success of modern organisations. Companies face challenges in how to manage new technologies in line with existing business practices (Probert et al., 2000); this is especially relevant in the automotive industry where new technologies can provide a significant commercial advantage. This chapter presents a summary of the literature detailed in submissions 1 and 2, providing an overview of technology strategy and innovation within the automotive industry, with a focus on the strategic importance of in-vehicle technology to customer satisfaction.

Successful technology management approaches require fluent communication between the technical and business aspects of the company in order to ensure that both technology and market-driver opportunities are realised. This in turn requires a clear understanding of both internal factors, such as business aims and culture; and external factors, such as market conditions (Phaal et al., 2013). Within the automotive industry, continuing introduction of new technologies is a vital means of differentiating products in a crowded and competitive market (Maniak et al., 2014). Global automotive market trends show that technologies, particularly Advanced Driver Assistance Systems (ADAS) and connectivity features, are highly sought after by potential vehicle purchasers and contribute strongly to purchase intention (IHS Markit, 2016). It is vital therefore that automotive manufacturers employ robust product innovation management processes to ensure successful product outcomes.

2.1. Technology Innovation in the Automotive Industry

The automotive industry is highly competitive and highly technical, with slim profit margins and significant Research and Development (R&D) spend requirements (Darveight, 2010). Companies wishing to attain or maintain a market-leading position in a high-technology field must be willing to commit significant levels of funding to R&D activities (Trott, 2012), requiring that the technology development strategy receives support from the highest levels within the organisation. The Association des Constructeurs Européens d'Automobiles (ACEA) states that €53.8bn was invested in R&D within the EU in 2016, making automotive the third-ranked sector by R&D spend (ACEA, 2018). An innovation-targeted strategy is not without risk: a study by management consultancy Oliver Wyman (Oliver Wyman, 2007) found that only 20% of automotive R&D investment results in profitable innovation with 40% either failing to reach market or failing to gain customer acceptance, 20% allocated to serial development and the remaining 20% required to meet legal requirements. A failure to understand customer requirements and to place emphasis on marketing innovations was found to contribute to failure in the marketplace, with only one in six innovations offered at the point

of sale being accepted by the customer, as shown in Figure 3. Of the profitable 20% of innovations, only half of these have the potential to create a significant change in the market, and it is these technologies which carry the largest innovation risk.

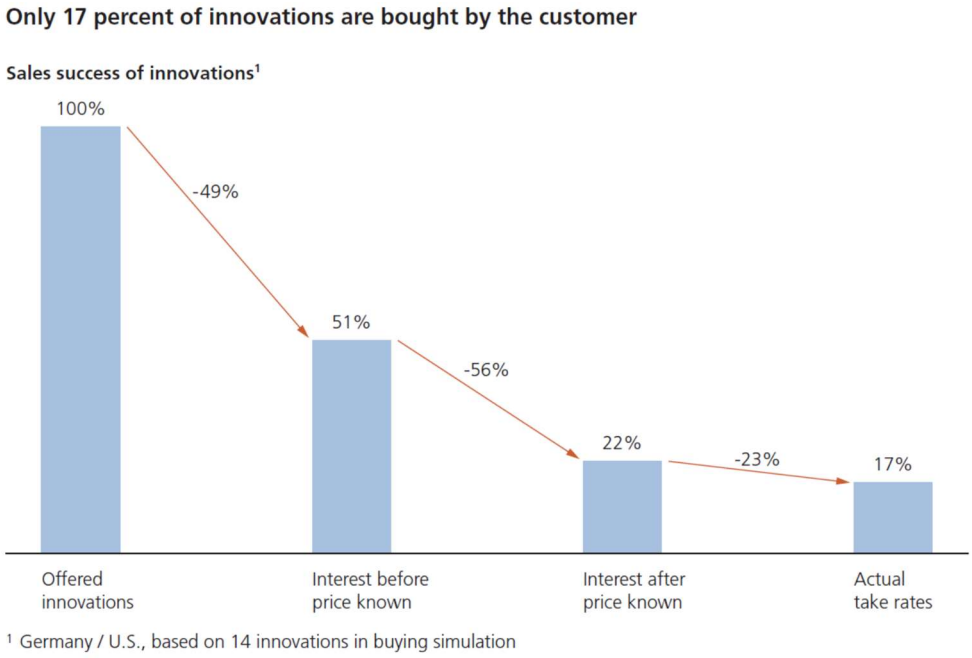


Figure 3 - Uptake of innovations in the automotive industry (Oliver Wyman, 2007)

Disconnects between manufacturers and customers are acknowledged as a limiting factor in developing successful products (Trott, 2012) and alignment with the needs of the customer is therefore vital to drive user acceptance and ensure the success of the innovation in the marketplace. In order to achieve this, the 'voice of the customer' must be integrated into the new product development process(Cooper, 2011, 1998, 1996), requiring the application of appropriate tools and processes; however, existing tools to enable this approach are often under-utilised (Yeh et al., 2010).

2.2. Automotive Product Development Processes

This section discusses product development processes as they relate to the development of automotive products and technologies. Examples are based on generic processes from the literature, with reference to JLR processes. Specifics of these processes are discussed in submission 1, chapter 5 and submission 2, section 3.2. The structure of the various HMI development capabilities are discussed in submission 1, section 5.3.1.

2.2.1. Mainstream Product Development Processes

JLR's main new product introduction process follows a 'stage-gate' methodology, featuring a series of decision points along the development path ('gates') at which the progress of the project is assessed against defined criteria. Gates also function as quality control points for the product development process, allowing underperforming projects to be terminated before excessive resources are committed, thus maintaining the quality of the project portfolio (Cooper et al., 2002b, Cooper, 2011). However, accurate and relevant information must be available to gateway decision makers in order to ensure that the correct conclusions are reached.

Successful product development processes feature a pre-development 'discovery' phase (Cooper et al., 2002a, Cooper, 2011). An example of a generic stage-gate model with a pre-development phase is shown in Figure 4.

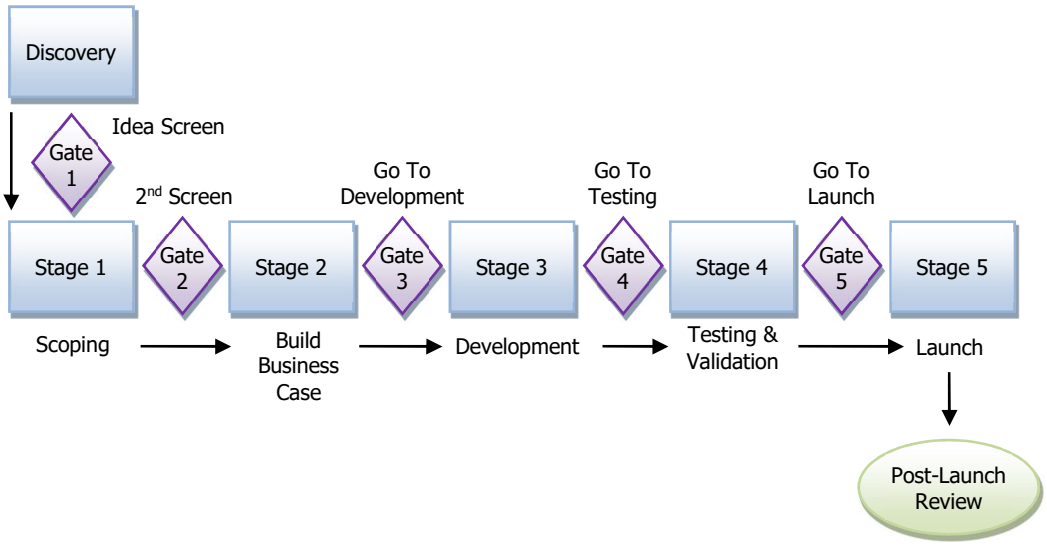


Figure 4 - Generic stage-gate product development model with pre-development phase (adapted from Cooper et al., 2002a)

The purpose of this phase is to translate high-level strategic plans into a concept plan and objectives that can be taken forward into development by the programme teams. The checks and reviews which occur in the pre-development phase serve to ensure that only the most viable ideas are taken forward into the mainstream development process (Cooper, 1988, 2011).

Stage-gate processes support the creation of successful products by placing emphasis on the work conducted in the planning and development phases. This allows product requirements to be clearly defined and aligned with market requirements, improving customer satisfaction. Furthermore, by clearly defining the criteria required for a project to pass a gateway, decisions

on the viability of a project can be made more objectively and from a business-wide viewpoint, reducing development time and increasing profitability (Schilling, 2010).

Next-generation Idea-to-Launch systems integrate 'Agile' project management practices to create an Agile Stage-Gate Hybrid model (Cooper, 2016, 2014). A key characteristic of this model is multiple iterative 'spiral' development paths that occur throughout the development timescale, as shown in Figure 4. This creates multiple points of contact with the 'customer' whereby feedback is used to refine and improve the product definition. By initiating this activity early in development, technical issues can be identified up-front and the cost of making changes to the product is minimised (Cooper, 2016). Furthermore, this approach overcomes one of the key criticisms of traditional stage-gate processes: that they are too rigid and do not manage uncertainty well (Cooper, 2017). The Agile Stage-Gate Hybrid model allows projects to progress even when project requirements are incomplete or unclear, relying on the experimental nature of the development loops to add definition (Cooper, 2016).

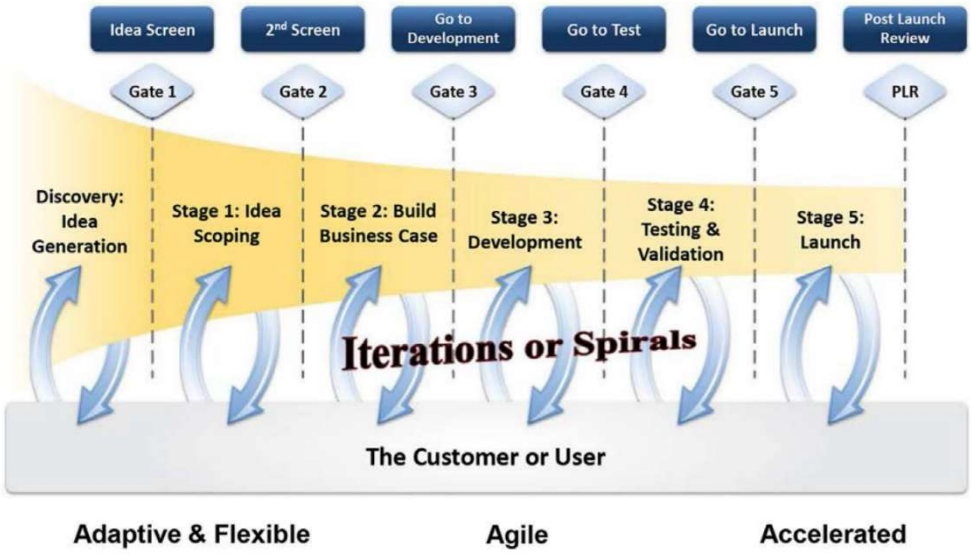


Figure 5 – Next-generation idea-to-launch process (Cooper, 2014)

2.2.2. 'Up-Front' Technology Development Processes

As product life cycles become shorter and the market becomes more competitive, the need to quickly and successfully integrate new technologies into innovative products and processes that meet the demands of customers becomes increasingly apparent. The development of innovative products is however characterised by high levels of uncertainty, especially in the early stages (Jahn and Binz, 2007). The inherent risk in NPD can be mitigated by concentration of development effort at the start of the process, known as 'front-loading'. Thomke & Fujimoto (2000) define front-loading problem solving as "a strategy that seeks to improve development performance by shifting the identification and solving of [design] problems to earlier phases of a product development process". Up-front NPD activities include thorough

competitive and market analysis, assessments of in-house capability and project feasibility, research into customer wants and needs, and financial analysis (Bhuiyan, 2011). The fuzzy front end of the NPD process can require up to 50% of the total product development time. It is in this phase that the scope of the project is set, with commitments to major resource allocations (Morse and Babcock, 2010)

The risks associated with the development of high-technology products can be further mitigated by employing a strategy of 'up-front' technology development. This encompasses work done to create, evaluate and mature specific technologies prior to their introduction to the main product development process, ranging from new idea generation to the point at which the idea is either moved forward into the main product development process, or development is ceased (Murphy and Kumar, 1997). This helps to ensure that a large proportion of problems are identified and resolved before mainstream development is initiated.

Figure 6 shows a typical stage-gate NPD process with an up-front technology development path (Cooper et al., 2002a). The up-front 'discovery' phase takes the form of a separate, offline development process with its own stages and gateways, providing a development framework which allows technical feasibility and market opportunity to be robustly assessed. If the project meets the pre-determined success criteria at the application path gate it is then fed into the main stage-gate process at the appropriate point, as determined by the relative maturity of the technology.

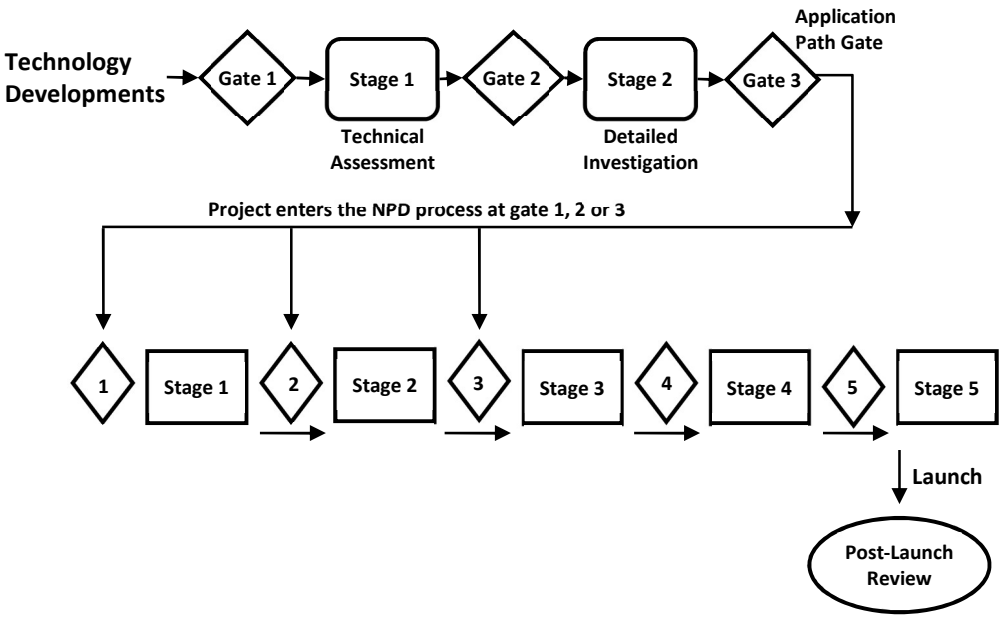


Figure 6 - Offline technology development in the NPD process (Cooper et al., 2002a)

In the case of automotive Product Development (PD) processes, this is the phase in which innovative components and systems are developed prior to allocation and integration into a vehicle programme (Jahn and Binz, 2007). An example may be a new driver assistance system, which will have dedicated hardware and software but must also be integrated into the existing vehicle systems. The up-front technology development process allows the core functionality of the new technology to be matured to the extent whereby feasibility is established, before expending additional development resource in integration and final maturation. Technology development within the automotive industry is discussed further in submission 2, section 2.2.

JLR utilise two types of advanced product creation processes: Milestone and Technology Creation Delivery System (TCDS), which are summarised below; see submission 1 section 5.3 and submission 2 section 3.2 for further details. Milestone projects have a long-term outlook, being used to develop knowledge and technologies which may be applied to production vehicles in the medium- to long-term. A project with a successful outcome may be passed on to mainstream development teams who will mature the technology for inclusion on a vehicle programme.

TCDS is a formal advanced development process, which exists in parallel to the mainstream PD process. It exists to facilitate rapid maturation of technologies which can then be transferred directly onto vehicle development programmes on the mainstream process. The main aim of TCDS is to reduce failure rates by early identification and rectification of failure modes, as illustrated in Figure 7. Potential issues which are identified in the research phase at the beginning of the development process are easier to rectify, although may be more difficult to identify. Conversely, once the product has reached the manufacturing phase its faults become clearer, but significantly harder and more costly to fix. It is clearly beneficial to fix problems at the earliest stages of development, which is aided by the application of a robust advanced development process.

As a strategy, TCDS helps to mitigate these risks by focusing effort on the early stages of development, placing emphasis on product definition and technical feasibility. One of the key benefits of this approach is that it prevents allocation of development resource to technology projects which have limited change of success. Making robust product concept selection decisions helps to ensure that PD effort is directed towards candidate technologies that have the best chance of market success. However, given the emotive nature of these decisions (Trott, 2012) selection must be supported by user-derived usability and UX data in order to justify the decision and to maximise the potential benefits of the technology in terms of customer satisfaction.

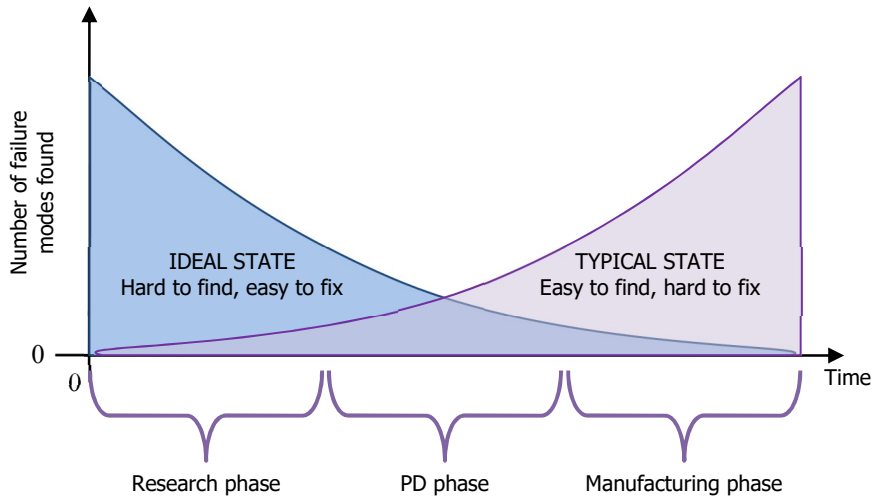


Figure 7 - Managing failure modes through TCDS

2.3. Summary

The introduction of new technologies to the market carries significant risk, but is necessary to achieve and maintain competitive advantage, especially in high-technology industries such as automotive. With recent developments in driver assistance systems, low-carbon drivetrains and connectivity, the pressure on vehicle manufacturers to introduce innovative technologies, both in terms of technology push and customer demands (market pull), has never been greater.

The constant demand for new technologies places significant emphasis on product development processes. The key aim of any product development process is to apply rigour to the activities associated with bringing a product concept to market. Stage-gate processes provide a series of checkpoints at which the feasibility of the project can be assessed, based on criteria determined at earlier stages of the project. In order to be effective however, the gate decisions must be a) based on valid data; and b) enforced objectively and rigorously.

Up-front technology development processes apply stage-gate methodology to the development of new technology concepts, allowing feasibility assessment and rapid maturation to occur offline from the main PD process. The aim here is to minimise the risk associated with the introduction of new technologies before they are integrated into a final product. The application of user-derived data in up-front development activities can ensure that the technology is correctly aligned with customer needs and that major UX pitfalls are avoided. This requires that there are processes in place for gathering relevant customer data in order to support the decision-making process.

3. Customer Acceptance of Technology

A user's intention to use a new technology is governed by how its intrinsic attributes are judged by the user, highlighting the influence of the individual on the outcome of the interaction. This chapter introduces and discusses fundamental concepts in the formation of product quality judgements, including usability, mental workload and User Experience.

The Technology Acceptance Model (TAM) (Davis, 1989), illustrated in Figure 8, shows that a user's intention to use a system is ultimately determined by the user's perception of the ease-of-use and usefulness of the system under test. Note that Perceived Usefulness is in itself influenced by Perceived Ease of Use; the user's view ability of the system to successfully achieve its designed function is in part determined by how easy the user finds task completion. The definition and evaluation of usability is discussed in section 3.1.

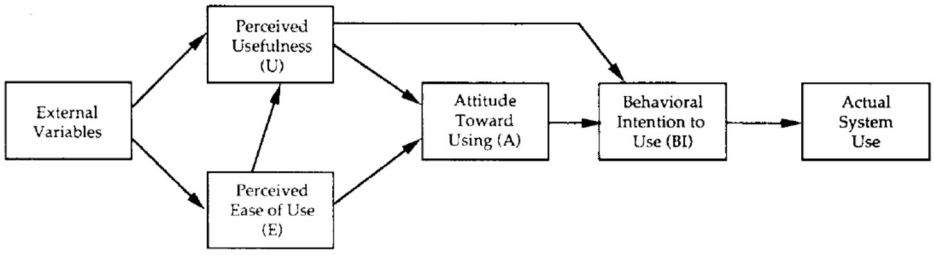


Figure 8 - Technology Acceptance Model (Davis et al., 1989)

Venkatesh (2000) proposed an expanded model that outlined the determinants of perceived ease-of-use, in order to integrate the concepts of control, intrinsic motivation and emotion into TAM (Figure 9). This featured two sets of influencing factors: Anchors and Adjustments. Anchors relate to the user's self-perceptions surrounding their relationship with technology, including their personal anxiety with using computers and their perceptions of external control. Adjustments included perceived enjoyment: the extent to which the act of using the system is enjoyable in its own right, regardless of the outcome; and objective usability: a measure of the actual (rather than perceived) effort required to complete a task. Objective usability is related to the concept of workload, discussed in section 4.3.1.

It is clear from these models that the expectations, experiences and biases of the user shape their view of the technology under consideration, along with their 'perceived enjoyment', an affective construct, thus clearly establishing a link between a positive product experience and intention to use. These concepts are reflected in the User Experience literature, discussed in section 3.2.

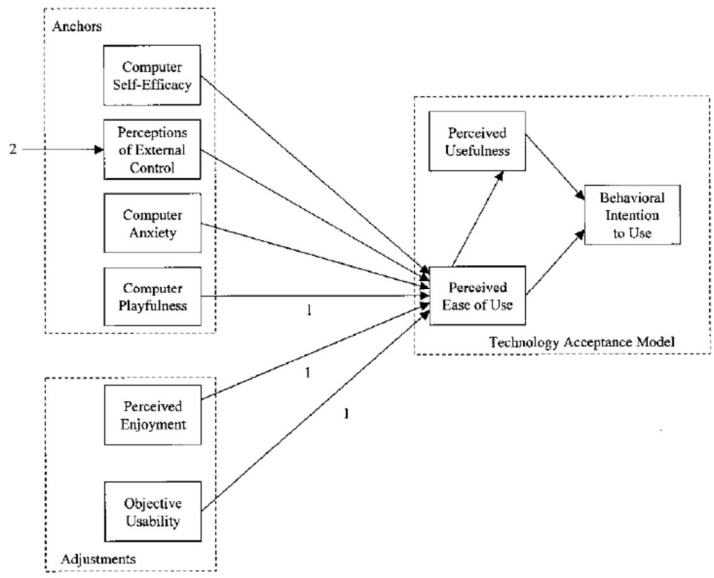


Figure 9 - Model of Determinants of Perceived Ease of Use (Venkatesh, 2000)

3.1. Usability

The importance of designing and testing for usability when developing interactive products is well established. Products with good usability can help users reach their goals more effectively and with greater satisfaction, while poor usability can seriously harm the overall perception of a product. Good usability is a key factor is ensuring strong market performance and competitive advantage.

The study of usability originates from the field of Human-Computer Interaction, with the term first used in relation to visual displays (van Kuijk et al., 2006). The traditional view of usability focusses largely on the functional aspects of product interaction, as reflected in the definition in ISO 9241-11 (ISO, 1998):

> *"The extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use"*

The ISO usability framework reflects the User-Centred Design (UCD) framework highlighted in section 3.3 in that it emphasises the importance of the context of use to both the goals and outcome of interaction with a product. Clearly, it is important that the context of use is reflected in usability evaluation. At the highest level, usability studies can be either *summative* or *formative* (Tullis and Albert, 2008, Lewis, 2012).

The aim of formative usability studies is to improve design and as such, an iterative approach is adopted whereby design flaws identified through testing are remedied and the product re-tested. This clearly requires that formative usability testing must occur before the design is

finalised, in order to allow the design flexibility required by the process. It is important therefore to recognise where there is no scope to improve a design before instigating in-depth formative studies, such as not to waste time and resources (Tullis and Albert, 2008).

Summative usability studies are used to evaluate a design once it is finalised and to determine the extent to which it meets its design objectives. This may include evaluating the product relative to previous versions in order to determine the degree of improvement; and relative to current competitors. Summative usability is focused on the measurement of task performance and the completion of task goals (Sauro and Lewis, 2012).

Usability, as a construct, cannot be measured directly; however different aspects of what comprises usability can be measured (Hornbæk, 2006). The ISO framework does not however specify which usability measures should be used in a given situation, leading to a reliance on guidance from the literature. Sections 5.2 to 5.5 of submission 2 detail a range of usability evaluation methods in four categories: instrumental methods, analytical methods, questionnaire-based methods and interviews. Instrumental methods involve the objective measurement of task performance metrics, such as task completion time and error rate. Questionnaire-based methods rely on subjective self-reporting by the user to obtain a qualitative measure of usability. Analytical methods use models of task performance to calculate completion times and critical paths, generally requiring expert application. Interviews can be used to obtain rich, qualitative data about usability issues that can provide significant insight, but requires additional effort to interpret.

The appropriate methods to apply will depend upon the technology under test and its level of development. While analytical methods are less costly to administer and can be applied to very early-stage prototypes with minimal functionality, they lack the ability to capture the subjective nature of the user experience, potentially missing key usability issues that may only become apparent though actual use. The use of experimental methods, supplemented with qualitative data captured though post-evaluation interviews can ensure that a comprehensive picture of critical usability issues is obtained.

3.2. User Experience

Delivering products that offer a high-quality experience for the user requires a focus on the needs of the user throughout the product development process. The question raised therefore is: 'how can a high-quality user experience be consistently delivered?' Establishing processes and methodologies to address this question requires an understanding of the key drivers of UX, approaches to designing for UX, how companies select technologies for development and what makes specific technologies acceptable to the user. This section explores the foundations of UX, outlining definitions, key models and design approaches to delivering products with high levels of customer satisfaction.

3.2.1. Definition of User Experience

Multiple definitions of UX exist, as discussed in submission 2, section 4.1. From these, a number of key themes can be identified:

- UX is determined by the user's total experience of a product, service or company, from before they interact with the product to beyond the point at which the user's goal is reached
- UX is subjective and determined by the perceived quality of a user's interaction with a product, service or company
- UX is shaped by the personal beliefs, expectations, experiences and biases of the user
- UX is based on an assessment of the functional attributes of the product (e.g. how well it works, how easy it is to use), combined with the user's emotional response to the product

This holistic approach to understanding users' interactions with products represents a significant shift from a largely functional view of usability, usefulness and satisfaction. UX encompasses all aspects of the end-user's interaction with the company, its services, and its products, from marketing through to after-sales service (Nielsen and Norman, n.d.). The user's evaluation of UX is influenced by the entirety of this interaction, beyond the bounds of physical interaction with a product, framed by user's prior expectations of the experience. These expectations may in turn be formed through prior interaction with a product or brand; the user's mental model of what a particular product should be is determined by their response to previous product interactions, both positive and negative.

The degree to which a user enjoys using a new product can also have a significant impact on their continued use intentions. Igbaria et al. (1994) showed that identified that both acceptance and satisfaction in a new software product increased when users enjoyed using it. Indeed, the

study found that the effect of enjoyment as perceived by the user was of a similar magnitude to that of perceived usefulness.

The Kano model, discussed in section 0, illustrates how user expectation is key to UX. For a 'Must-be' attribute, whereby the user expects a certain minimum level of performance, a failure to adequately deliver is seriously detrimental to the overall perception of product quality. Conversely, by identifying and leveraging 'Attractive' attributes, which may not be needs voiced by the user, it is possible to create a significant positive effect on UX.

3.2.2. Models of User Experience

As definitions of UX vary, there are a number of models that seek to portray the relationship between the customer, the product and the experience. Submission 2, section 4.2 details four key models: firstly, the Nielsen-Norman Group model of User Experience (nnGroup 2008; in Nilsson, 2010), shown in Figure 10, depicts how the overall UX extends from the fundamental ability of the product to meet the needs of the user, through their sensory and affective experience of the product, ultimately shaping their judgements about both the product and the brand. The model effectively demonstrates the holistic view of interaction with a product offering that underpins UX, encompassing factors beyond the functional usability of the product and integrating the emotional response of the user.

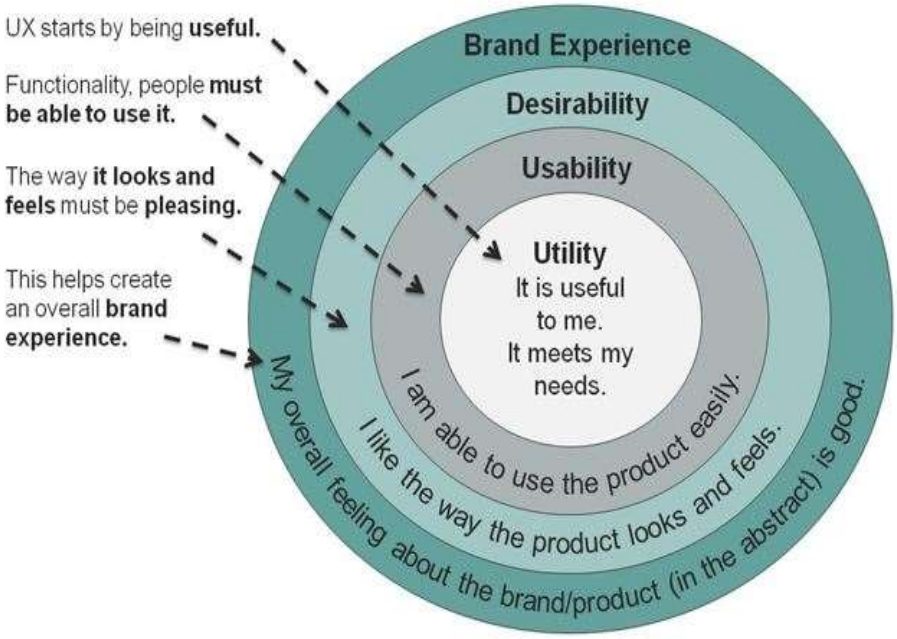


Figure 10 - Usability vs. UX (nnGroup 2008; image from Nilsson, 2010)

Secondly, Hassenzahl's Model of Ergonomic and Hedonic Quality (Hassenzahl, 2001) demonstrates how perceptions of the functional (ergonomic quality) and emotional (hedonic quality) combine to form judgements of appealingness, driving subsequent behavioural and emotional consequences. Hassenzahl's later Model of User Experience (Hassenzahl, 2004b), shown in Figure 11, elaborates upon the differing perspective of the designer and user, and how it relates to the intended characteristics of the product. The model highlights the importance of context in product evaluation, showing that the circumstances in which a user interacts with a product may strongly influence the outcomes.

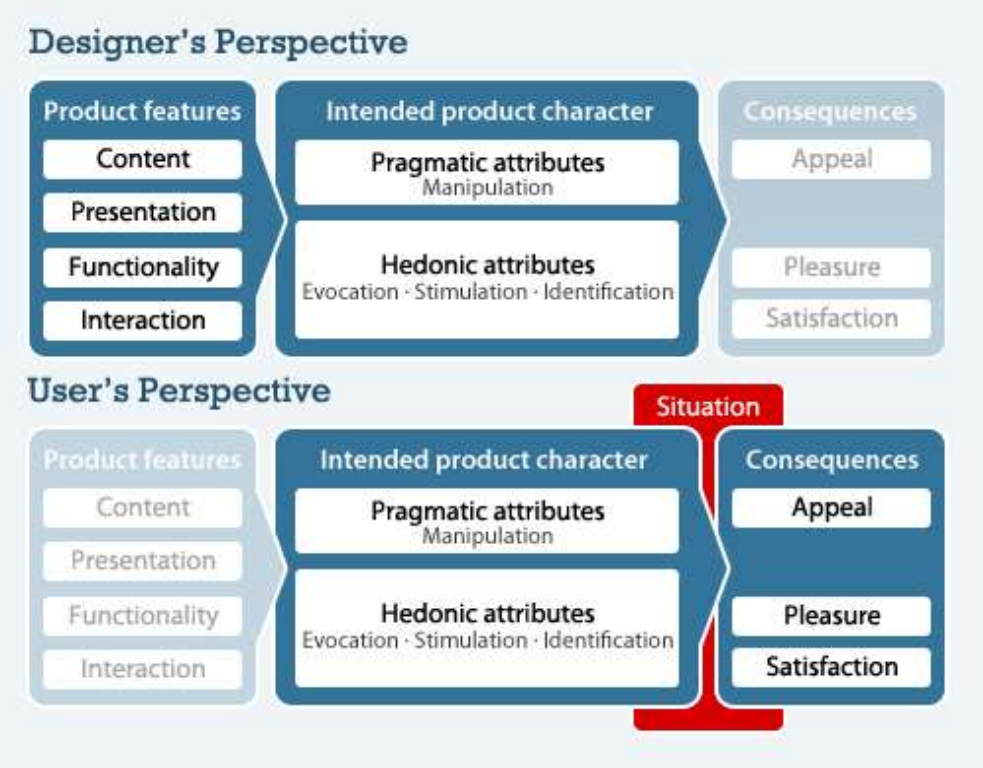


Figure 11 - Hassenzahl's model of User Experience (Hassenzahl, 2004b, image from Fredheim, 2011)

Finally, Desmet and Hekkert's Framework of Product Experience (Desmet and Hekkert, 2007) describe how a user's emotional experience of a product is driven by their sensory interactions and the meaning that they attach to the product, based on their interpretations and associations. As with Hassenzahl's model, both the situation and the individual are key to the emotional outcome: the same stimulus (i.e. product interaction) may elicit a different emotional response in an alternate situation. Therefore, it is essential to carefully consider the context of interaction when attempting to understand the UX of a given product.

3.3. Designing for User Experience

A focus on UX necessitates a User-Centred Design (UCD) approach, representing a shift away from the traditional, functional view of product quality to one that views the end-user's perception as the main determinant of quality. Consequently, processes to design for UX also require consideration of these factors. It is important therefore that product development is led by an understanding of the user and their needs, such that the attributes of the product can be tailored to ensure that these needs will be met. UCD is discussed in detail in submission 2 section 4.3 and summarised below.

Vrendenburg et al. (2001) outlined the key differences between a traditional and a UCD process, illustrated in Figure 12. The clearest difference between the two approaches is extent to which customers are involved in the process: a traditional PD approach is centred on the functional attributes of the product, with the involvement of customers limited to the collection of requirements at the beginning of the project. This tends to lead to a focus on the technology under development rather than the specific needs of the user.

In contrast, the UCD approach has the wants and needs of the customer at its core, with a consistent focus on the customer throughout the development process. Ideally, this focus on the user extends beyond the point of manufacture of the product into all aspects of the product offering, including marketing and aftersales service. It is important however that the focus on the user also extends beyond the existing customer base to encompass the target market in its entirety. A UCD approach necessitates that the intended product character is designed first with the required product functionality subsequently defined.

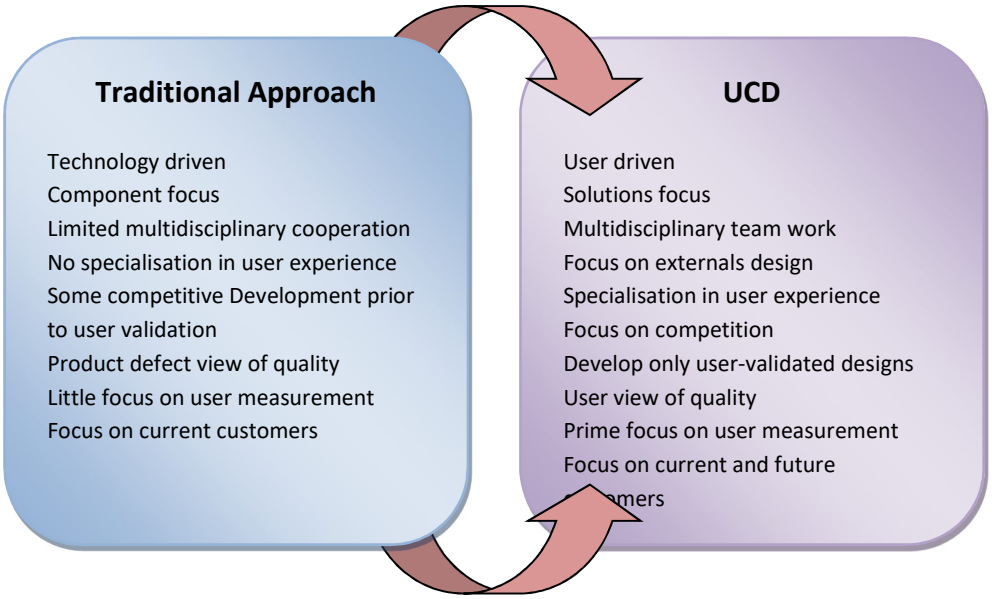


Figure 12 - Contrasting the traditional approach to design with UCD (Vredenburg et al., 2001)

A UCD approach also determines a shift in the quality paradigm. In a functional approach, quality can be seen as the lack of technical defects that affect the reliability of the product. A focus on quality defined in these terms however neglects problems encountered by the user. In contrast, a UCD-based definition views the quality of the user experience as the prime determinant of product quality. This ensures both that the product meets the requirements of the user, and that a suitable level of product reliability is attained (Vredenburg et al., 2001). Furthermore, the involvement of users throughout the PD process allows the product to be validated at various stages of development, ensuring that at launch the final product is fully aligned with user requirements; a lack of user validation within the PD process can lead to problems only being discovered once the product is in the marketplace, increasing the cost and complexity of resolving them (Cooper, 2011, Owens and Davies, 2000).

ISO 9241-210 (ISO, 2010a) provides a generic framework for the application of UCD, which can be tailored to specific products, illustrated in Figure 13 below.

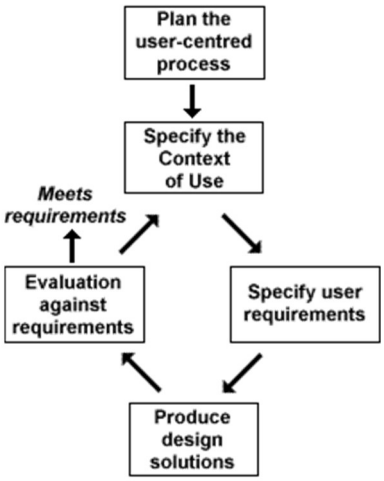


Figure 13 - ISO 9241-210 User-Centred Design process (ISO, 2010a)

The first step within this process is to specify the context of use; indeed, the process as a whole flows from the establishment of the context of use, acknowledging the extent to which this influences identification of goals and subsequent judgement of product quality (Bevan, 1995). Clearly, the context of use must also be considered in the evaluation phase, such that user feedback accurately reflects the conditions in which the product will be used once in the market.

Successful application of a UCD approach is dependent on obtaining high-quality data from perspective users, thus requiring appropriate evaluation methods. While traditional usability approaches often focus on instrumental measurements of the functional attributes of the product (e.g. task completion time, error rate), information derived from UCD evaluations is likely to be subjective in nature, requiring a range of qualitative and quantitative methods to

build a full picture of the quality of interaction in order to inform and support the development process.

3.4. Summary

The processes that drive customers' acceptance of new technologies are complex and multi-faceted, incorporating both intrinsic and extrinsic product attributes along with personal beliefs, goals and motivations. The literature discussed in this chapter identifies how the user's perceptions of a product are built from core capability in terms of product utility and usability: products that are effective and easy to use will be perceived more positively; conversely, products that are difficult to use are likely to be perceived more negatively. User Experience, however, moves beyond this purely functional view, encompassing emotional responses and expectations based on prior experience.

In order to understand UX therefore it is necessary to address both functional and hedonic aspects of the product interaction. It is important to note that the perspective of the designer is often quite different from that of the user: the former can influence the character of the product but the consequences experienced by the latter are determined by situation of the interaction. Hence the situation, i.e. the context of use, must be reflected in evaluating the quality of the experience.

This broader view of product quality prescribes a revised approach to product design to one that views the end-user's perception as the main determinant of quality. User-Centred Design practices utilise iterative evaluations of product concepts to ensure that the 'voice of the customer' is reflected in the product from the earliest stages. Key to this approach is an effective methodology for gathering contextual, customer-derived data.

The implication therefore is that the design and manufacturing processes for in-vehicle technologies must consider the quality of the user experience as a success factor, alongside traditional product performance metrics. Consequently, there must be an effective mechanism by which elements of UX can be evaluated, and that information fed back into the product development process to improve the product.

4. Automotive Human Machine Interface

The Human Machine Interface (HMI) can be defined simply as "the point at which the user and the technology meet" (IEC, 2007). The interface facilitates the flow of information from the user to the system under control and vice versa (ISO, 2006). While often used to refer to graphical user interfaces (GUIs), the term applies in general to the medium by which a user interacts with a system, tangible or intangible, encompassing the means through which tasks are accomplished when using a product (Raskin, 2000).

In the context of the automobile, HMI can be considered to be the controls and displays required to control the speed and direction of the vehicle, operate secondary functions such as the turn signals and windscreen wipers; and to interact with information and entertainment functions which may improve the driver's experience, but do not influence control of the vehicle (Tönnis et al., 2006). However, the paradigm for primary vehicle control has changed little over the past 100 years (Frawley, 2012), with all modern vehicles using variations of the ubiquitous pedal/steering wheel arrangement. HMI is therefore currently used more readily to refer to the interface with technology within the vehicle: the secondary and tertiary systems with which the driver interacts in the course of driving but which are not the core means of affecting the primary driving task.

As the levels of in-vehicle technology and its associated information increase, functionality has become more centralised, characterised by a main multi-function display with a corresponding multi-function control mechanism to facilitate inputs. More recently, multi-function displays have become common in the dash cluster, with the trend moving towards configurable displays replacing traditional physical instruments. These developments increase the flexibility of information display, but also have the potential to increase driver distraction.

4.1. Strategic Importance of In-Vehicle Technology & HMI

In-vehicle technology represents a specific priority for vehicle manufacturers, especially in the premium sector (Güttner and Sommer-Dittrich, 2008). With relative parity across the market for functional attributes such as build quality or fuel economy, technology can be the factor that drives a customer's purchase (Heaps, 2010). Recent market data shows that in vehicle-technology, specifically ADAS and connectivity features, strongly contribute to purchase intention (IHS Markit, 2016). This desire for technology features is spreading from established to emerging markets and is expected to grow in lower-priced segments as well as premium (Robbins, 2012).

Within the domain of in-vehicle technology, the infotainment system, comprising audio, communications and navigation functionality, holds particular significance. Bob Joyce, Group

Engineering Director at Jaguar Land Rover, identifies infotainment as a critical area of focus for research and development activity, highlighting a recent significant expansion in capability (Hibbert, 2013). The complexity of modern infotainment systems is outlined by the fact that current generation vehicles feature up to 10 million lines of code in infotainment-related systems, double the amount required for engine management functions. Joyce states that the importance of in-vehicle technology will continue to increase:

> *"Cars have become more connected but, trust me, we are only just on the wave of the next generation. Touchscreens, phones, the cloud, telematics and the connected world are going to transform the cars we drive"*

This trend is reflected in the emergence of mobile device integration solutions. Google Auto & Apple CarPlay were estimated to be available in 46% of new vehicles in Europe and 52% in the US in 2018 (Canalys, 2018). These solutions offer integration of smartphone capability through an installed in-vehicle display, allowing access to app functionality while constraining the extent of user interaction to reduce driver distraction. Both systems incorporate cloud-based 'smart assistant' voice interaction (Apple Siri & Google Assistant) bringing naturalistic voice interaction to the car and reducing visual-manual distraction.

Advances in naturalistic interaction are furthered with the advent of gesture recognition, providing non-contact interaction with the aim of improving ergonomics and reducing eyes-off-road time. VW's 'Discover Pro' infotainment system (Volkswagen, 2016), offers gesture control in addition to a centre-console touchscreen. Mercedes' 'MBUX Interior Assistant' provides voice interaction, gesture control and augmented reality navigation (Mercedes-Benz, 2018). BMW's 'Natural Interaction' system combines gesture control, voice command and gaze recognition to provide multimodal interaction options, and is expected to be available in 2021 (BMW, 2019).

The advent of autonomous vehicles will see a seismic shift in the way in which we interact with our vehicles, with the focus moving away from the primary driving task and towards non-driving activities; and the role of the driver shifting from control to monitoring. This change brings many challenges in terms of HMI, with the nature of interaction with the vehicle varying significantly from the current norm. Key issues of interest include how to manage driver re-engagement (Cunningham and Regan, 2015), what types of information should be displayed (and how) (Debernard et al., 2016); and how to build trust in automated systems (Naujoks et al., 2019).

4.2. Customer Satisfaction and HMI

It is clear that in-vehicle technology is highly important to both the customer and the manufacturer. However, it is necessary to consider how users' interactions with the technology will affect their perceptions of the product. Kano's model of Attractive Quality ('The Kano Model') (CQM, 1993, Kano et al., 1984) describes the relationship between the delivery of product attributes and user satisfaction based on the nature of the attribute, and is shown in Figure 14. Attributes are viewed as belonging to one of three categories: 'attractive quality', 'one-dimensional quality' or 'must-be quality', each having a different function of delivery to satisfaction.

'One-dimensional' quality attributes have a linear relationship between delivery and satisfaction: the more the attribute is delivered, the greater the satisfaction. An automotive example would be fuel economy: more is always better, and less is always worse. One-dimensional attributes can produce incremental improvements or decrements and do not support step-changes in customer satisfaction.

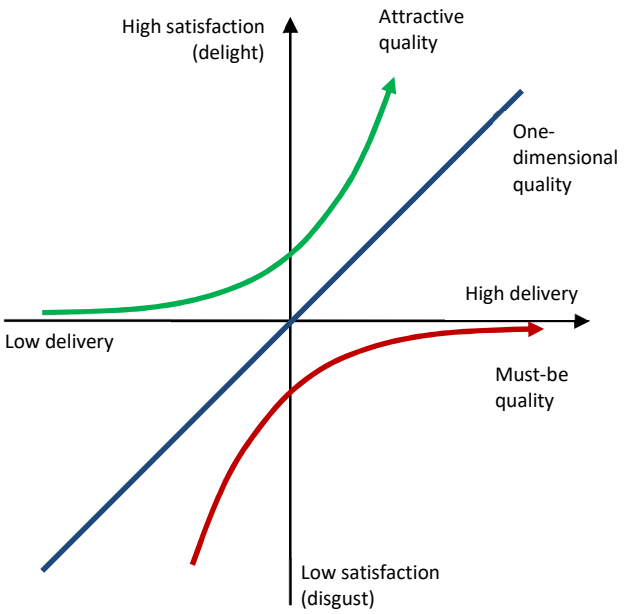


Figure 14 - Kano's model of Attractive Quality

'Must-be' quality attributes have a non-linear relationship between delivery and satisfaction, such that a lack of delivery causes a significant decrease in satisfaction. An automotive example may simply be basic reliability, which is largely taken for granted in modern vehicles. These attributes represent features that the customer expects, and as such significant increases in delivery will not produce increases in satisfaction.

‘Attractive’ attributes, also known as ‘delight features’, represent the elements of the product that go beyond the customer’s expectations of quality; as such, a lack of delivery of attractive attributes does not significantly harm the perception of quality. However, it is the delivery of delight features that can drive the overall response to the product, providing significant increases in the perception of quality.

The Kano model indicates that it is possible to create significant increases in satisfaction by providing product attributes that exceed the expectations of the customer, but that failing to deliver basic requirements can conversely damage the perception of quality. Indeed, it is necessary to establish the basis of functionality and usability (and arguably safety) before a product can provide a pleasurable user experience (Jordan, 2000). Section 3.3.3 in submission 1 details a case study relating to Ford’s ‘MyFord Touch’ infotainment system which experienced significant usability, reliability and safety issues when launched, creating a strongly negative customer reaction and causing significant damage to brand value. This illustrated the risks inherent in a technology-focused innovation strategy and emphasised that manufacturers must seek to optimise the user experience of in-vehicle technology before the vehicle reaches the market.

4.3. Driver Distraction

The use of technology in the car is a major source of driver distraction. The act of driving a car on a public road carries a degree of accident risk, due to a range of factors relating to the vehicle and its occupants (internal risks); and the environment and other road users (external risks). Data derived from real-world observation of driver behaviour indicates that approximately 80% of accidents are related to driver inattention (Klauer et al., 2006), indicating that human factors play a major role in driving accidents. Vehicle manufacturers face a challenge in how to enhance and promote the safety of their vehicles, with safety-related innovations forming a large part of certain manufacturers’ marketing strategy.

Driver distraction refers to the diversion of the driver’s attention away from the primary task of vehicle control. A full discussion of driver inattention, its causes and effects is given in submission 1, chapter 4; key points are summarised here.

Young et al. (2003) define four main types of distraction:

- Visual distraction – Visual attention is of primary importance when driving (Wierwille, 1993). Visual distraction can occur when the driver’s view is blocked by objects, when the driver’s gaze is diverted away from the roadway by other visual targets (either inside or outside the vehicle); or when the driver’s visual attentiveness is lost, i.e. the driver “does not see” a hazard.

- Auditory distraction – Auditory distraction is analogous to the second type of visual distraction, where the driver's attention is diverted away by signals not originating from the road environment. An obvious example of this is mobile phone use, which has been shown to have significant correlation to increased accident risk (Klauer et al., 2006, Yee, 2007).
- Physical distraction – Physical distraction occurs when the driver moves their hands from the primary driving controls in order to engage in another task. This can include operating other controls within the vehicle (e.g. heating and infotainment controls), or non-driving related tasks such as eating and drinking.
- Cognitive distraction – Cognitive distraction occurs when the driver's mental attention is diverted away from the driving task. This often manifests as an increase in reaction time. Again, an example is holding a conversation on a mobile phone while driving, which imposes both cognitive and auditory workload (as well as potential manual and visual workload in holding/operating and looking at the device).

Regan et al. (2011) seek to differentiate driver distraction from other forms of inattention, applying the term 'Driver Diverted Attention' to driving-related and non-driving-related sources of distraction. This taxonomy effectively segregates technology-based distraction from environmental and biological factors by acknowledging the division of driver attention as an element of the larger issue of inattention as a whole. As such, Regan et al. adopt the following definition of driver inattention:

"Driver Inattention means insufficient or no attention to activities critical for safe driving"

Driver distraction can arise from a variety of sources, both internal and external to the vehicle. Figure 15 shows the calculated odds ratios for in-vehicle activities, based on naturalistic driving data. The figures represent the relative probability of being involved in an accident while undertaking one of the listed activities compared to driving alone; for example, dialling a hand-held device increases accident risk by a factor of 2.8.

The continuing proliferation of in-vehicle technology increases the potential for the driver to be distracted from the primary task of vehicle control (Stutts et al., 2001). New technologies bring with them new sources of information that must be monitored and managed by the driver. OEM devices afford the manufacturer a degree of control over the presentation of information to the driver and can potentially make use of contextual information provided by other vehicle systems to mitigate distraction by limiting information display. Mobile devices such as smartphones present further challenges, although there are initiatives to integrate this functionality into vehicles, such as Android Auto and Apple CarPlay as mentioned previously.

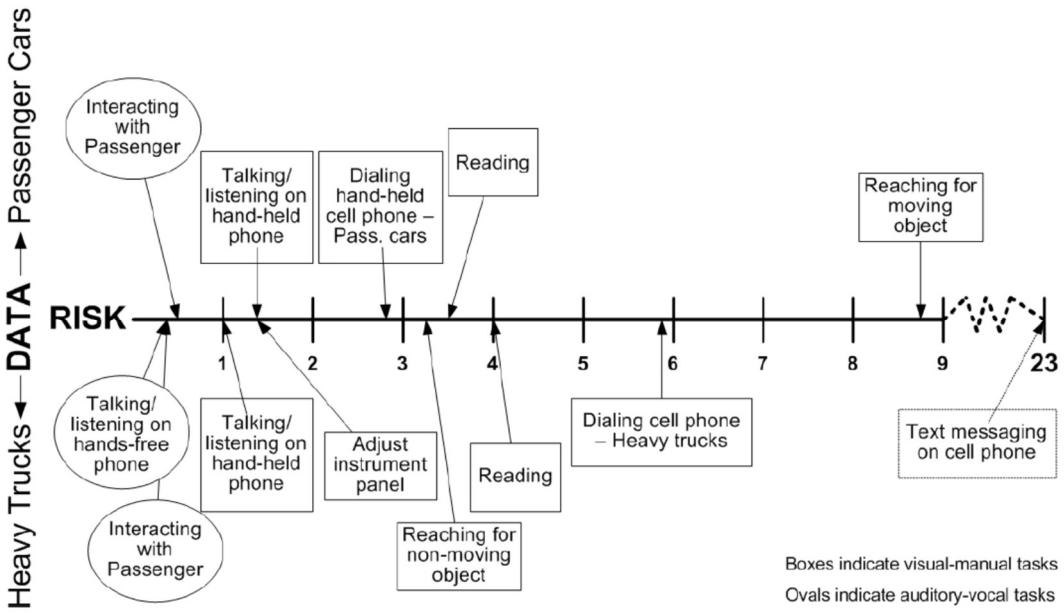


Figure 15 - Risk odds ratios for in-vehicle activities (NHTSA 2013)

Finally, there are a range of HMI issues relating uniquely to low-carbon electric vehicles, which may have an effect on driver attention. Most important of these is 'range anxiety': the driver's anxiety about the car's ability to cover the required distance before it must be recharged (Woodcock et al., 2012). As low-carbon vehicles become more prevalent, the number of 'naive' users will increase, hence accurate and reliable information must be relayed to the driver taking into account environmental factors and driving style, in order to minimise range anxiety issues.

4.3.1. Mental Workload

The assessment of mental workload is also highly important in understanding the impact that technologies have on the user. The concept of mental workload is widely studied within Human Factors research (Young et al., 2015), but is considered difficult to precisely define (Mehler et al., 2012, Miller, 2001, de Waard, 1996). Young and Stanton (2005) state that: *"mental workload of a task represents the level of attentional resources required to meet both objective and subjective performance criteria, which may be mediated by task demands, external support and past experience".* This emphasises the influence of the user's own conditions on the perception of workload while drawing a clear distinction between objective and subjective measures; it is easy to envisage a situation whereby a user perceives a high workload for a task which is objectively relatively simple, because the necessary workflow does not meet their expectation of how the task should be completed. Furthermore, perceived

workload is influenced not only by the demands of the task within the context of interaction, but by the user's own motivations and coping strategies (Mehler et al., 2012).

Exceeding the limits of resource capacity will generally result in a reduction in the subject's ability to perform a given task (i.e. a performance degradation)(Wickens et al., 2013). As technology levels increase and systems become more complex, the workload demands placed on operators are also increased. For in-vehicle technologies, interaction tasks that impart workload onto the driver will potentially degrade the driver's primary task performance; that is: to reduce their ability to safely control the vehicle. The assessment of workload is therefore highly important during the design of complex systems (Stanton et al., 2013), such as new in-vehicle technologies.

Mental workload assessment techniques can be placed in three main categories: physiological, subjective and performance-based (Miller, 2001). Physiological measurement involves monitoring changes in a subject's physical condition in response to changes in resource demand, such as heart rate, eye movement and brain activity (Stanton et al., 2013). Subjective techniques involve the use of scales or rankings to determine the self-reported subjective workload experienced by a user. Performance-based workload techniques involve examining the capacity of a user to through the application of a primary or secondary task. By measuring the performance of the task(s) and observing how that performance varies under increased workload, an estimate of the mental workload experienced by the user is obtained (Miller, 2001). Performance-based and subjective approaches are discussed in detail in submission 2, sections 6.1 and 6.2.

4.4. Industry Guidelines for Minimising Driver Distraction

A number of organisations have been involved in the development of guidelines and standards aimed at reducing distraction from automotive HMI. Within Europe, the Commission of the European Communities Statement of Principles on Human Machine Interface (HMI) for In-Vehicle Information and Communication Systems (European Commission, 1999, updated 2007) provide a series of design goals and installation principles for HMI designers which aim to limit distraction caused by in-vehicle systems, with examples of good and bad practice for each. The Statement of Principles references a number of ISO standards relating to ergonomics and the assessment of driver distraction (see submission 1 section 4.3.2 for details). From Japan, the Japan Automobile Manufacturers Association (JAMA) Guidelines for In-Vehicle Display Systems, version 3.0 (2004) outline limitations on information display. In addition, complex interaction tasks, such as inputting phone numbers and map scrolling, are prohibited when the vehicle is in motion. The guidelines also make reference to ISO standards for measuring glance behaviour.

In the USA, the National Highway Traffic Safety Administration (NHTSA) published the Visual-Manual NHTSA Driver Distraction Guidelines for In-Vehicle Electronic Devices (NHTSA, 2013); these are discussed in detail in submission 1 section 4.3.3. The guidelines were developed in recognition of the fact that the most distracting in-vehicle activities are visual-manual in nature, and incorporate the latest driver distraction research with a sole focus on safety. The publication draws heavily from previous works, in particular the Alliance of Automobile Manufacturers (AAM) Statement of Principles, Criteria and Verification Procedures on Driver Interactions with Advanced In-Vehicle Information and Communication Systems (2006), and aims to provide standardised test procedures that will help to minimise variation across tests and allow easier comparisons across tasks and systems.

One of the key features of the Visual-Manual distraction guidelines is the recommendation for 'lock-outs': the forced disabling of specific tasks while the vehicle is being driven, based on performance criteria. Acceptance criteria are applied on the basis of visual glance time, with limits on mean glance duration and total glance time for a given task. In addition, a range of tasks are recommended as 'per se' lock outs that will be disabled under driving conditions at all times, on the basis that they are either not appropriate for use while driving, or that they are not suited to the proposed assessment methods.

The focus of the phase 1 guidelines is on visual-manual distraction for installed (OEM) systems only; phase 2 guidelines relating to visual-manual interactions with aftermarket/portable devices were released in 2016 (NHTSA, 2016) and forthcoming publications will extend the scope to auditory-vocal interactions. While the guidelines are not legally binding, NHTSA state that they expect vehicles manufactured by AAM members and launched 2 years following the publication of the guidelines to be fully compliant.

4.5. Summary

In-vehicle technology has become a major point of focus for vehicle manufacturers as they compete to win market share by offering the latest functionality. While the potential for competitive advantage is apparent, it is vital that the impact of new technologies of the experience of the user is well understood. Many of the new systems introduced will have a tangible interface with which the driver can interact, potentially providing new sources and types of information. As the main means by which the user interacts with the vehicle, the Human-Machine Interface has a major impact on the quality of the User Experience. When implemented poorly, the effect on the overall perception of the quality of the brand can be significant.

The influence of in-vehicle technology on driver safety is clear, with a proven link between driver distraction and accident risk. Current guidelines and standards outline a range of approaches for evaluating the distraction potential of HMI solutions, with recent publications moving towards establishing acceptability criteria for task time and visual workload along with standardised testing procedures, while complementary guidelines for auditory-vocal interfaces are currently under development. Emerging technologies will continue to create challenges for designers and engineers as new functionality is added to the vehicle and mobile device integration standards are matured.

The previous chapter identified the importance of evaluating the UX on in-vehicle technologies as part of a user-centred design process. However, HMI design must also be carefully considered in terms of minimising the impact on the driver's ability to concentrate on the primary driving task. An effective technology evaluation processes must therefore address both areas in parallel, as part of a holistic approach.

5. Making the Case for a New Evaluation Toolset

The preceding chapters have identified the strategic importance of in-vehicle technology, in terms of commercial advantage, driver safety and customer satisfaction. As noted, levels of in-vehicle technology are continually increasing as manufacturers seek competitive advantage for their products, bringing new sources and types of information into the vehicle that the driver must process, representing an increased potential for driver distraction. Manufacturers also face significant pressure to provide innovative features at low cost and within short design cycles (Bhise et al., 2003). The increase in technology level invariably brings an increase in usability issues which are more likely to occur with complex systems that feature high levels of functionality (Harvey et al., 2011a).

It the integration of user-derived information into the NPD process is therefore vital to support the development of new technology concepts. This chapter presents a precis of the rationale given in submission 2, chapter 7 for the development of a new evaluation toolset to facilitate the holistic evaluation of in-vehicle technology concepts.

Broström, Bengtsson & Axelsson (2011) identify a distinction between research approaches employed in the evaluation of IVIS. Firstly, a formative approach is often employed to evaluate safety characteristics though driver behaviour studies. This type of evaluation is typically conducted by universities and research institutions. Vehicle manufacturers, however, tend to employ summative approaches to customer satisfaction and technology acceptance though user trials.

In the majority of these cases the focus is very much on the functional performance of the technology, rather than the user's experience of interacting with it. In other studies however, a combination of performance and usability metrics is often utilised. For example: Sumie et al. (1998) investigated the usability of menu-based interfaces, using objective measures along with subjective ratings of task difficulty, design preference and perceived safety. Döring et al.(2011) applied subjective ratings of ease of use, self-reported distraction and liking to the LCT-based evaluation of a steering wheel gesture input surface. Harvey et al. (2011b) performed simulator-based comparisons of touchscreen and rotary controller interfaces, using driving performance and secondary task performance metrics alongside subjective usability measures; and Rydström et al. (2005) and Broström et al. (2011) combined secondary task completion measures with expert heuristic usability evaluation in the assessment of IVIS interfaces.

Analytical approaches are also apparent: Manes & Green (Manes, 1997) compared task completion times for button and knob controls against a GOMS (Goals, Operators, Methods

& Selection) model. Pettitt et al. (2007) applied the Keystroke Level Model (part of the GOMS family of methods) to predict the visual demand of IVIS interaction, to provide a comparison to the occlusion method; and Harvey & Stanton (2012) applied Hierarchical Task Analysis and multimodal Critical Path Analysis to the evaluation of a touchscreen interface.

UX-centric examples focus more heavily on static evaluation of design features. Wellings et al. (2008, 2010, 2012) utilised the semantic differential technique, factor analysis and content analysis to determine key parameters for automotive switch feel quality. Gaspar et al. (2014) employed a similar approach to the evaluation of automotive audio interfaces; and Park et al. (2013) applied a magnitude estimation technique to create a model of UX for a mobile device, including dimensions of usability, user value and affect.

The importance of the user's experience of in-vehicle technology is clear, with a failure to meet the user's expectations shown to have a negative effect on the perception of quality of the product. It is therefore important to engage the user in the design process such that the product can be aligned with their needs. By engaging users at the earliest stages, i.e. in the discovery phase of the NPD process, the benefits of user engagement can be maximised. Failure to meet the needs of the user in the marketplace can have a detrimental effect on perception of both the product and the brand as a whole.

This focus on up-front development activities is highly significant for JLR, who as a company have adopted a development strategy with in-vehicle technology at its core. The UX literature discussed in section 3.2 identifies that the user's experience of interaction with automotive HMI extends beyond the functional into the emotional, hence evaluation methods must consider measures of affect alongside ease of use (Edmunds and Dorn, 2012). Traditional usability engineering approaches do not adequately consider the hedonic aspects of the experience and their relationship to utility and usability (Hassenzahl et al., 2001).

Hassenzahl (2004b) encapsulates the dilemma facing the designers of in-vehicle technologies for modern automobiles: if the major design goal is the efficiency and/or effectiveness of the product (as might be expected for a system designed to minimise driver attention requirements), then objective measures of task performance are most appropriate. Conversely, if the core goal of the designer is to create a rich user experience, then the focus of evaluation should be on the perceptions of the product by the user, requiring tools that can capture users' subjective impressions.

The UX of automotive HMI is, as with all products, also dependent on the user and the context of use (Burnett, 2008) and this must also be reflected in product evaluations conducted to support product development. Hence evaluation processes must consider aspects of affective

response to the product alongside functional usability and safety, and that the conditions under which product are tested must reflect their intended usage in the field.

While examples of technology evaluation approaches exist in the literature as identified above, there is often a lack of detailed description of the evaluation methodology, or discussion of why particular methods or approaches were selected. Submission 2, section 7.1 describes four evaluation frameworks, designed for the aeronautical and automotive sectors, which include a cross-section of the usability and workload methods described in submission 2 chapters 5 and 6. These encompass both analytical and experimental methods, generating both qualitative and quantitative data, representing a diverse range of tools that can be applied at appropriate stages of product development.

It is clear however that both functional, safety-related aspects and User Experience are critical to the success of in-vehicle technology and should be represented within the technology evaluation process. Simply put: there is no point developing an in-vehicle technology that provides a compelling user experience if it heavily distracts the driver in use; nor is there marketing value in a product that minimises distraction but is unappealing to the user. Given also that usability issues which remain unidentified until late in development can significantly increase costs and risk adding delay to the development timeline, it is essential that these issues are addressed in parallel throughout the product development process.

Jander, Borgval & Ramberg (2012) defined a set of requirements that must be met in order for an HMI evaluation toolset to be considered fit for purpose; these are discussed further in submission 2 section 9.1 and are summarised below:

1. **The ability to diagnose specific problems or weaknesses in design**
2. **The ability to benchmark different design solutions**
3. **The ability to support holistic system evaluations**
4. **The ability to distinguish if a system can be HMI approved or not**
5. **The ability to be utilized for simulator evaluations**
6. **The ability to perform evaluations iteratively**
7. **Low to moderate complexity**
8. **User-centred**

An effective evaluation toolset must facilitate the gathering of data on both functional and affective aspects of users' interactions with a range of technologies, whilst reflecting the context of evaluation. In addition, the toolset must assist practitioners in conducting studies, while generating clear findings that can be used to support product development decisions.

However, none of the frameworks identified in submission 2 chapter 7 fully meet the evaluation toolset criteria. There is therefore an apparent need for an evaluation toolset that addresses these concerns. This must integrate appropriate methods to create a holistic view of the relative merits of candidate technologies, while representing the context of use.

5.1. Summary

The importance of integrating the 'voice of the customer' into the NPD process is well established and has specific relevance to in-vehicle technology. Obtaining relevant and accurate information requires an evaluation toolset that can something encompass all key quality and performance metrics while reflecting the context of use of the concept under test. Traditionally, HMI evaluation has focused on functional, objective measures which, while important, do not fully reflect the experience of the user. As identified in earlier chapters, the subjective aspects of UX are also important and hence should be included in the assessment process.

The methodologies and toolsets identified in the literature do not adequately meet identified requirements, indicating the need for a new technology evaluation toolset, designed to provide a holistic view of the user's experience of interaction. Data gathered using the toolset will be used to support NPD decision-making processes, in turn supporting the technology-focused strategic goals of the company.

6. Research Methodology

The preceding chapters outlined the basis for the research, both from the literature and from an industrial perspective. This chapter outlines the research question, objectives and methodology for the project, including a discussion of the research philosophy.

The research methodology was revised early in the project to place greater emphasis on the development of UX-based technology evaluation tools to support decision making in the up-front NPD process. The original methodology is described in submission 1 chapter 6, while the revisions are explained in submission 2, chapter 8. Presented here is a summary of the revised methodology, supported by relevant justification from the earlier work. Issues relating to validity and reliability are discussed in submission 1, section 6.6.

6.1. Research Question

The project focuses on the use of driving simulators to provide a context-rich environment in which novel technologies can be evaluated. The evaluation process must take into account the issues of user experience and driver safety as previously discussed. The research question adopted therefore is as follows:

> *"How can simulation tools be used to support advanced product development processes to deliver HMI solutions that optimise the user experience while minimizing safety risk?"*

The main research themes and sub-questions are identified in table 5, section 6.2 of submission 1.

6.2. Objectives

The objectives for the project are derived from the research question and themes:

1. Determine the key objective performance indicators for driver behaviour
2. Develop a context-rich environment to enable evaluation of novel HMI technologies
3. Identify and evaluate methods for capturing the User Experience of in-car technology
4. Create and test a protocol for the evaluation of technologies from the perspectives of UX and driver safety
5. Determine how the evaluation protocol can be integrated into the up-front NPD process for in-vehicle technology

The objectives of the research are be addressed through a combination of theoretical and empirical approaches. While the underlying research philosophy remains unchanged, the revised research design is described below.

6.3. Research Design

The research design is illustrated in Figure 16. The research consists of three phases, described below:

- Phase 1 focuses on development and validation of evaluation tools, encompassing hardware and software tools along with data gathering and analysis methods. This includes development of the evaluation protocol and selection of appropriate methods, as outlined in chapter 10 of submission 2.
- Phase 2 consists of a series of technology case studies conducted on concepts from both the Milestone and TCDS advanced product creation processes, representing technologies at different technology readiness levels and requiring different degrees of maturation and differing outputs.
- Phase 3 applies the learning from the case studies conducted within phase 2 to revise the toolset, which will then be validated and integrated into the NPD process. The work from this phase is detailed in the Innovation Report (this document).

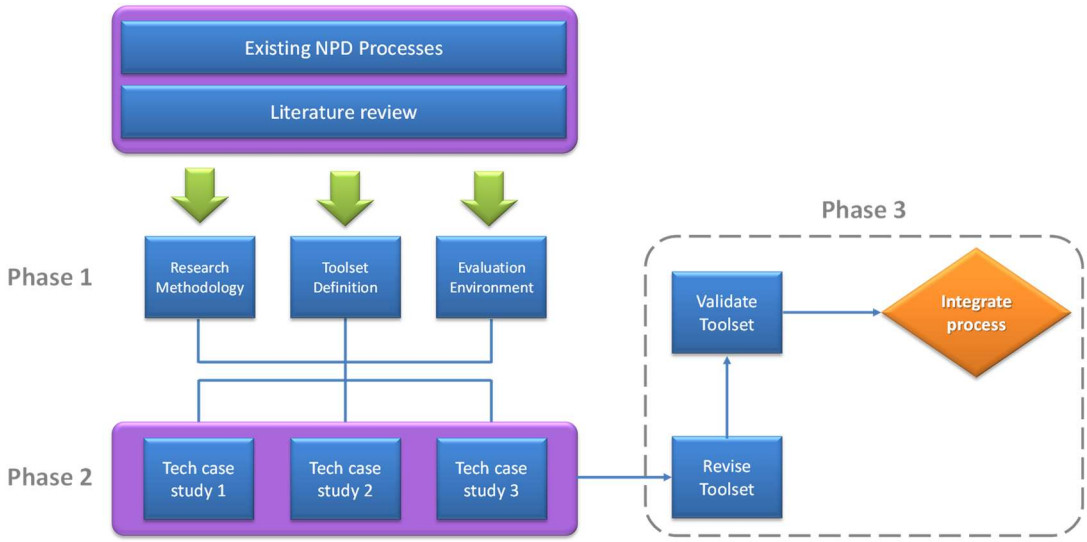


Figure 16 - Revised research design

The technologies selected for the case studies are shown in Table 1 below.

Table 1 – Technologies selected for case studies

Case Study	Technology	Format	PD Process	Modality	
				Input	Output
1	Voice dictation system	Software	Milestone	Vocal/Manual	Auditory/Visual
2	Touchscreen (Capacitive/Infra-Red/Resistive)	Hardware	Milestone	Manual	Visual/Manual
3	Autostereoscopic 3D display	Hardware	TCDS	-	Visual

For the toolset to be viable it must be applicable to the wide range of technologies likely to be investigated by JLR's HMI research team. As such, the technologies selected for the case studies feature a range of input/output modalities, include both software and hardware solutions, and have different technology-readiness timescales, aligned to both the Milestone and TCDS development processes. The outputs from these case studies will enable refinement of the evaluation toolset, with a view to its incorporation into the development workflow as a new business process.

6.4. Research Philosophy

The research philosophy, describing the research paradigm and the epistemological and ontological assumptions made by the researcher is summarised below. For a full description see submission 1, section 6.4.

The research deals in the observation, measurement and interpretation of human behaviour (Tidwell, 2006), therefore an empirical approach is employed. Empiricism typically implies a positivist paradigm, with an emphasis on qualitative methods and numerical analysis. However, the user-centred nature of this research requires the integration of more descriptive approaches in order to understand the nature of user experience. Therefore the overall methodology can be considered a mixed-methods design, incorporating both qualitative and quantitative methods to provide a detailed understanding of the research problem (Teddlie and Tashakkori, 2009, Creswell, 2006, Johnson and Onwuegbuzie, 2006). Adoption of this strategy represents a shift in philosophy towards post-positivism, a step away from the purely objective positivist perspective (Teddlie and Tashakkori, 2009). This acknowledges the element of uncertainty that exists in human-centred enquiry and the value of subjective knowledge, while retaining a neutral stance in order to avoid bias. A post-positivist approach is appropriate for the investigation of user experience due to the inherent influence of context

and users' expectations on their perception of product interactions: in effect, their 'reality' is personal to them and will differ from that experienced by another user.

6.5. Summary

Following the revisions to the research methodology defined above, the focus of the project is now centred on the development and testing of a toolset that will support the holistic evaluation of early-phase prototype technologies. The toolset was developed based on methods identified in the literature and tested through application to three technology evaluation case studies. The technologies chosen for the study reflect a range of concept readiness, architecture and development timescales, in accordance with the design parameters for the toolset discussed in the following chapter.

7. Evaluation Toolset Development

This chapter summarises the proposed evaluation toolset and protocol, which is described in detail in submission 2, chapters 9 and 10.

7.1. Toolset Definition

Chapter 5 provided a justification for the development of a new technology evaluation toolset, based on needs and requirements identified in the literature. These requirements form the basis of the design of the toolset. This section provides an outline of the methods and approaches selected; a detailed justification can be found in submission 2, section 9.2.

Jander et al. (2012)'s requirements explicitly identify the important of a user-centred approach. Given the importance of customer input into the NPD process and the subjective nature of the user experience, a **user-centred experimental approach** is preferred. An experimental approach is chosen over analytical methods due to the relevance of the data and depth of insight obtained (Stanton et al., 2013).

As previously identified, replicating the context of use becomes highly important when conducting user trials. A **driving simulator-based approach** was therefore adopted. The simulator allows technology solutions to be evaluated in a safe but realistic context, with environmental and workload conditions approximating those found in a real-world driving exercise (Godley, 2001, Reed, 1999, Hoyes, 1997).

Driving simulators are increasingly used in research, largely due to their ability to provide a consistent environment for evaluations (Kaptein et al., 1996). Driving simulation offers a number of benefits: firstly, it allows evaluations to be conducted under repeatable conditions, with close control over intrinsic and extrinsic factors such as traffic and weather conditions. Secondly, the use of simulation minimises the risk in the evaluation of conditions under which vehicle control may be significantly compromised, such as high levels of workload or distraction (Bach et al., 2009). Finally, it enables the use of key objective methods to measure driving performance as a means of evaluating the impact of the system under test.

The primary requirement for a technology evaluation toolset is that it is effective, i.e. that is can be successfully used to differentiate between different solutions and to identify design issues. This will be achieved through the use of established methods with proven sensitivity, reliability and validity. A mixture of objective and subjective methods are used in order to reflect actual performance levels as well as users' opinions of IVIS under test (Harvey et al., 2011c) and triangulation of methods will ensure that findings are not skewed by data anomalies. To minimise the complexity of the toolset, standardised test scenarios are

employed; this will also assist in drawing comparisons between technologies by ensuring a consistent evaluation platform with a known context in order to give a 'baseline' for system performance.

Also discussed in chapter 5 is the importance of considering UX in parallel to safety. A holistic evaluation approach is achieved through the application of subjective UX methods alongside performance, usability, and workload measures. This integrated approach is absent from the evaluation frameworks identified in the literature.

The ability to distinguish if a system can be HMI approved is determined through comparison of system performance to appropriate pre-determined pass-fail criteria. Harvey et al.'s framework requires that evaluation goals are determined at the start of each evaluation cycle. However, the safety criteria to be met by final products are determined by industry guidelines (such as the NHTSA Guidelines discussed in section 4.4) and are consistent across systems, being based on safety limits. Acceptability criteria for subjective measures can be determined with relative to one or more reference system; or on a statistical basis (e.g. requiring a minimum percentile rating to pass). Acceptance criteria are discussed in section 14.2.

Jander et al. state that the toolset should support the iterative evaluation of technology concepts. Simulator-based user evaluation necessitates a certain level of functionality in order to support user interaction, hence cannot be used at the very earliest stages of product development. However, the evaluation toolset can be developed to support the product development workflow such that evaluations can be implemented at the earliest possible phase; this places specific demands on the design of the driving simulator (see section 8.1 for discussion).

7.2. Toolset Applicability

Blandford & Green (2008) identify that: *"Evaluation methods can be broadly classed on three dimensions: 1) With or without the active involvement of users; 2) With or without a running system; 3) with or without a realistic context of use. Most methods fit within this three-dimensional space."*

Using Blandford & Green's taxonomy, the methods selected for the evaluation toolset all fit within the 'with users'/'with system running'/'with context' category, as illustrated in Figure 17. The user is at the heart of the proposed toolset; while analytical methods can be used to evaluate interface concepts earlier in the design phase, the aim of the toolset is to derive user-based information to identify UX issues and inform the ongoing PD process. As identified above, this necessitates a minimum level of interactive functionality from the system under test.

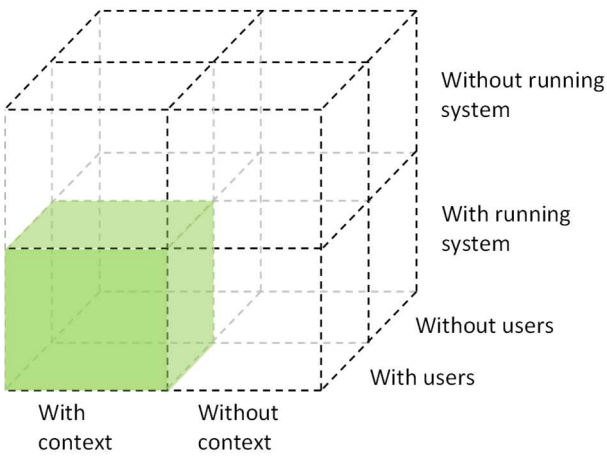


Figure 17 - Technology applicability for evaluation toolset

The minimum technology development level required negates some of the perceived advantages of desktop analytical methods, which are well suited to low-cost evaluation of very early stage prototypes, before any significant functionality is developed. Furthermore, analytical methods are not appropriate for the evaluations required in case studies 2 and 3 (see Table 1, section 6.3), whereby it is the fundamental underlying performance of a display technology that is under test rather than a graphical user interface.

The tasks to be evaluated will be determined in large part by the level of functionality of the system under test: while ideally these should reflect the full range of a product's attributes of interest, this may not be possible for early prototypes. It should be noted that the level of prototype fidelity can vary widely based on the product type and development stage (McClelland, 1995), and this will affect the validity of the results of the evaluation (Sauer and Sonderegger, 2009, Sauer et al., 2010).

7.3. Dependent Variables and Data

A holistic evaluation approach requires that a range of measures are applied in order to create a full picture of the user's experience of interaction with the system. Employing multiple methods, both objective and subjective, also allows triangulation of data, providing greater confidence in the validity of the outcomes.

Firstly, objective performance measures, relating to both the driving and technology tasks are used. Performance measures generate quantitative data that are captured in software or derived from video analysis.

Lateral vehicle control is measured using the standard deviation of lane position (SDLP), while longitudinal control is determined through the measurement of headway to a leading

vehicle, in terms of distance or time. These metrics are specified in the NHTSA Visual-Manual distraction guidelines (NHTSA, 2013).

Task performance is measured through task completion time and error rate (e.g. incorrect button presses, unsuccessful completion). This may be achieved through software logging of user inputs (where possible), or through video capture and coding post-evaluation.

Participants' visual behaviour is also captured. This may be achieved using video, allowing for the coding of glance events and calculation of eyes off road time (EORT) and glance frequency. However, conducting this type of analysis manually is arduous and time consuming, therefore it is preferable where possible to utilise automated eye tracking equipment to determine gaze vectors and calculate visual metrics.

Secondly, quantitative subjective usability, workload and UX data are collected post-task using a questionnaire approach.

Usability is assessed using the System Usability Scale (SUS) (Brooke, 1996, 2013). SUS is simple to apply and to understand, exhibits favourable reliability (Bangor et al., 2008, Lewis and Sauro, 2009) and is shown to be sensitive to differences between interface solutions (Bangor et al., 2008).

Task workload is assessed using the Raw Task Load Index (Byers et al., 1989) variant of the NASA Task Load Index (TLX) (Hart and Staveland, 1988, Hart, 2008). This approach benefits from extensive evidence of validity while offering benefits in ease of application over the original NASA-TLX method.

The user's affective response to the product is measured using the Hedonic Rating Scale (Peryam and Pilgrim, 1957, Stone and Sidel, 2004) - This is a simple rating of liking, derived from sensory evaluation techniques; the simplicity of the scale means that it is easy for both researchers and participants to use (Lim, 2011), and that it is well suited for integration into a questionnaire-based approach as part of a wider suite of measures. The method has previously been successfully applied by the author in the evaluation of automotive technologies (Pitts et al., 2012b, Pitts et al., 2012a).

In addition to the hedonic rating data, a post-trial interview is conducted with each participant. This approach can provide deep insight into the issues that drive preference, adding qualitative detail to the quantitative measures accrued. The interviews are semi-structured, leading with a focus on the user's affective response to the technologies under test, with follow-up questions seeking to identify the main reasons behind that judgement. This approach will also allow the derivation of preference ranking of the candidate technologies. The interview approach is described in submission 2, section 10.2.

7.4. Summary

This chapter provided an outline of the approach, environment and methods that form that toolset, identifying how each element of the toolset addresses key requirements derived from the literature.

In summary, the key features of the toolset are:

- A **User-Centred, Experimental** evaluation approach
- Uses a **Driving Simulator** to provide a context-enhanced evaluation environment
- Is based on a **Standardised Evaluation Protocol** to minimise bias and variability
- Employs a combination of **Qualitative and Quantitative evaluation methods**
- Utilises triangulation of data from **Driving Metrics, Usability, Workload and UX measures**

The tools described here are applied through the evaluation protocol defined in Chapter 9.

8. Driving Simulator Development

As identified in section 7.1, driving simulators are widely used in research settings due to the repeatability and safety benefits offered. Simulators also place the participant in the context of driving, recreating aspects of the user experience and approximating the workload conditions of a real-world driving task. Furthermore, a virtual driving environment supports the use of driving-related metrics to measure performance. It must be remembered however that a driving simulator is an approximation of the real-world activity; the researcher must always use caution when seeking to generalise the findings of their work to the wider world and be aware of the limitations of their system.

The original basis for the WMG simulator resulted from the HMI workstream of the Low Carbon Vehicle Technology Project (LCVTP). LCVTP was a collaborate project between a number of industry and academic partners, funded by the European Regional Development Fund and Advantage West Midlands. The HMI workstream was led by WMG and supported by JLR, Tata Motors European Technical Centre (TMETC) and Coventry University. Work conducted under this project set the initial specification for the simulator, including the budget, software selection, hardware procurement and visualisation design.

This chapter provides a summary of the key work undertaken to develop the WMG driving simulator from that basis to fit the requirements of the evaluation toolset. The development of the simulator is covered in detail in Submission 3.

8.1. Simulator Specification

The term 'driving simulator' can be used to describe a wide range of solutions, from a desktop system featuring a single monitor and a gaming wheel/pedals, to a fully-immersive system offering six degrees of motion and costing multiple millions of pounds. The exact design and specification of a given system will be largely determined by its intended use and by the available budget.

Generally speaking, increased realism demands increased cost and complexity. For example, a wrap-round display will provide greater immersion, but will require additional projectors, graphics cards and PCs; and potentially additional hardware/software for warping and blending graphics card outputs to create a coherent image. Similarly, including a motion platform will replicate some of the kinaesthetic cues that are absent from a fixed-base system, but adds increased cost and complexity.

The WMG simulator is shown in Figure 18. It was designed to minimise system complexity where possible while providing an immersive experience and observing budgetary constraints. The system therefore utilises a fixed-base architecture with no vehicle cabin motion. The system runs OKTAL SCANeR Studio software on a network of four PCs, described in section 8.1.1. Visualisation is provided via three flat projection displays to the front of the vehicle, described in section 8.1.2.

Figure 18 – WMG driving simulator

The context of use is highly important to the overall perception of a product and should be reflected in the evaluation process. Changes to the vehicle cabin interior were minimised such as to maintain as much of the original vehicle context as possible while providing the connectivity required to integrate the vehicle cabin into the virtual environment (see section 8.1.3). The primary driving controls were modified to interface them with the simulator PC hardware and allow participants to drive the virtual vehicle, described in section 8.1.4.

A key design requirement was that the simulator must support evaluations of different types of technologies in differing stages of development, including third-party software and hardware concepts (such as the case study technologies identified in section 6.3). In addition, the simulator should be compatible with the HMI development tools and workflows identified in submission 1, section 5.4, such that concepts can be quickly deployed onto the simulator for evaluation. This required modifications to the vehicle's controls and displays, described in section 8.1.5.

8.1.1. Software

The virtual environment is generated using OKTAL SCANeR Studio[1] software. This is a dedicated driving simulation package with a range of features that make it suitable for this application. SCANeR Studio is designed to be run across a network of PCs, with different software modules allocated to different machines on the network. In this way, processing workload is distributed across the network, ensuring that performance limitations are not encountered.

The software includes tools for terrain creation, allowing experimenters to create and apply their own simulated environments. The software is compatible with standard 3D model formats, allowing externally-created objects to be imported into the virtual environment. SCANeR Studio also features native scripting functionality, allowing creation of detailed scenarios featuring interactive events based on trigger conditions. This includes control of traffic vehicles and environmental conditions such as lighting and weather. In addition, a comprehensive API supports integration of 3rd party software and hardware, allowing real-time dynamic data exchange with the simulation; the dash model described in section 8.1.5 is an example of this capability.

The simulator must support the collection of all relevant metrics for the evaluation methods selected, including objective data relating to lateral and longitudinal control of the virtual vehicle. SCANeR Studio supports the collection and output of a wide range of parameters relating to the virtual vehicle and its position within the virtual environment. Data can be easily exported for analysis in standard ASCII format. Data can also be imported and written to the recording file across the network using custom data channels.

8.1.2. Visualisation

The visualisation system for the simulator consists of three flat projection displays positioned in front of the vehicle cabin, with the centre display perpendicular to the centreline of the vehicle and the side displays angled forwards by 45°. The use of flat screens simplifies the projection setup as no image warping or blending is required. The display hardware was designed and supplied by Holovis International Ltd. Figure 19 shows images from the original CAD.

[1] https://www.avsimulation.fr/solutions/

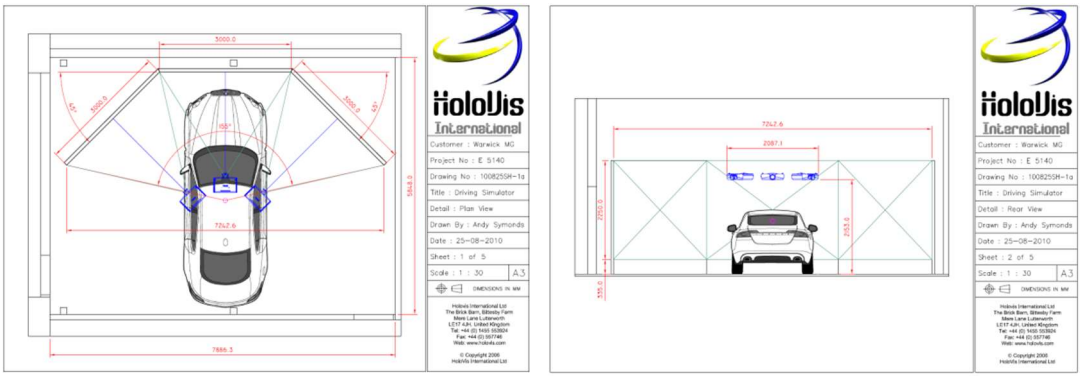


Figure 19 - Visualisation system CAD images

Each display surface measures 3.00m x 2.25m (4:3 aspect ratio), producing a viewing angle of 135°. Images are produced using three Sony VPL-FE40 projectors suspended from the frame above the vehicle cabin. These operate in SXGA+ resolution (1400x1050px), creating a total display resolution of 4200x1050px. Each projector is fed from its own visual server, connected to the SCANeR network.

A 5.1 channel surround sound system provides immersive audio from the virtual environment. This includes engine, wind and road noise from the driven vehicle, along with sounds from traffic vehicles. The positioning of the surround channels ensures that passing traffic vehicles provides an immersive effect. The subwoofer is positioned under the vehicle cabin to emulate the source position of low frequency engine and road noise.

8.1.3. Simulator Cabin

The basis for the simulator is the cabin of a 2009 Jaguar XJ premium saloon car, supplied by JLR. The powertrain of the original vehicle was removed and the body was cut rearward of the B-pillar, retaining the full front section of body including the full interior and the main section of the headlining. This ensures that the simulator retains the 'feel' of the original vehicle, providing a context-rich environment to facilitate the evaluation of in-vehicle technologies. The cabin sits on its original front wheels and is mounted to a rigid frame at the rear; the simulator is therefore a fixed-base configuration with no cabin motion.

The original seats are retained, with the ECUs modified by JLR engineers to allow operation of the original adjustment controls with an external 12V supply while disconnected from the main loom. Again, this helps to maintain the feel of the original vehicle. Care was taken to ensure that cabling was not visible from the driving position.

8.1.4. Primary Driving Controls

The primary driving controls of the vehicle are connected to the simulation software, allowing users to drive the vehicle within the virtual environment. The original steering wheel is retained, with the airbag removed and trim replaced. The steering column is connected to a

force feedback system from OKTAL, providing realistic steering force and vibration playback from the virtual environment. The system communicates with the simulation PC over a dedicated LAN interface.

The original pedal set was removed and replaced with the pedals from a Logitech G25 gaming wheel, inverse mounted on a custom bracket; this was connected via USB to the simulation PC, identifying in Windows as a gaming controller. The simplicity of this setup ensured compatibility and reliability, while allowing easy calibration of pedal position using built-in tools.

Given that the virtual vehicle model features an automatic transmission and the simulation scenario involves steady-state driving, there is little requirement for the driver to change gear. For simplicity, the gearshift is therefore operated by the experimenter via a USB control pad. For future applications, a more interactive gearshift mechanism may be developed. The engine start button is connected to a dedicated USB interface designed for simulator use, identifying as a game controller in Windows. This is configured using built-in tools to engage the ignition in the virtual vehicle model when pressed.

8.1.5. Additional Controls and Displays

The simulator is designed to be flexible to support the evaluation of a range of interface and technology concepts. A number of modifications have therefore been made to support the integration and operation of both hardware and software concepts.

The production XJ features a TFT configurable dash display which displays context-specific information to the driver. This includes secondary displays, such as the speedometer and rev counter; and tertiary information such as navigation instructions and media information. The display features an embedded controller that receives signals from other nodes across the CAN bus and displays pre-determined information accordingly. For the purposes of the simulator however, additional flexibility is required to support the display of novel interface concepts. The standard display unit was therefore replaced with a development unit supplied by Visteon Corporation, based on the same display panel, but with the embedded display controller replaced with a DVI-LVDS converter. This allows the dash display to be connected to the simulator PC as an additional monitor, registering in Windows as an extended desktop, and allows the experimenter to display a wide range of graphical content to the driver.

To replace the dash interface, a virtual dash model was developed. This was built as a SCANeR Studio module using NI LabVIEW to access the SCANeR API. The model accesses engine speed, vehicle road speed and gear engaged data, and displays them on a simple GUI featuring a gear indicator, rev counter and speedometer. The GUI was based on graphics

extracted from a JLR-produced Adobe Flash model of the original dash application, retaining some of the authentic look. The display and dash model are shown in Figure 20.

Figure 20 - Configurable dash display showing LabVIEW dash model

The centre console touchscreen unit (HLDF) features the electronics for the audio and navigation systems. Again for flexibility, this was removed and replaced with an 8" resistive USB touchscreen monitor. As with the dash display, this can be simply connected over VGA/HDMI to any PC, allowing any graphical content to be displayed. The touchscreen element connects via USB and acts as a Human Interface Device (HID), similar to a standard mouse. The display and its installation is shown in Figure 21.

Figure 21 - Centre console USB touchscreen installation

To complement the central touchscreen, the steering wheel controls were also modified to allow them to interface with HMI concepts. Normally, these switchpacks would communicate with the dash module over LIN; however, the dash module was removed in the display modifications described above. The switches were therefore modified to allow them to be connected to a second wireless USB control pad mounted inside the steering wheel in the airbag cavity. This interface provided up to 16 discrete button inputs which can be configured in software using built-in tools. The interface installation is shown in Figure 22.

Figure 22 - Steering wheel control interface installation

8.2. Driving Environment and Scenario

The driving environment and scenario were developed to present the participant with a simple driving task with relatively low workload requirements that would require a minimal amount of learning to achieve consistent performance. Given that the core aim of the scenario is to derive information relating to the driver's interactions with in-vehicle technologies, it was important to ensure that the participant was afforded 'space' in their resource to allow them to engage with the technologies under test. A challenging driving environment would demand a much higher proportion of the participant's attention, either limiting their ability to conduct secondary tasks or causing catastrophic failures in primary vehicle control when their attention is diverted away. A simple driving scenario also helps to minimise the occurrence of simulator sickness.

A motorway environment was therefore designed, utilising a lead vehicle to provide a reference for longitudinal control variation. The environment was developed using the terrain creation and scenario authoring tools within SCANeR Studio. Tijerina (1996) identifies that steady motorway driving represents the conditions under which drivers are most likely to divert their attention to secondary tasks

The vehicle model used in the scenario is a 'Large Executive Car', the closest approximation to the Jaguar XJ donor vehicle available in the software. The roadway features three driving lanes plus a hard shoulder in each driving direction, separated by a central reservation. Armco-style barriers and street lights are positioned along the central axis of the road. The ground outside the paved roadway rises to form high banking, reducing the visible distance. A screenshot from the environment is shown in Figure 23.

Figure 23 – Simulator scenario screenshot

The roadway follows a curved trajectory, featuring a series of alternating clockwise and counter-clockwise bends with 3000m radius. This ensures that lateral position must be attended to constantly; any degradation in lateral tracking will be emphasised. The use of a curved trajectory is in contrast to the recommended straight layout featured in the NHTSA guidelines (NHTSA, 2013); however in this situation the driver could experience significant levels of distraction from the primary driving task without an increase in lateral deviation. Road curvature and lane widths are based on UK motorway specifications, while road marking comply with the Traffic Signs Manual 2003, chapter 5 section 4.

The driving task itself is simple: participants were required instructed to follow a lead vehicle travelling at a constant speed of 70mph, at what they perceived to be a safe distance. The lead vehicle is constrained to the inside lane; additional automated vehicles are included for context but the scenario does not require negotiation of traffic nor are there any reaction events from the lead vehicle such as emergency braking. After a stabilisation period, the secondary tasks are introduced and completed by the participant. When all tasks are complete the participant is asked to pull over to the hard shoulder and stop the car, signifying the end of the drive.

8.3. Summary

The WMG driving simulator, commissioned under LCVTP, has been developed under this project to meet the requirements of the evaluation toolset identified in the preceding chapters. Starting with the vehicle cabin, visualisation hardware and simulation software delivered under LCVTP, additional developments to controls and displays have been applied to support the integration of early-stage technology concepts into the vehicle, in both hardware and software.

A bespoke virtual environment and evaluation scenario were created, based on UK motorway specifications. These were designed to provide a consistent driving workload, encouraging participants to engage with the technologies under test, while capturing variations in their driving behaviour. Data from the simulation can be exported and objective performance metrics calculated, forming one facet of the holistic evaluation approach previously defined.

9. Evaluation Protocol

This chapter outlines the protocol that forms the basis of the technology evaluation approach, determining how the case studies will be conducted. Figure 24 illustrates the workflow for the evaluation studies. The protocol is based on the approaches defined by Stanton et al (2013), Harvey et al. (2011c), McClelland (1995) and Bhise et al. (2003), and is described in detail in submission 2, chapter 10.

The protocol consists of three main sections: Pre-trial, Trial and Post-Trial. The earliest activities in the pre-trial phase are the selection of the candidate technology/technologies; and definition of the research question. It is important to have a clear research question for the study such that the aims and objectives can be clearly defined and communicated. Once the research question is established, the use cases and tasks to be performed can be defined, reflecting the range of functionality of interest for the given technology. The final step of the pre-trial phase is to identify and recruit appropriate participants.

The Trial phase forms the body of the workflow. Each participant in the sample begins with an induction, during which the study is explained and consent is obtained (see section 9.2 for a discussion of ethics issues). Pre-study questionnaires are then completed, obtaining background data and a baseline reference for simulator sickness by completing the Simulator Sickness Questionnaire (SSQ) (see section 9.2.3). Once complete, the participant undertakes a practice/baseline drive in order to familiarise them with the simulator and obtain a reference for their driving behaviour. Baseline drives are conducted in the single-task condition, i.e. driving only.

The remainder of the trial phase follows an iterative process. The protocol employs a within-subjects experiment design, in which each participant experiences all of the technologies under test, discussed in section 9.1. For each of the technologies, the participant will be trained on the tasks to be conducted, then will begin a trial drive in the dual-task condition. The assessment of whether the participant is trained to an adequate level to undertake the trial drive is subjective and made by the experimenter. Tasks will be completed and data recorded until the tasks are exhausted. At the end of the drive the participant completes a post-task questionnaire, including the hedonic rating, workload and usability measures. The process is then repeated for each of the remaining technologies under test. Once all technologies have been tested, a post-study SSQ is completed and the post-study interview is conducted. The trial phase is then repeated for each of the remaining participants.

The Post-Trial phase involves the collation and analysis of the data from the study. Qualitative and quantitative data collected throughout the study (as shown on the right of the diagram) are

brought together to form the metrics defined in section 7.3. Comparison between the single- and dual-task conditions will provide an indication of the effect of the in-vehicle technology on driver performance, using the statistical methods described in section 9.1.3.

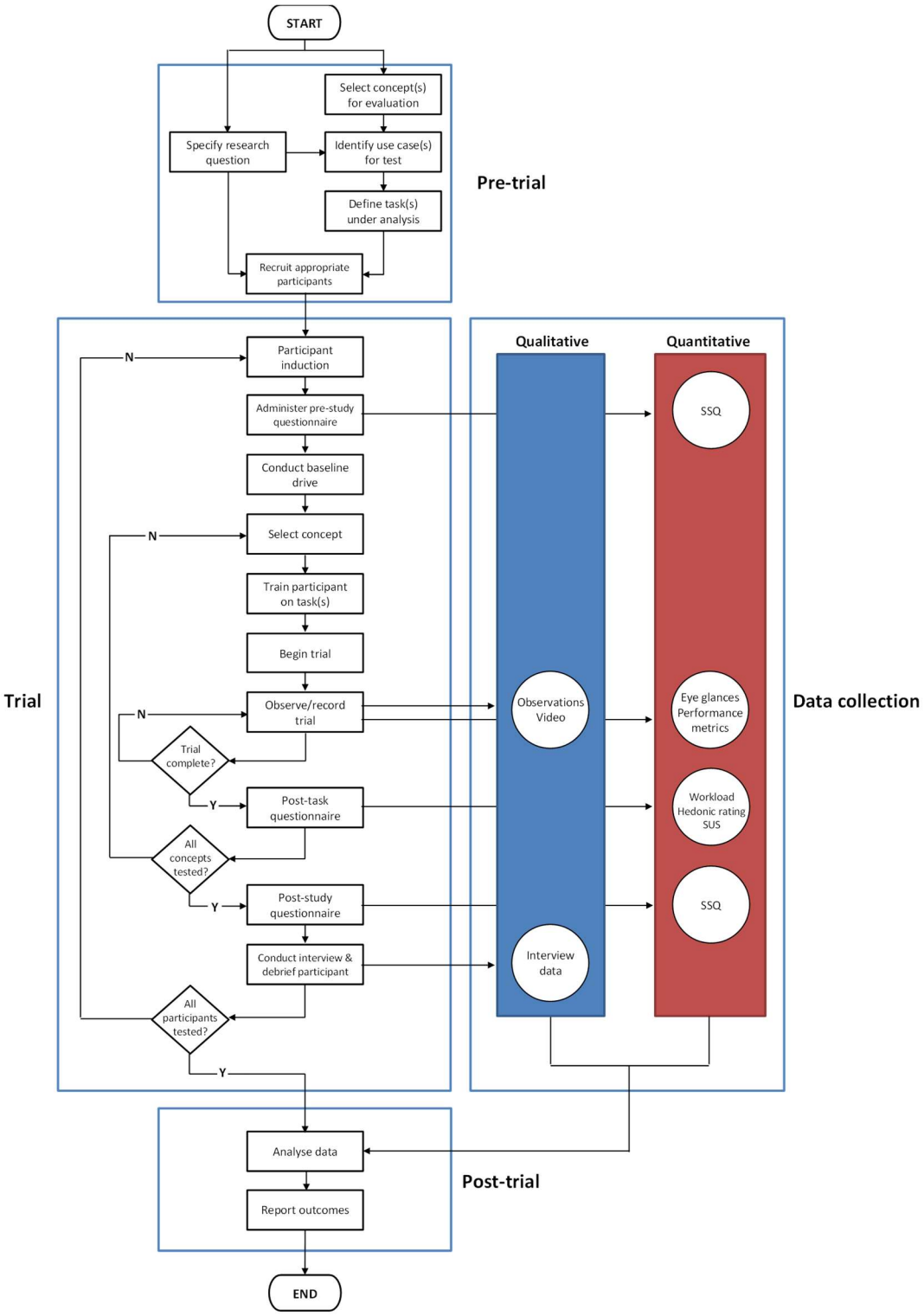


Figure 24 - Evaluation study workflow

9.1. Experiment Design

A within-subjects experiment design is employed, whereby participants complete a number of measured drives, corresponding to the number of levels in the variable under test (normally 2-4 drives), therefore each participant experiences all of the experimental conditions. As an example, an earlier study (Pitts et al., 2012a) focused on the evaluation of haptic feedback touchscreens, with haptic feedback states as the independent variable; participants completed one practice and two experimental drives, corresponding to the two haptic feedback conditions.

Participants are provided with training on both the driving task and the technology task prior to conducting dual-task drives. It is important to ensure that participants are accustomed to driving in the simulator and that their driving performance has stabilised before data collection commences. Similarly, participants must be allowed time to practice technology use cases in order to form a mental model of the operation of the system that will allow them to complete tasks efficiently (Bengler, 2007). The experiment design is balanced for presentation order to avoid bias. Instructions to participants are discussed in section 9.1.1 below.

Driving sessions are limited in duration to around 15 minutes in order to reduce the risk of simulator sickness. Allowing for introductory procedures, training rest time between drives and data collection; total involvement time for each participant should be no more than two hours.

9.1.1. Instructions to Participants

At the start of the trial phase, participants are briefed on the purpose of the study and the tasks that they will undertake, both in terms of the primary driving task and the secondary technology interaction task. For the primary task (described in section 8.2), participants are instructed to follow the lead vehicle "at what they perceive to be a safe distance", encouraging them to adopt their normal driving behaviour. A combined practice/baseline drive is then undertaken: the baseline is taken from a stable section of the data once the driving performance had stabilised. There were no specific criteria for 'stable' driving performance.

Participants are then given training on the secondary tasks in a single-task condition (i.e. non-driving). Participants are instructed on each of the tasks to be conducted and allowed to ask questions. Once training was complete, the trial drives are undertaken, one for each treatment in the within-subjects experiment design. No specific instructions are given in terms of prioritising either the driving or secondary task. No training/practice is conducted in the dual-task condition prior to the trial drives.

9.1.2. Data Collection

As discussed in section 7.3, objective data is recorded from the simulation software or derived from video recording. Subjective rating data is collected using a questionnaire administered after each evaluation drive; an example questionnaire is included in submission 2, appendix 2. Once all experimental conditions have been completed, a semi-structured interview is conducted, based on the reasons for the participant's preferences for the technologies tested. The interview is recorded (audio or video) with notes taken by the experimenter. The forced-choice preference assessment also supports a ranking of the technology options as a quick-glance indicator of relative performance.

9.1.3. Data Analysis

Objective data collected is analysed using parametric statistical methods, such as within-subjects ANOVA, allowing a clear indication of difference between experimental conditions. Subjective questionnaires data yielding ordinal data are analysed using non-parametric methods, including Friedman's test (non-parametric ANOVA), Wilcoxon's Signed Ranks test and contingency table approaches. Analysis is performed using Microsoft Excel, IBM SPSS and Mathworks MATLAB.

The post-evaluation interviews are transcribed and subjected to thematic analysis. This provides an understanding of the key themes that drive users' perceptions and preferences which is highly valuable in compiling a complete picture of the interaction experience.

9.1.4. Participants

The sample size required for the study is determined by the demands of objective data analysis, with a larger sample likely to improve statistical confidence in the findings. A sample size of 24 is recommended for simulator-based evaluations of in-car technology by the NHTSA guidelines on visual-manual driver distraction (NHTSA, 2013). Power calculations detailed in submission 2 section 10.4 indicate that this sample size will provide statistical power of greater than 0.9 when using a paired test (as would be used with a within-subjects experiment design) at $\alpha = 0.05$; a sample size of between 20 and 30 will therefore provide statistically significant results.

Participants should be adults aged between 18 and 65; 18 is the minimum age recommended in NHTSA guidelines, and participants over 65 years of age have an increased risk of simulator sickness (see section 9.2.3) (Brooks et al., 2010). Cacciabue & Carsten (2010) identify driving experience as a factor influencing intended speed and error propensity. Furthermore, as the evaluation environment will be based on UK road characteristics and to avoid variance in training levels arising from international driver training programmes, participants will be

required to have a minimum of one year experience of driving in the UK. Health-based constraints on participant recruitment are described in section 9.2.3.

9.2. Ethical Considerations

There are three main ethical considerations relating to this study: informed consent, data protection and simulator sickness. A description of each issue and plans for mitigation are detailed in turn below. Ethical approval was granted by the University's Biomedical Science Research Ethics Committee (BSREC); copies of the application form and approval letter are in submission 2, appendix 5.

9.2.1. Informed Consent

The issue of consent to participate in the research will be addressed through the use of information sheets and consent forms. The information sheet is provided to the participant prior to the user trial and details the purpose, requirements, duration and risks of the study, along with contact details for the researchers. The participant will be free to cease their involvement in the study at any time without penalty.

The consent form must be signed by the participant before the start of their involvement, confirming that the participant has understood the requirements of the study and the associated risks. Examples of the information sheet and consent form are included in submission 2 appendix 3.

9.2.2. Data Protection

Information recorded during the course of the user trials will be used solely for the purposes of the research. Personal data recorded from participants (e.g. name/age) is limited to the consent form, with an ID number allocated to each participant which is then associated with study data. No personal information will be used in publication of findings from the study. No video footage or still images from the study will enter the public domain without the express permission of the participant featured.

Hard copy forms and questionnaires are securely stored in a locked cabinet in accordance with University of Warwick data protection policies, while electronic data are securely stored on the WMG file server in an anonymised form. Access to raw versions of both hard copy and electronic data is restricted to the researcher and principal investigator. Data will be retained for the duration of the EngD project, at which time electronic and hard copies will be destroyed.

9.2.3. Simulator Sickness

While the demands of operating the driving simulator are not significantly different to those experienced in everyday driving activities, some participants may experience the symptoms of motion sickness while operating the simulator. Symptoms include headache, sweating, nausea, dizziness and drowsiness (Brooks et al., 2010). Simulator sickness is believed to be caused primarily by sensory disparity between the visual and vestibular systems, whereby visual cues indicate movement while the body remains stationary (Zaychik and Cardullo, 2003). The symptoms of simulator sickness are temporary and subside with time once the participant is removed from the virtual environment (Cobb et al., 1999).

The risk of simulator sickness is mitigated using a multi-stage approach. Firstly, potential participants are screened during the recruitment process to exclude those who may be more susceptible. Applicants who report susceptibility to migraine, epilepsy, motion sickness, vertigo, postural instability, blurred vision are not allowed to take part in the study. As an extra precaution pregnant applicants and those with heart conditions are also excluded. Prior to commencing the study and obtaining consent, the researcher must ensure that the participant is in their normal state of health (i.e. not suffering from any illness), that they are aware of the risk factors and again confirm that they suitable to take part.

Secondly, each participant must complete a Simulator Sickness Questionnaire (Kennedy et al., 1993a), before and after they participate in the study. The purpose of this is twofold: to make the participant aware of the potential symptoms such that they can stop if they feel unwell; and to monitor the health effects of interaction with the driving simulator. An example of the simulator sickness questionnaire is included in appendix 4. Participants will be asked to sign the questionnaires to confirm their condition.

As duration of exposure is known to increase the risk of simulator sickness (Kennedy et al., 2000), the duration of each drive is minimised and rest periods provided between each drive. During the user trials, the experimenter monitors the condition of the participant for signs of discomfort or distress, such as sweating, burping or yawning. Should such signs become apparent the researcher will cease the trial. All participants are also instructed to avoid driving their own vehicles or operating heavy machinery for a period of at least 30 minutes following their participation in the study, to ensure that any possible symptoms have subsided; Cobb et al. (1999) identified that symptoms generally subsided within 10 minutes of removal from the virtual environment.

It is important to note that in a previous study which made use of the precautions listed above (Pitts et al., 2012a), no problems with simulator sickness were encountered.

In the instance that a participant does suffer from a degree of simulator sickness, a number of contingency measures may be implemented, following the recommendations of Brooks et al. (2010). Firstly, participants will be removed from the simulation environment to a quiet location where they can recover. Drinking water and light snacks will be made available, and cleaning equipment will be on hand. The researcher will ensure that the department First Aider is aware of the potential risk and is available to assist in case of any issues. The participant will be encouraged to remain for a minimum of 1 hour to allow symptoms to subside before leaving, and the researcher will follow-up with the affected participant to check on their condition.

9.3. Summary

The evaluation protocol outlined above defines the process for conducting technology evaluations in a simulated driving environment, describing how the data collection measures defined in chapter 7 will be implemented. A within-subjects experiment design is employed, with subjective data collected at the end of each drive and at the end of the study. Data collected will be statistically analysed using appropriate parametric and non-parametric methods.

The protocol includes an assessment and treatment of the ethical concerns and risks of the study approach, including factors to mitigate the risk of simulator sickness. The protocol was reviewed by the University's Biomedical Science Research Ethics Committee (BSREC) and approved.

9.4. Introduction to Case Studies

The following chapters describe the findings of the three case studies identified in Table 1 (pg. 36). For each case study, a brief summary of the background is given along with the major findings for each of the key measures. The three case studies represent technologies on both the Milestones and TCDS development pathways and hence have different times-to-market, as illustrated in Figure 25 below; note that specific timings will depend on the scope of the programme.

Case study 1 details the evaluation of a software-based voice dictation system for sending text messages. This initial application of the toolset used the evaluation protocol outlined in chapter 9 and trialled video capture of eye glance behaviour with manual post-study coding of glance events. Case study 2 was structured as a comparative study between three different touchscreen technologies. Here, the focus was on the relative performance of the devices; in this case eye glance data was not collected as all three systems used the same graphical

interface (discussed in section 11.2). Both of these projects were conducted under the Milestones technology development process. Case studies 1 and 2 are described in full in submissions 4 and 5 respectively.

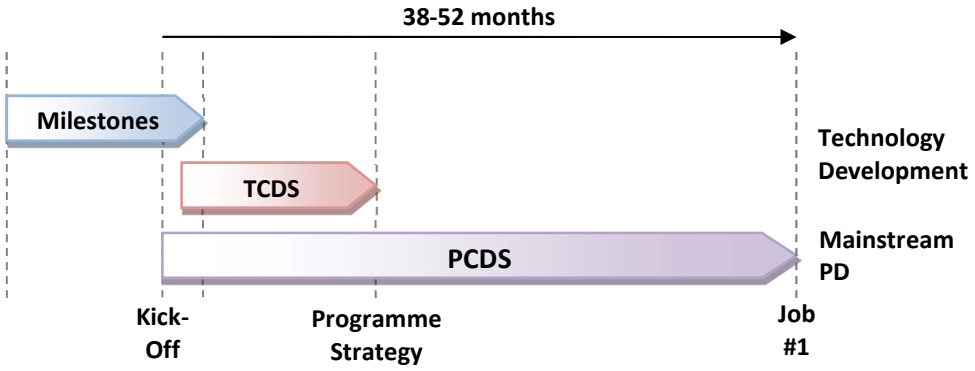


Figure 25 – Representative time-to-market for case study technologies

Case study 3 concerns the application of 3D displays to the automotive cockpit, and differs in approach to the previous case studies. The activities discussed here were conducted as part of a TCDS project, with the overall aim of effectively defining the design space in which 3D graphics could be applied in an instrument display, in terms of the depth of the 3D content. This required a more acute focus on specific aspects of human interaction, including stereopsis (the ability to perceive 3D images), depth perception and fatigue effects.

Case study 3 was conducted in three parts: the first focused on objective measures of performance, using the driving simulator to impose representative visual, cognitive and manual workload. This study offered an opportunity to trial an eye-tracking system for glance data collection. The second and third studies focused on users' responses to 3D stimuli and were conducted in a static condition using the simulator to provide context, utilising an alternative experimental approach. The three studies are reported in submission 6 and summarised in this chapter.

Chapter 13 presents a discussion of the case study findings in the context of the evaluation toolset, including revisions made to the protocol and the simulator.

10. Case Study 1 – Voice Dictation System

With smartphone use widespread, consumers are seeking to extend their connected lifestyle into the automotive environment. The use of mobile communication devices while driving is known to increase accident risk, with the sending of text messages shown to be the highest risk activity (Dingus et al., 2006), and UK legislation has made using a handheld mobile device while driving a specific criminal offence[2]. Vehicle manufactures are seeking to mitigate this issue by integrating smartphone functionality into the vehicle's in-built IVIS through standards including Apple CarPlay[3] and Android Auto[4], providing a degree of control over the quality of experience and the functionality available to the user. However, recent industry guidelines (NHTSA, 2013) seek to limit the extent of visual-manual interaction for non-driving tasks, in order to reduce distraction risk.

This poses challenges to vehicle manufacturers in terms of how to manage users' behaviour while providing desirable functionality and an enjoyable user experience. While SMS use in the UK peaked in 2012, text-based mobile communication has grown rapidly with the introduction of instant messaging apps such as Facebook Messenger and WhatsApp, with the latter handling 30 billion messages per day in 2015, compared to 20 billion sent by SMS (Sparkes, 2015).

Research has identified that visual tasks cause more interference with driving than auditory tasks (e.g. Dingus et al., 2011, Collet et al., 2010), and that manual interfaces cause more interference than auditory interfaces (e.g. Shutko and Tijerina, 2011, Dingus et al., 2011), especially those requiring visual feedback for data entry tasks (Wickens et al., 2013). Clearly, a visual/manual approach to text entry carries a high distraction risk. A potential solution therefore is to engage the auditory/vocal modality in the text messaging process. The Multiple Resource Theory (Wickens, 2008) indicates that separate processing resources exist for spatial and verbal tasks; and for visual and auditory modalities. Thus, by transferring the workload of text messaging tasks to the verbal/auditory domain, there is a reduced risk of interference with the predominantly spatial/visual driving task.

[2] The Road Vehicles (Construction and Use) (Amendment) (No 4) Regulations 2003, Regulation 110, (1) & (2).
[3] https://www.apple.com/uk/ios/carplay/
[4] https://www.android.com/intl/en_uk/auto/

10.1. Description of Technology

The case study focuses on a novel concept for editing and sending text-based messages while driving through voice interaction. This took the form of a functional early-stage software concept developed by JLR and deployed in a Windows PC environment. The key aim of the system is to minimise the visual and cognitive workload associated with text messaging. As such, the interface is designed to comply with draft NHTSA guidelines for Visual-Manual distraction with regards to the display of text, requiring that messages are limited to a maximum of 30 characters at a time and that scrolling text is prohibited. The concept splits messages into chunks of no more than 30 characters and uses text-to-speech to play back the content. A combination of auditory-vocal and visual-manual interactions are then used by the driver to manipulate the message content. A screenshot from the interface is shown in Figure 26 and a full description of system functionality is provided in submission 4, section 3.1.

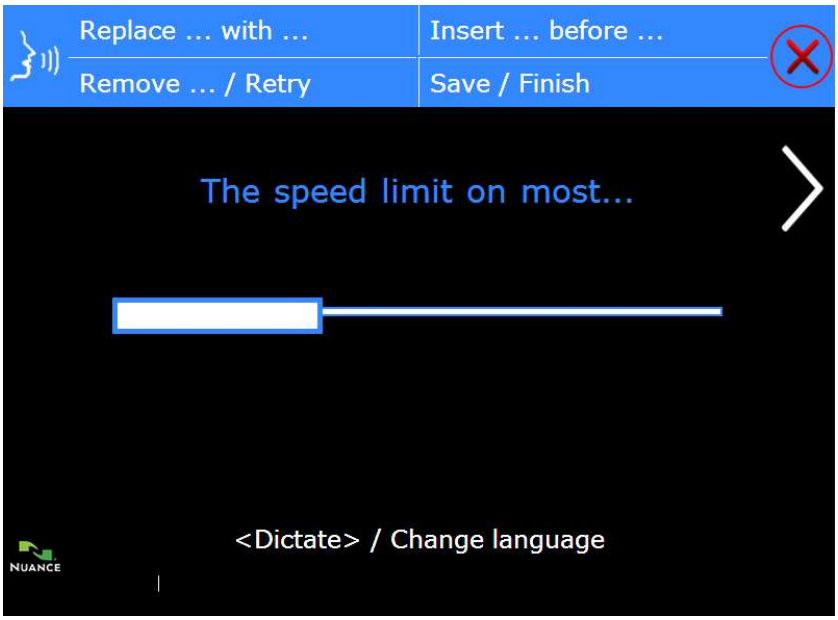


Figure 26 - Voice dictation system concept screenshot

Given that the study was not intended as a test of voice recognition software per se, a 'Wizard of Oz' approach was taken whereby the system's responses were pre-programmed and initiated manually by the experimenters, thus removing any variability in the performance of the voice recognition software.

10.2. Study Methodology

The research questions for the study were as follows:

1. *What is the effect of message formation and dictation on driving performance?*
2. *What is the effect of message error correction activity on driving performance and eye glance behaviour?*
3. *What is the effect of message content on message correction performance and eye glance behaviour?*
4. *What is the effect of error position on message correction performance and eye glance behaviour?*
5. *What are the relative workload requirements of message dictation and correction?*

In order to address these research questions the study was structured such as to disassociate the content creation and message correction aspects of the messaging activity. This was achieved by participants completing two experiment drives, one performing a content creation task and one performing a message correction task. For the creation task, participants were asked a series of simple questions regarding aspects of their daily lives that would elicit brief answers. The aim was to engage a degree of cognitive demand through memory recall. Users were asked 12 questions in total.

For the message correction task, the participant was presented with one of three pre-determined messages, shown in Table 2. The message was separated in to three sections of 30 characters or less and played back audibly using text-to-speech. An error was introduced into one of the message sections, requiring the user to correct the error and send the message using a combination of voice commands and button inputs, using the steering wheel buttons via the USB interface described in section 8.1.5. Participants completed twelve tasks: three error and one no-error condition for each of the three messages. The tasks are described in more detail in submission 4, section 4.3.2.

Table 2 - Message content and error locations

Correct message (Error locations shown in red)	**Length**	**Error 1**	**Error 2**	**Error 3**
There are **sixty** minutes in an hour, twenty-four **hours** in a day and seven **days** in a week	85	fifty	houses	years
The **speed** limit on most motorways in Great **Britain** is seventy **miles** per hour	74	weight	brain	meters
The Latin **alphabet** has twenty six letters, starting with **A** and ending with **Z**	74	elephant	D	H

A total of 32 participants were recruited for the study, with one data set lost due to technical issues. The remaining participants were aged between 20 and 59 years (mean = 37.3 years, SD = 12.2 years). All participants were UK drivers with a minimum of 1 year experience; nearly all (28) users had experience of using integrated and/or mobile infotainment (communications, media and navigation) devices in their cars. The majority of participants (24) were recruited from a database of existing participants within JLR. Approximately two-thirds (21) of the participants worked in a technical role.

The study was conducted using the protocol described in chapter 9.

10.3. Results

Task completion data indicates that sending a message containing an error took approximately 2.6 times as long as sending a message without an error (27.9s vs. 10.4s). A one-way ANOVA shows this result to be significant ($F(3,87) = 100.8$, $p<0.001$). The message itself did not have a significant effect on task time ($F(2,58) = 2.276$, $p=0.12$). A post-hoc Tukey HSD test indicated that when an error was present, its position did not have a significant effect on task completion time, illustrated in Figure 27.

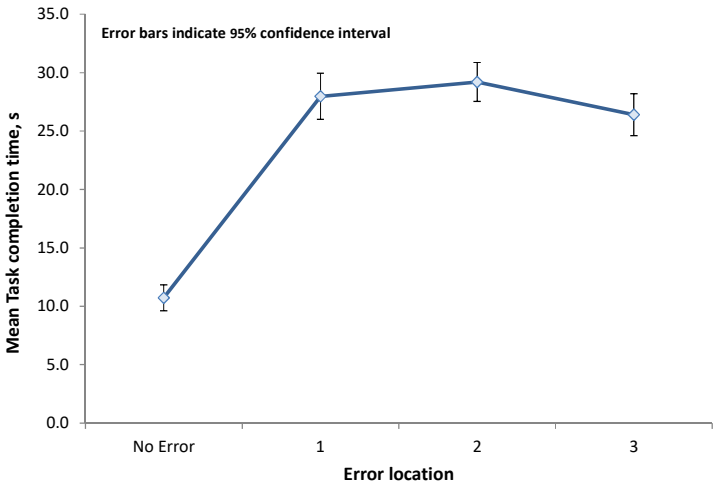


Figure 27 - Task completion time for correction task

For the correction tasks, the mean individual glance duration was 0.82s, with only 0.2% of glances exceeding 2.00s. Messages containing errors required approximately 3 times the total glance time (TGT) of messages without errors, but the message itself did not have a significant effect on the total glance time per task, as shown in Figure 28. When an error was present, its position did not have a significant effect on TGT per task.

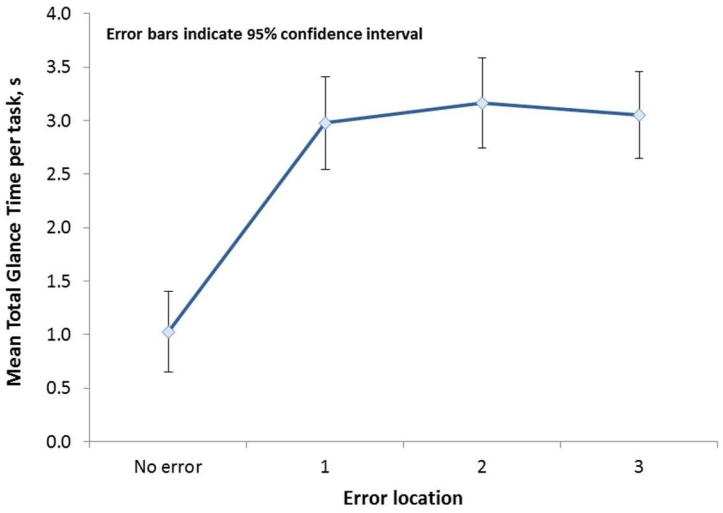


Figure 28 - Glance time per task for correction task

Figure 29 shows a comparison of SDLP measured in the baseline, dictation and correction conditions. A within-subjects ANOVA indicates a difference across the group ($F(2,27) = 5.18$, $p<0.01$); a post-hoc Tukey HSD test indicates that conducting both the driving and correction tasks did not significantly increase SDLP compared to the baseline measure, but there was a small but significant increase in SDLP for correction compared to dictation (mean 0.230 vs. 0.194 respectively). There was no significant increase in headway variation ($F(2,27) = 1.096$, $p=0.34$) or mean speed ($F(2,27) = 0.491$, $p=0.61$). Mean distance headway however increased significantly between the baseline and correction conditions (94.55m to 112.83m, $p<0.01$); this suggests an increase in cognitive workload imposed by the correction task.

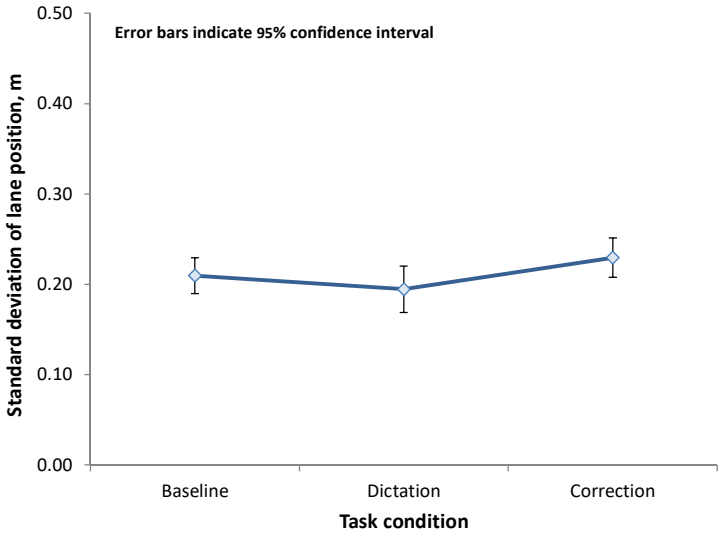


Figure 29 - Comparison of SDLP, case study 1

The overall NASA-RTLX workload measure was increased for both dictation (+19.5%) and correction (+44.3%) tasks relative to the baseline condition. Similar patterns of variation are observed for the individual subscale items, except performance and effort, which showed no significant difference to the baseline condition. The median hedonic rating was 7, corresponding to 'moderate' liking. The mean SUS score was 76.6 (SD = 12.8), with 100 indicating perfect usability.

The post-task interview revealed some valuable qualitative insights into the user experience of the concept. Users requested greater flexibility in their vocal inputs to the system, for example by allowing synonyms of key command words for a more naturalistic interaction. Some users were frustrated by the controlled flow through the correction task, citing issues with latency between the message segments and frustration with speed of message playback. Users identified three approaches that could improve this: 1) being able to interrupt the system with a verbal command; 2) swiping the scroll bar on the touchscreen to move through the message; and 3) being able to 'wrap round' from the end to the beginning of the message with a single (forward) button press. Users also suggested changes to the design of the interaction modalities: having the correction 'pick list' augmented with auditory feedback and allowing messages to be sent with a button press rather than a verbal command.

10.4. Study Conclusions

The study identified that using the message correction concept induced a degree of workload which was reflected in driving behaviour and identified in the subjective workload measures. Glance time data shows that visual workload is also relatively low, with almost all glances less than 2 seconds in duration. The hedonic rating and SUS scores indicate user acceptance of the system while qualitative findings suggest a number of design changes that may further improve the user experience.

11. Case Study 2 – Touchscreen Technologies

The touchscreen interface has become part of our daily technology experience, driven largely by the prevalence of the technology in smartphone and tablet devices. It was estimated that 97% of all smartphones would feature touchscreen interfaces by 2016 (ABI Research, 2011), with annual smartphone sales topping one billion units per annum in 2014 (IDC, 2014). This prevalence has driven an increase in automotive touchscreen applications, with shipments exceeding 50 million units in 2017 (IHS Markit, 2017).

Touchscreen interfaces offer a number potential usability benefits (Sears et al., 1992). Firstly, the flexibility of the interface allows multiple functions to be controlled with a single physical configuration. Input is direct as the display and input elements of the interface are co-located, enabling an intuitive mode of interaction that is easy to learn even for novice users (Wu et al., 2011). Touchscreen interfaces also provide the opportunity for vehicle designers to reduce the number of physical controls visible in the vehicle interior, creating a cleaner and less cluttered design, with the Tesla S providing a recent example.

The increased prevalence of touchscreen devices raises questions relating to the use of touchscreens in automotive applications, where the market is largely dominated by resistive touchscreens, an incumbent but arguably outdated technology. The aim of this study is to investigate the relative merits of resistive, capacitive and infra-red touchscreens in terms of their user experience.

11.1. Description of Technology

The study utilised three technology evaluator touchscreen devices developed by Continental AG. The devices consisted of a screen unit mounted in a common bezel, connected to a display controller driven by an embedded Linux device running a custom GUI application. This was designed to emulate a typical automotive user interface and featured a range of functionality commonly found in modern vehicles, including navigation, audio, telephone and text message functions; the tasks are discussed further below. The menu screen for the interface is shown in Figure 30.

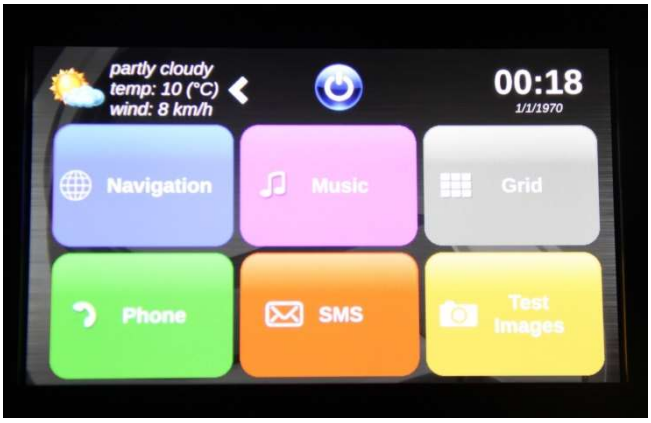


Figure 30 - GUI menu screen

The devices were designed to be visually identical, each with the same bezel and running the same user interface, with display brightness levels normalised. The display units were mounted in the centre console of the simulator vehicle cabin in the horizontal and vertical position of the original touchscreen; due to the size of the display bezel and the mounting arrangement, the touchscreen protruded slightly when fitted, as shown in Figure 31.

Figure 31 - Touchscreen installed in simulator cabin

11.2. Study Methodology

The study aimed to address the following research questions:

1. Can drivers differentiate between different types of touchscreen technology in use?
2. Do drivers display a preference for touchscreen technology types?
3. Is the performance of touchscreen tasks influenced by interface technology?
4. What are the key aspects of user experience for touchscreens in an automotive context?

The study utilised a 3x10 within-subjects experiment design, with display type and task type as the independent variables. Participants completed a baseline drive with no secondary tasks,

followed by three evaluation drives using one of the displays per drive. The experiment design was counterbalanced for display presentation order. The average duration of each drive was approximately 10 minutes.

Participants completed a total of ten touchscreen tasks per drive involving phone, audio, navigation and messaging functions. The tasks are described in detail in submission 5, section 4.3. Tasks were initiated by the experimenter via verbal instruction. Tasks were varied slightly across the three displays to avoid repetition but retaining the same number of inputs to complete (e.g. button presses).

A total of 20 participants were recruited for the study, 17 male and 3 female. Two male participants withdrew from the study due to the effects of simulator sickness. For the remaining 18, age ranged from 21 to 60 years (mean = 39.4 years, SD = 13.6 years). All participants were UK drivers with a minimum of 1 year experience; all had experience of using touchscreen devices and owned touchscreen-based smartphones, with 8 being regular users of factory-installed in-car touchscreen systems.

The study was conducted using the protocol described in chapter 9. Data from an earlier touchscreen-based study (Pitts et al., 2012a) indicated that touchscreen task completion time was directly correlated with glance time ($r=0.70$; $p<0.001$), indicating that visual workload will be reflected in the total task time. Given that the user interface, embedded hardware and task demands were identical for each device under test, visual workload was expected to be the same for each test condition; hence visual metrics are not reported.

11.3. Results

Due to a technical failure of the resistive display part way through the study, the data are effectively split into two sets, with twelve participants having experienced all three displays and six participants having experienced the capacitive and infra-red (IR) displays only. While this split must be enforced for all subjective data, it is legitimate to consider objective task completion time data across all 18 participants when comparing the capacitive and IR displays, as it is assumed that task performance using one device does not influence subsequent performance using another. The findings are therefore presented below with this structure in place.

Figure 32 shows the mean task completion time per task for the three displays experienced by twelve participants. Mean task time ranged from 9.1s (radio preset task, resistive display) to 60.9s (handwriting text entry task, IR display). A two-way within-subjects ANOVA showed that while interaction was significant ($F(18,198) = 1.93$, $p<0.05$), main effect for task did show significant differences ($F(9,99) = 58.34$, $p < 0.001$), but main effect for display type was not

significant (F(2,22) = 0.35, p = 0.71). Again, this is expected as the UI and tasks were identical across all three devices.

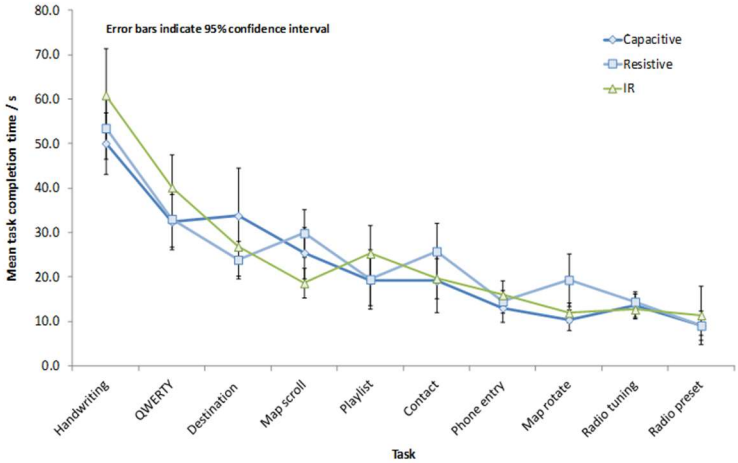


Figure 32 - Mean task completion time (3 displays, 12 participants)

Figure 33 shows the mean Standard Deviation of Lane Position (SDLP) for the baseline and task conditions. SDLP increased by 78% from 0.21m to 0.40m when users were performing touchscreen tasks while driving (F(30,3) = 15.828, p<0.001); however post-hoc paired t-tests indicate that no significant difference was observed between the different touchscreen types (p > 0.3 for all pairs).

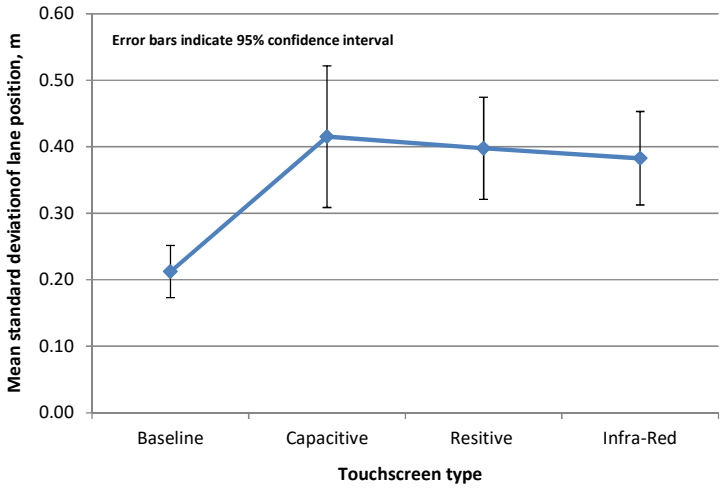


Figure 33 – Standard deviation of lane position (3 displays, 12 participants)

A similar pattern was observed for longitudinal control, as shown in Figure 34: mean time headway increased by 73% from 2.2s to 3.7s when touchscreen tasks were introduced, compared to the baseline driving condition (F(30,3) = 8.673, p<0.001). This is a reflection of the cognitive workload required to complete the tasks while driving, with the increase in headway to the lead vehicle serving as a coping mechanism for the user. As with SDLP, no significant difference in headway was observed between the touchscreen types. It is notable

that there is large variance in the data, as shown by the error bars in Figure 34. This is partly due to the small sample size (only 12 participants experienced all three displays), but may also be due to variations in vehicle following strategy between participants.

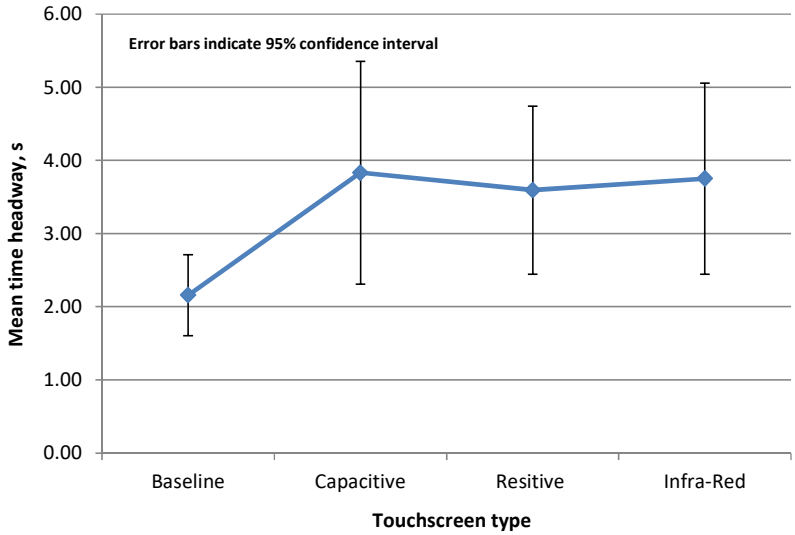


Figure 34 - Mean time headway (3 displays, 12 participants)

Preference data indicates that the resistive screen is clearly least preferred, with no significant difference between the IR and capacitive screens. Hedonic rating data shows a significant difference between all three displays, with a mean rating of 5.7 versus 3.0, where a score of 6 equals 'like moderately' and 3 equals 'Dislike slightly'. No significant difference was found across the displays in the SUS scores ($F(2,11) = 2.20$, $p = 0.124$). The mean SUS score was 80.83, 73.33 and 62.50 for the capacitive, IR and resistive displays respectively.

Figure 35 shows the subjective overall workload and frustration scale data for all three displays across 12 participants, with the reported workload for the baseline drive included as a fourth variable level. While workload was significantly increased by the introduction of touchscreen tasks relative to the baseline ($F(3,33) = 21.20$, $p<0.001$) there was no significant difference in workload across the display types. Post-hoc analysis of the 'frustration' sub-scale of the NASA-TLX questionnaire using the Tukey HSD test shows that using the resistive display was significantly more frustrating than the capacitive ($W = 1.5$, $p<0.005$).

Analysis of the qualitative data confirmed the preference findings shown above: the resistive display was strongly disliked, with users citing a lack of sensitivity and perceived latency in response. The resistive display required a higher input force compared to the other two, which affected the experience of scrolling tasks - in particular, there was a tendency for a list scroll gesture to result in an inadvertent selection of the list item under the user's finger. In addition, the resistive display featured a different surface feel, which was described as 'papery' and 'cheap'.

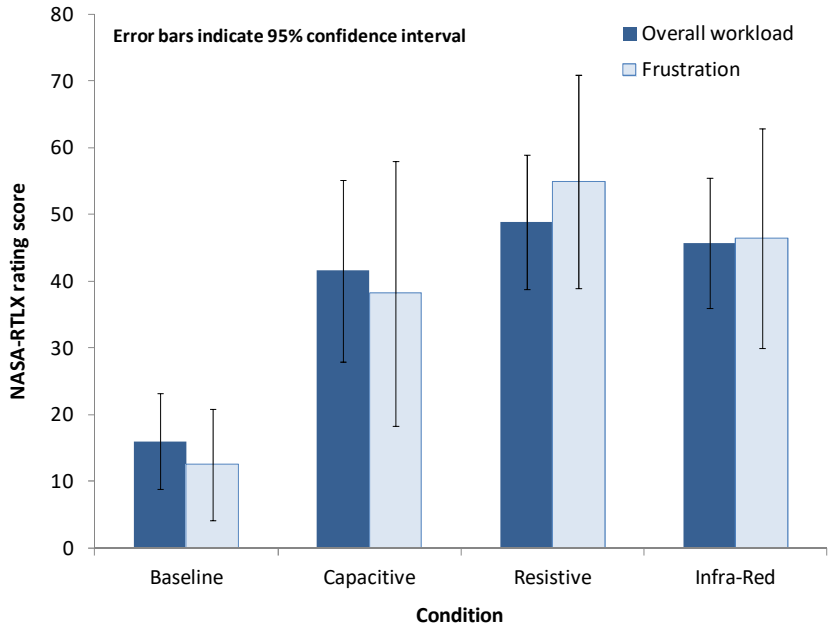


Figure 35 – Subjective ratings of overall workload and frustration (3 displays, 12 participants)

The comparison between the capacitive and IR displays was less distinct, with 5 of the 12 users that experienced all three displays stating that they could not identify a clear difference between the two. Users cited an increase in responsiveness leading to improved confidence in the system and a perceived decrease in response time relative to the resistive display.

However, 11 of the total 18 users believed that the IR display was less responsive than the capacitive. Specifically, buttons in the lower-right corner of the display frequently failed to register inputs as expected. Indeed, while there were no strong reasons for disliking the capacitive display, the IR display was identified by some users as lacking sensitivity, with some tasks requiring multiple repeat button presses to complete. One user identified a potential increase in driver distraction arising from the additional attentional demands; and another found the display 'frustrating', comparing it to an inexpensive aftermarket navigation system.

11.4. Study Conclusions

The study identified that the resistive touchscreen technology widely used in current vehicles does not meet users' expectations for User Experience. The main issue was a lack of responsiveness from the display which was both interpreted as added latency, and caused repeated errors for tasks that involved 'drag' gestures, such as playlist and contact list scrolling. Both infra-red and capacitive displays offer significant improvements over resistive; however, it is the capacitive display that provides the best performance in terms of responsiveness and affective response, with the IR display exhibiting some specific usability flaws. The findings suggest that the ubiquity of touchscreen interfaces has led to users' expectations of touchscreen behaviour to be driven by their experience of mobile devices, and

that touchscreen usability has become a ‘must-be’ attribute. Given the importance of in-vehicle technology to the perception of quality of the vehicle, manufactures must carefully consider the display technology employed for future vehicles.

The task completion data showed that there was no significant difference between the three displays in terms of task workload. This is not unexpected, due to the underlying similarities between the systems. As stated in section 11.2, the assumption was made that these similarities would also negate differences in the visual workload. While the task completion data would support this assumption, it is not possible to state for certain that differences do not exist without analysis of glance behaviour. The absence of this data is a limitation of this case study; it is recommended therefore that glance data is collected for all future studies.

The findings from this study also illustrate how the qualitative data can substantiate effects identified by the qualitative measures. While differences were observed in the hedonic rating and workload measures that suggested preference, the nature of these differences was characterised though the interview responses of the participants, highlighting issues with perceived responsiveness, confidence and even sensory quality that would not have been identified though purely instrumental or analytical approaches.

12. Case Study 3 – 3D Displays

Stereoscopic 3D displays have become more prevalent over recent years as the market for 3D televisions has developed. The gaming industry has also seen a growth in 3D display devices and content, with improvements in graphic processing capability allowing stereoscopic gaming at previously unavailable resolutions and frame rates. Recently, new virtual reality devices such as Oculus Rift have provided developers with new tools to create immersive 3D experiences. Beyond the entertainment sector there are numerous applications in which stereoscopic displays have proven benefits. Examples include flight simulation, medical imaging, remote inspection of hazardous environments and design visualisation (IJsselsteijn et al., 2005). The use of 3D displays for IVIS in road cars however is a developing field. It is important to understand how users will respond to new in-vehicle technologies and to determine the design space within which developers can work to provide a high-quality user experience.

12.1. Case Study 3a – 3D Display Effectiveness

As discussed in section 9.4 three discrete studies were conducted into 3D display technologies. The first of these was designed to provide a validation of the potential for the use of 3D displays within a vehicle environment and was conducted in accordance with the protocol described in chapter 9

12.1.1. Description of Technology

The 3D display used for study 1 was a 10.1" parallax barrier autostereoscopic demonstrator unit. This was mounted in front of the driver, in the approximate location of the standard instrument cluster. Figure 36 shows the position of the 3D display mounted in the simulator cabin. The display device features an embedded Android controller which is used to drive the display and serve graphical content. The display of images was controlled by the experimenter from a remote PC over Wi-Fi using a custom application written by a JLR developer in C++.

Visual glance data was collected using a Dikablis head-mounted eye tracker[5], worn by participants during the trial drive. The coded targets seen in Figure 36 are used to calibrate the eye tracker.

[5] http://www.ergoneers.com/en/hardware/eye-tracking/

Figure 36 - 3D display mounted in simulator cabin

12.1.2. Study Methodology

There are multiple applications where image depth can be used to convey information to the driver, such as navigation and hazard perception. Study 3a was designed to determine users' ability to identify depth-encoded target images when presented in forced-perspective 2D (known as 2.5D (McIntire et al., 2014)) and stereoscopic 3D formats. The hypothesis was formed that 3D presentation would assist the user in the identification task.

The research question for study 3a was:

> *"Can autostereoscopic displays relay depth-encoded information more effectively than a 2D display, in the driving context?"*

Participants were shown an image consisting of three toroidal objects arranged on a vertical plane at zero parallax and labelled A to C, shown in Figure 37 – note that target C is displaced in positive parallax. The images were created by a JLR developer using a 3D modelling graphics package and then exported in raster format. In the test condition, one of the objects was displaced from the others in positive parallax, i.e. away from the driver. The participant was then asked to identify which, if any, object was displaced. Two levels of parallax offset were used (low/high), with each object offset once at each level; a zero offset condition was also included. It addition, the images were rendered in both stereoscopic 3D and 2.5D. In total, 14 experimental conditions were used; a randomised complete block experiment design was utilised whereby all participants experienced each condition once.

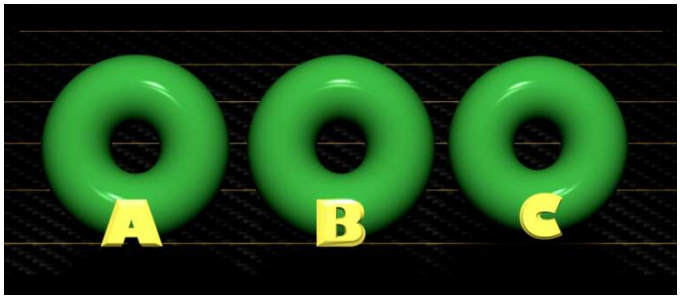


Figure 37 - Target identification task screen shot

The display tasks were completed while engaged in the simulated driving task. Measurements of task completion time, correct response rate and glance time were collected.

27 participants were recruited for the study, 24 male and 3 female. All participants were UK drivers with a minimum of 1 year experience and had normal or normal-corrected vision.

12.1.3. Results

The driving data was compared using a paired t-test (2-tailed) between the baseline and task conditions. No significant differences were observed for SDLP ($t(21) = 1.137$; $p=0.26$), mean distance headway ($t(21) = 1.642$; $p=0.12$), or distance headway variation ($t(21) = 0.073$; $p=0.15$). Comparisons of the 2D and 3D task sections of the driving data similarly showed no significant differences (SDLP: $t(21) = 1.23$; $p=0.23$; mean headway: $t(21) = 0.140$; $p=0.89$; headway variation: $t(21) = 1.525$; $p=0.14$).

The findings identified differences in the correct response rate between the different experimental conditions. A two-way within subjects ANOVA analysis was performed: interaction between render type and target depth offset was significant ($F(2,52) = 12.48$, $p < 0.001$), with significant main effects for both factors ($F(1,26) = 12.235$, $p < 0.005$; $F(2,52) = 16.649$, $p < 0.001$ respectively). Simple main effects analysis shows that significant differences exist only at the 'low' depth offset level ($F(1,26) = 25.490$, $p < 0.001$), with mean correct response rates of 0.48 and 0.85 for 2D and 3D respectively. These findings are illustrated in Figure 38 below.

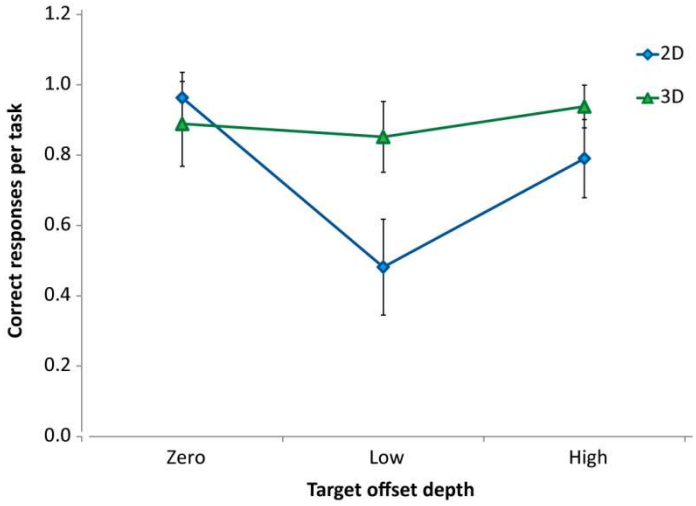


Figure 38 - Correct response rates for target identification task.

The total glance time per task, defined as the sum of the durations of individual glances to the 3D display, was calculated and the data analysed using two-way ANOVA. Interaction was again significant ($F(2,52) = 6.059$, $p < 0.005$), as was the simple main effect for image depth for the 3D condition ($F(2,52) = 12.055$, $p < 0.001$). A post-hoc Tukey's HSD test showed that the low and high parallax 3D images required over 1 second shorter TGT per task than the zero-parallax images (mean 1.43s vs. 2.45s). These results are illustrated in Figure 39.

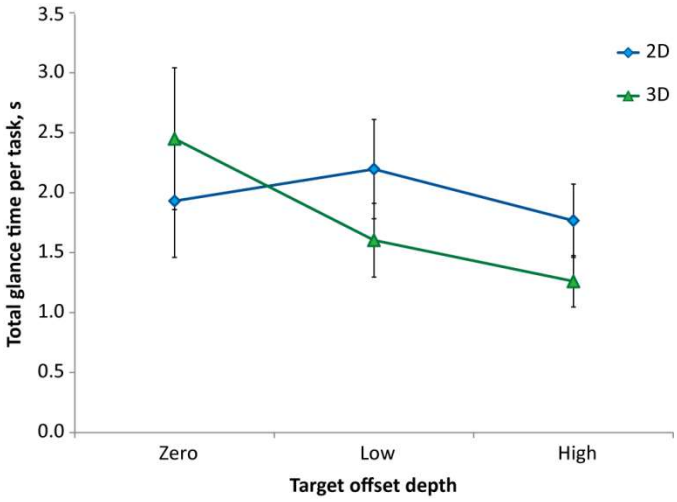


Figure 39 - Total glance time per task.

The task completion time data indicated that only the main effect for image depth was significant for the 3D condition $F(2,52) = 12.113$, $p < 0.001$), with post-hoc tests indicating a difference between overall mean task completion time for the high and zero parallax images (1.92 vs. 3.23s).

12.1.4. Study Conclusions

The findings indicate that content certain levels of depth-encoded information can be interpreted more accurately when rendered as stereoscopic representations, compared to 2.5D rendering. Furthermore, 3D-rendered content which extended beyond the convergence plane was identified with less eyes-off-the-road time than zero-parallax content. This indicates that stereoscopic displays have potential safety benefits by assisting drivers in processing depth information for critical applications. Participants were also able to more accurately identify content with 'low' levels of parallax offset when presented in the 3D condition, indicating that the ability to discern fine details in depth information is assisted by stereoscopic presentation. It is important however to undertake further studies to determine how image perception is affected by changes in parallax offset levels.

12.2. Case Study 3b – 3D Display Depth Limits

The level of depth featured in stereoscopic content is related to the quality of the perceived image and to the degree of visual discomfort experienced by the viewer. Most people are able to comfortably view stereo 3D content at low levels of parallax without negative effect. However, as the level of parallax offset increases, the viewer's visual system has to work harder to maintain the sharpness and fusion of the stereoscopic image, creating a trade-off between image quality and discomfort. At higher levels of parallax, the viewer will experience a blurred image, a loss of fusion resulting in double vision, or both (Lambooij et al., 2009).

Lambooij et al. define visual fatigue as "physiological strain or stress resulting from exertion of the visual system"; and continue to identify that, while the terms 'visual fatigue' and 'visual discomfort' are used interchangeably in the literature, the former refers to the degradation of performance of the visual system which can be measured objectively, while the latter is the subjective perceptual effect experienced by the subject. It is assumed therefore that subjective measurements of visual discomfort will serve as an analogue for objective measurements of visual fatigue.

The aim of this study was to establish limits of acceptability for image parallax offset, based both on image quality and visual discomfort. An alternative experiment protocol was used, applied in a static (non-driving) condition, using the simulator cabin for context.

12.2.1. Equipment

The display used for the study was a 13.3" lenticular lens-type autostereoscopic display unit from SeeFront[6]. This display utilises a built-in tracking system to identify the position of the viewer and adjust the display parameters to optimise the 3D effect at the viewing position; this provides a 3D viewing distance range of 310mm. The display was mounted in driving simulator cabin in front of the driver, as per study 3a. The display was connected to a laptop which output the image content over DVI.

12.2.2. Participants

A total of 22 participants were recruited for the study, 19 male and 3 female. As previously, all participants were UK drivers with a minimum of 1 year experience and had normal or normal-corrected vision. Five participants owned 3D TVs at home and one used a gaming device with an autostereoscopic display (Nintendo 3DS).

12.2.3. Study Methodology

The research question for study 3b was:

> *"At what depth level can 3D content be presented without compromising image quality and viewer comfort?"*

Lambooij et al. (2009) identify that users compensate for image disparity through the accommodation-vergence system, with higher levels of accommodation effort related to higher levels of visual discomfort; and that discomfort is increased with prolonged viewing, i.e. for autostereoscopic 3D images there is a trade-off between image quality and visual effort. Users can 'force' convergence of a 3D image by applying more effort over a longer time, with a resultant increase in discomfort. To exclude images that could not be resolved within a typical glance in an automotive use case, the experiment constrained users' glance durations using occlusion spectacles (see submission 6 section 4.1.3 for more details). Shutter open time, the duration for which the participant can see the 3D display, was set to 1.5 seconds per image in accordance with SAE J2364 and ISO 16673:2007 (SAE, 2004, ISO, 2007).

The study was conducted in the driving simulator cabin as per study 3a, but in a static (non-driving) condition. Participants were shown a series of images created by a JLR developer, consisting of two instrument dials with the left-hand dial set at zero parallax and the right-hand dial offset in the positive or negative direction. Participants received a verbal notification of image onset by the experimenter then the occlusion shutters were opened for the set period. Participants were then asked to indicate whether the right-hand dial was located closer in space

[6] http://www.seefront.com

(negative parallax) or further away (positive parallax) than the left-hand reference dial. An example image is shown in Figure 40; note that the two halves of the image representing the left and right-eye views are divided and scaled by the 3D display to provide full-screen stereoscopic content with 16:9 aspect ratio.

Figure 40 - Example image from Study 3b

A total of 31 images were used: 15 each of positive and negative parallax offset. Parallax offset was defined in percentage of convergence plane distance (%CPD); where the CPD is the distance from the viewer to the zero parallax plane, i.e. the screen surface. For example, at a viewing distance of 50cm, -10% CPD would see an image appear 5cm in front of the zero parallax plane. The images used in the study ranged from -15.0 to +22.5 %CPD, spanning a range that extended beyond the boundaries of comfortable viewing, including a zero parallax condition. Image presentation order was randomised and reverse balanced, such that each image was seen twice.

Participants were asked to provide a rating of visual discomfort after the presentation of each image. As the rating would be conducted in-task, a simple approach was adopted based on the wording of the Simulator Sickness Questionnaire (SSQ) (Kennedy et al., 1993b). Bouchard et al. (2007) determined that seven of the sixteen sub-scale items of the SSQ form an underlying factor relating to oculomotor performance: fatigue, headache, eyestrain, difficulty focusing, difficulty concentrating, fullness of head and blurred vision. As it was not practical to rate these items individually during the experiment, participants were instructed to rate 'overall visual discomfort', encompassing all of the above aspects. The rating scale used in shown in Table 3.

Table 3 -Visual discomfort rating scale

Adjective rating	Rating score
None	0
Slight	1
Moderate	2
Severe	3

Participants also provided a rating of image quality for each task, comparing the right-hand (offset) dial to the left. The quality rating scale was based on the adjectival categorical judgement method for stimulus comparisons defined in section 6.2 of ITU-R BT.500-13 (International Telecommunication Union, 2012), with slight modifications to the wording to emulate the wording of the SSQ-derived visual discomfort scale described above. The image quality rating scale wording is shown in Table 4.

Table 4 - Image quality rating scale

Adjective rating	Rating score
Much Worse	-3
Moderately Worse	-2
Slightly Worse	-1
The Same	0
Slightly Better	+1
Moderately Better	+2
Much Better	+3

12.2.4. Results

Figure 41 shows the median image quality ratings for each parallax offset position. Median image quality degrades as parallax level increases in both the positive and negative directions. A non-parametric Kruskal-Wallis test was applied to the data which indicated that the difference across all measurements was significant ($\chi^2(30) = 495.5$, $p < 0.001$). Post-hoc Wilcoxon signed ranks tests with Bonferroni correction identified that a full point of median quality degradation was observed at -4.0 and +4.5% CPD. In general, negative parallax images were subject to a greater degree of degradation than positive parallax images. Order effects across the data set were shown to be non-significant ($\chi^2(61) = 60.9$, $p = 0.48$).

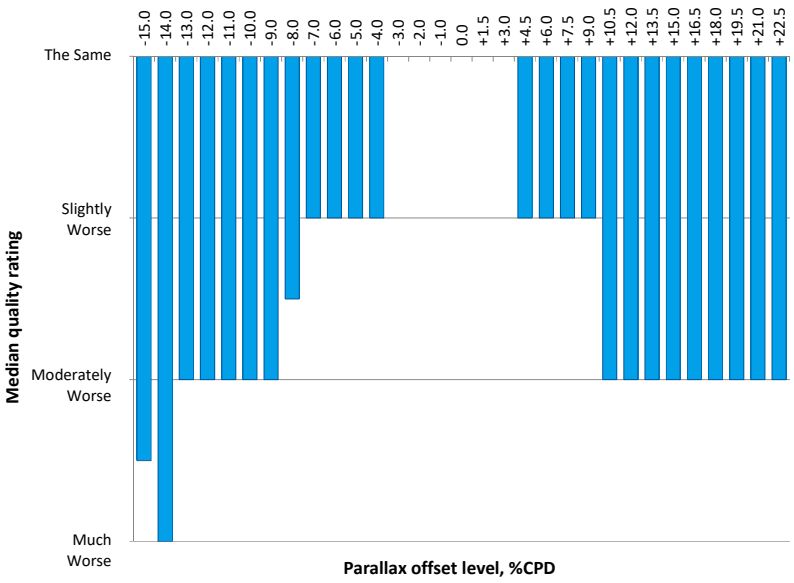


Figure 41 - Median image quality ratings by parallax offset level

84% of all reports of visual discomfort were 'slight'; while reports of higher levels of discomfort did occur for images at higher parallax offset levels, these were not statistically significant. However, regression analysis showed a weak but significant trend for increased frequency of reported discomfort with increased exposure ($F(1,61) = 115.6$, $p < 0.001$).

Finally, the number of times participants reported zero image offset was calculated. While some participants were not able to perceive the 3D effect at low parallax levels, this effect was not statistically significant.

12.2.5. Study Conclusions

The findings of study 2 clearly show that users will see a reduction in image quality as the level of 3D offset is increased, even for relatively small offset values. The limits identified in this study were carried forward into the design for study 3c.

The trend identified in visual discomfort rating provides evidence of an important issue for implementation of 3D technology in vehicles. While real-world use cases would likely feature both a lower frequency of exposure and less 'demanding' 3D content, i.e. lower parallax levels, OEMs must be aware of the potential for cumulative effects of viewing 3D content over prolonged periods of time. Indications were also present of a threshold effect for perception of depth in 3D images. While not statistically significant in this study, further investigation is warranted.

12.3. Case Study 3c – 3D Display Depth Resolution

The final study conducted was designed to investigate the minimum levels of parallax offset required to create a perceptible 3D effect. As described above, evidence was found in study 3b of a perception threshold effect; it is important therefore to understand where the lower limits of the design space are. Two hypotheses were formed: that there is a relationship between the positions of objects in depth and the amount of offset required to create a discernible depth difference; and that this effect would vary based on the type of image object. The study was designed to test these hypotheses using representative automotive graphical content.

12.3.1. Equipment

The experimental setup for study 3c was the same as for study 3b, with the display mounted in the driving simulator cabin in the position of the instrument cluster and the participant's vision constrained using occlusion glasses, with shutter open time set to 1.5s per image.

12.3.2. Participants

30 participants were recruited for the final study. 5 data sets were rejected due to data quality and 1 participant was withdrawn due to technical issues. Of the remaining 24, 18 were male and 6 female. Again, all participants were UK drivers with a minimum of 1 year experience and had normal or normal-corrected vision.

12.3.3. Study Methodology

The research questions for study 3c were:

> *"What is the minimum difference in 3D position between two objects required to perceive a difference in depth between them?"*
>
> *"Does the minimum difference required vary with starting position and object type?"*

Participants were shown a series of images created by a JLR developer, based on one of three image types: dials, icons or text; examples of these are shown in Figure 42. Note that for icon and text images, the dummy dials shown were fixed at zero parallax. As in study 3b, each image consisted of a pair of objects. In this case however, both objects were set to a nominal parallax level, in a range that extended just beyond the image quality limits determined in study 3b.

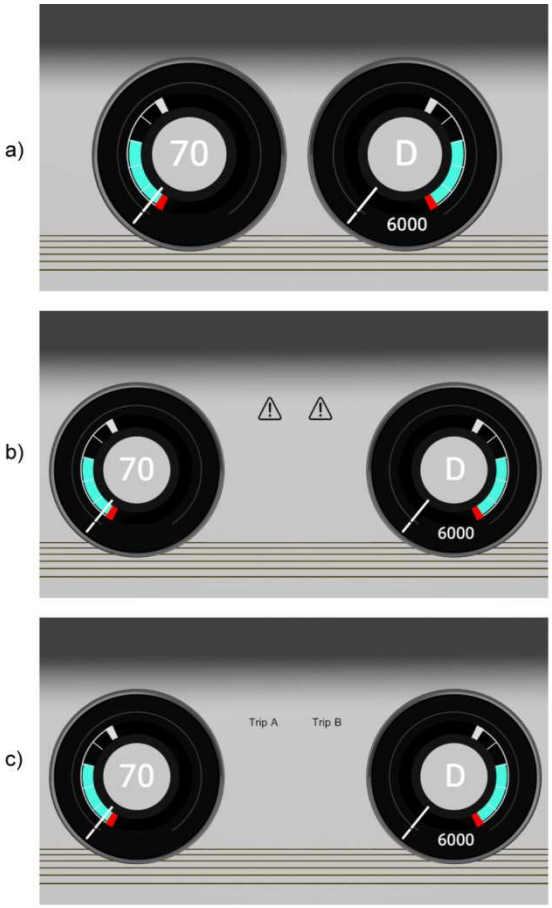


Figure 42 - Study 3 image examples: a) dials; b) icons; c) text

An additional parallax offset of up to ±1.5% CPD was applied to the right-hand image in the direction of the nominal offset; 1.5% CPD was experimentally determined to be a clearly-perceptible offset level. Both positive and negative offsets were applied from zero nominal parallax, hence inclusion of both +0.0 and -0.0 categories.

The experiment utilised a 'staircase' approach to determine the minimum required depth offset, whereby for each nominal condition, participants were shown a sequence of five images from a set of 16 starting at either zero or maximum parallax offset. Participants were required to state whether the right-hand object was nearer, the same or further away than the left-hand object. The next image presented was determined by the participant's response: offset was either increased or decreased until the minimum perceptible offset was identified. In this way, the image type and nominal parallax form the independent variables, while the minimum offset forms the dependent variable.

8 nominal parallax levels between were applied for each of the three image types, giving 24 conditions. Presentation order of conditions was determined using a counterbalanced 24x24 Latin Square design and initial offset direction was balanced across all participants.

12.3.4. Results

Minimum offset data for each nominal parallax position and image type were collated. Responses were ignored where participants stated that they could see an offset when none existed, or that they could not see an offset when the maximum level was applied; this comprised 29 of 576 responses (5%). Figure 43 shows the outputs from the study, in terms of the minimum parallax offsets required at each nominal parallax position to create a perceptible 3D depth difference between two objects, for each of the three object types tested. Error bars indicate 95% confidence interval.

A linear mixed model analysis approach was implemented using SPSS. This indicated that the relationship between nominal parallax and minimum perceptible offset varies across image types: interaction between nominal parallax and image type is significant ($F(14,523) = 2.983$, $p < 0.001$). Differences across nominal parallax levels are significant ($F(7,523) = 7.700$, $p < 0.001$), as are differences across image type ($F(2,523) = 13.052$, $p < 0.001$). However, post-hoc pairwise comparisons of image type indicate that there is no difference between icons and text (mean = 0.50, $p = 0.75$), with dials being significantly different (mean = 0.38, $p < 0.05$). Hence, we can consider the image types in terms of 'large' and 'small' objects.

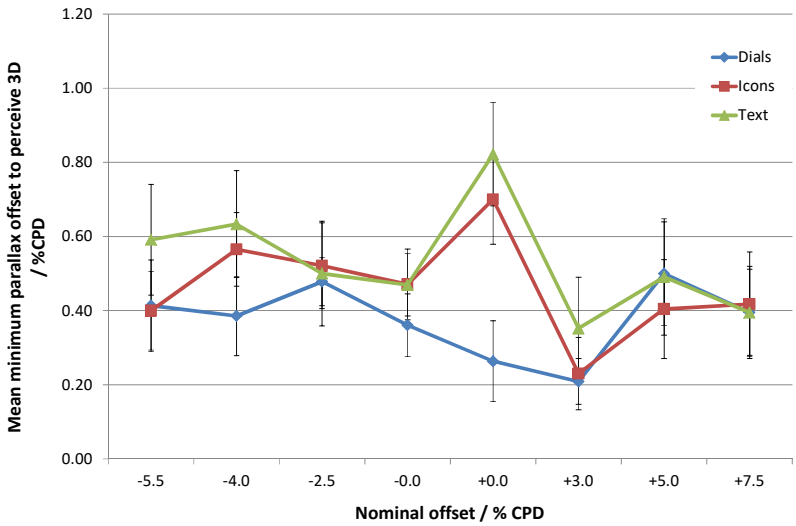


Figure 43- Minimum parallax offset levels to perceive 3D effect

The notable difference at +0.0 suggests an effect due to the design of the image stimuli. As show in Figure 42, the dial images featured two large objects presented in close proximity with a common nominal parallax. Icons and text however featured two small objects presented in proximity at the horizontal centre of the display and at a common nominal parallax, with dummy dial objects positioned on the left and right of the display at zero parallax. In the +0.0 condition, the left-hand text/icon image sits level with the larger objects; it is possible therefore

that in this case attention is drawn back to the reference plane, making it more difficult to identify when the smaller objects were offset in positive parallax.

Figure 44 below simplifies the findings from study 3c as reported in Figure 43. The data are consolidated into two series for 'large' and 'small' objects, as discussed above. The values are then normalised across the series where no statistical differences exist. In addition, the findings from study 3b are applied to create quality degradation zones, whereby the median quality difference from the zero parallax condition is greater than one scale point; i.e. a 'more than slight' degradation in image quality.

It can be seen from this figure that the minimum parallax offset for object differentiation varies for large and small objects, and for positive and negative parallax. Large objects require at between +0.25% and +0.43% CPD to create a 3D offset in positive parallax; and between -0.40% and -0.47% CPD to create an offset in negative parallax. Similarly, small objects require between +0.25% and +0.75% CPD to create an offset in positive parallax; and between -0.47% and -0.59% CPD to create an offset in negative parallax. Parallax offsets outside the range $-5.0 \leq x \leq +6.0$% CPD will result in a notable degradation of image quality.

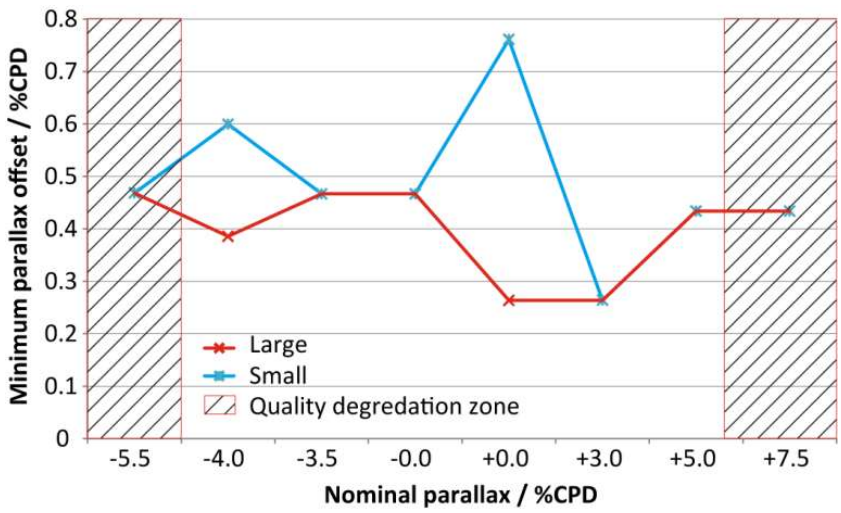


Figure 44- Design parameters for acceptable image quality and minimum offset for small and large object types.

12.3.5. Study Conclusions

The findings of study 3c confirm that the levels of parallax offset required to create a perceptible 3D depth difference between two objects is related to their position in the 3D space, the size of the object and its proximity to other objects that may serve a reference. In general, a lower degree of offset is required for large objects such as dials, compared to small icons or text. However, there is an apparent interaction between large and small objects when they are aligned on the same plane, an important consideration for interface design. By combining these findings with the results from study 3b the upper and lower limits of the

design space have been defined in terms of 3D effect perception and image quality, for both large and small objects.

Designers must consider the compatibility of 3D display technologies with their customer base. Between 5 and 10% of adults are affected by stereoblindness, which would render the display ineffective. Visual abilities will also degrade with age, meaning that older adults will be less sensitive to stereoscopic depth cues (IJsselsteijn et al., 2005). The display system could therefore be designed to include a degree of control over the 3D depth of the content, including the option to disable it. While it may not be able to replicate the UX of a 3D interface in 2D, it is important that a lack of 3D content does not adversely affect the safety of the system.

The methodology applied to studies 3b and 3c was designed to answer questions relating to specific aspects of the user-centred performance of the technology. While differing from the protocol outlined in chapter 9 and applied to case studies 1 and 2, certain elements were carried forward; specifically: the use of the simulator cabin to provide a context-specific evaluation environment and the derivation of the visual discomfort scale from the SSQ.

13. Discussion of Findings

The case studies presented in the preceding chapters provided valuable insights that illustrated the performance of the technology concepts from the users' perspective, while addressing key functional capability questions. The three technologies investigated varied in terms of their architecture, technology readiness and development timescale, encompassing both hardware and software-based concepts. In each case, the evaluation toolset was shown to be suitably flexible to support the evaluation, producing high-quality data. This chapter discusses the findings of the case studies, in terms of the application of the toolset, data quality, qualitative learning and revisions to the toolset.

13.1. Differences in Toolset Application between Case Studies

As noted in section 9.4, the evaluation and analysis approach varied across the three case studies. While case studies 1, 2 and 3a largely followed the toolset protocol, the specific requirements of the technology development project required that an alternative approach was taken to studies 3b and 3c. Here, the aim was to define the design space for 3D content, determined by users' ability to perceive content clearly and without discomfort and required an abstracted approach that negated the use of the dual-task simulator-based approach earlier defined.

The question remains, however, of how well the 3D display technology may be received by the user in a final product. Image clarity and visual discomfort will have an obvious effect on the UX of the system; additional testing conducted once representative content and tasks are available would again provide the detailed and valuable insights into the 3D display technology that are necessary to determine user acceptance. In this case, the evaluation toolset could be successfully applied; clearly then, there are limits to the applicability of the toolset based on technology readiness. This also demonstrates the value of an iterative evaluation approach whereby evaluation outcomes are used to further improve the concept.

Case study 2 provided comparisons across three candidate technologies that have potential to be used on a vehicle programme within a 1-3 year time frame. The findings from this study, that the incumbent (and most cost-effective) technology provided significantly poorer UX compared to the other candidates, provide clear and concise evidence to support product development decisions. Furthermore, the identification of specific usability differences between the remaining two candidates justifies the inclusion of the qualitative data element; without it, these subtle differences would have been lost.

There were also differences in application with respect to eye glance measurement. Case study 1 used a manual coding approach to identify eye glance events from video captured during the

study. While effective, this approach is labour-intensive and prone to variability. For repeatability, multiple, skilled coders should be involved in the process: this reflects the practice observed at MIT AgeLab, where a pool of part-time workers (mainly postgraduate students) are employed to manually code video from simulator and instrumented vehicle studies. Conversely, case study 3a employed automated glance identification using an eye-tracking system. While this simplifies the analysis, the system employed in this case required the user to wear tracking glasses, which can be intrusive and may affect the behaviour of the user during the study. The approach taken at University of Michigan Transportation Research Institute (UMTRI) was to use a remote eye-tracking system permanently installed in the simulator cabin – this is less intrusive to the user but requires careful setup and calibration to ensure data quality. Both eye-tracking systems also carry a significant associated cost for the hardware and software.

13.2. Data Quality

To act as an effective research tool, a questionnaire must exhibit both reliability and validity (Adams and Cox, 2008). Reliability relates to the consistency of the measure and validity relates to the ability of the instrument to measure the phenomenon that it is designed to measure (Nunnally and Bernstein, 1994). The subjective measures selected for inclusion into the toolset have all been tested for reliability and validity in the course of their development, but it is important to evaluate them as applied within the case studies. In addition, the consistency of the driving data from the simulator should be reviewed. This section addresses the reliability, validity and consistency of the data acquired by the toolset.

13.2.1. Usability and Workload Measures

The validity of both the NASA-TLX workload and SUS usability rating scales is well established in literature, forming a strong part of the rationale for their inclusion in the toolset. Reliability is commonly tested using Chronbach's Alpha (Cortina, 1993, Nunnally and Bernstein, 1994). Chronbach's Alpha is a measure of internal consistency, and ranges from 0 to 1, where 1 indicates perfect reliability. Typically quoted 'acceptable' values for Chronbach's Alpha range from 0.7 to 0.95 (Tavakol and Dennick, 2011).

Chronbach's Alpha is a function of the number of test items and the average inter-correlation among the items for a given scale:

$$\alpha = \frac{N\bar{c}}{\bar{v} + (N-1)\bar{c}}$$

Where N is the number of items in the scale, $\bar{c}$ is the average inter-item covariance amongst the items and $\bar{v}$ is the average variance of the items.

Table 5 shows the Chronbach's Alpha coefficients for the SUS data collected in case studies 1 and 2, calculated using SPSS. The overall mean value of 0.826 shows acceptable reliability, as expected. This both verifies earlier reliability values reported in the literature and confirms that the instrument was correctly applied in the case studies.

Table 5 – Chronbach's Alpha coefficients for SUS scores

	Experimental condition				
Study	**Baseline**	**Condition 1**	**Condition 2**	**Condition 3**	**Mean**
Voice dictation	-	-	-	.763	0.763
Touchscreen	-	0.923	0.883	.862	0.889
				Overall mean	0.826

Table 6 shows the Chronbach's Alpha coefficients for NASA-RTLX workload scores. Again, the overall mean of 0.848 indicates acceptable reliability.

Table 6 – Chronbach's Alpha coefficients for NASA-RTLX scores

	Experimental condition				
Study	**Baseline**	**Condition 1**	**Condition 2**	**Condition 3**	**Mean**
Voice dictation	0.837	0.853	0.809	-	0.833
Touchscreen	0.886	0.854	0.917	0.797	0.864
				Overall mean	0.848

13.2.2. Hedonic Rating

As a single-item scale, the reliability of the Hedonic Rating Scale cannot be verified using Chronbach's Alpha. However, Wanous & Reichers (1996) outline a method for determining the reliability of a single-item scale through correlation to another variable of known reliability, using the formula outlined by Nunnally (Nunnally and Bernstein, 1994):

$$\hat{r}_{xy} = \frac{r_{xy}}{\sqrt{r_{xx}}\sqrt{r_{yy}}}$$

Where r_{xy} is the correlation between variables x and y; r_{xx} is the reliability of variable x, r_{yy} is the reliability of variable y, and $\hat{r}_{xy}$ is the assumed true underlying correlation between x and y if both were measured perfectly, i.e. there is no measurement error.

Rearranging to solve for r_{yy} gives:

$$r_{yy} = \left(\frac{r_{xy}}{\sqrt{r_{xx}}\,\hat{r}_{xy}} \right)^2$$

Table 7 shows the correlations between the overall SUS scores and the Hedonic Rating scores for case studies 1 and 2. All correlations were significant at the α = 0.05 level. From these and the Chronbach's Alpha scores given in Table 5, the reliability of the Hedonic Rating scale can be estimated.

Table 7 – Spearman's rho correlations between overall SUS and Hedonic Rating Scale scores

	Experimental condition			
Study	**Condition 1**	**Condition 2**	**Condition 3**	**Mean**
Voice dictation	-	-	.397	0.397
Touchscreen	0.881	0.616	.838	0.778
			Overall mean	0.588

The correlation between the SUS and Hedonic Rating scores was lower for the single voice dictation study condition (0.397) than for the three touchscreen study conditions (mean = 0.778). This may be due to the nature of the technology: all of the participants in case study 2 were familiar with touchscreen interfaces while fewer were familiar with voice interfaces in case study 1, which may have affected their perception of the system.

Table 8 provides two estimates of the reliability of the scale, based on the two observed correlation mean values. A value of 0.90 is applied for the estimated true correlation, as suggested by Wanous & Reichers (1996). Taking the overall mean of these values provides a reliability estimate of between 0.516, improving to 0.841 if only the touchscreen study correlations are applied.

Table 8 - Estimated reliability of Hedonic Rating Scale

Observed correlation	**Estimated reliability**
Overall	0.516
Touchscreen study only	0.841
Estimated true correlation	**0.90**

Although lower than the 'minimum' 0.70 value prescribed in the literature, the lower estimate of 0.516 provides positive indications for the reliability of the scale. At the higher estimate, scale reliability is comparable and exceeds the minimum 'acceptable' value identified in the

literature. Additional data points would lead to an increased confidence in the estimate, which would be achieved by observing the correlations from later applications of the toolset.

13.2.3. Driving Metrics

The design intent of the evaluation toolset was to provide a consistent and repeatable platform on which technology evaluation studies could be conducted. A key aspect of this is the repeatability of the driving environment. In order to draw comparisons across studies, the nominal driving behaviour of each participant group, characterised by the driving metrics measured in the baseline (single-task) driving condition should not differ significantly. Mean values for three key driving metrics: SDLP, mean distance headway and standard deviation of distance headway ('headway variation') were calculated for case studies 1, 2 and 3a, and compared using one-way ANOVA.

Both SDLP ($F(2,67) = 3.08$, $p>0.05$) and mean distance headway ($F(2,67) = 1.88$, $p>0.05$) did not vary significantly across studies. Headway variation showed a greater degree of variation (±20%) but did not reach significance at $\alpha = 0.05$ ($F(2,67) = 3.08$, $p>0.05$). These findings show that the simulator system was capable of delivering consistent driving performance measures across multiple groups, providing evidence of its reliability.

13.3. Qualitative Learning from Evaluation Studies

Qualitative feedback obtained from users during the studies provided positive evidence of behavioural validity of the simulator setup. Users were able to engage in the driving task in a realistic way, with many requesting to fit their seatbelt when driving, and use their mirrors and turn signals when setting off. This however highlighted a limitation of the original setup: the lack of rear-view displays meant that the sense of immersion was 'broken' when users looked to their mirrors and saw a blank wall. This was exacerbated by the space in which the simulator was located: hard surfaces and light colours led to significant light scatter from the projectors, and isolation from exterior noise was insufficient.

While engaged in the driving task, some users commented on specific aspects of the control setup within the vehicle cabin. The use of gaming pedals compromised pedal position and feel. While this took a short time to acclimatise to, it was not perceived as an issue during the driving task. Users also found the steering feel to be lighter than expected, which may have affected their control of the vehicle.

A number of users commented on the perception of speed in the simulator, stating that it felt like they were travelling more slowly than the indicated speed. The fixed-base architecture leads to a lack of the kinaesthetic cues that a driver would normally experience, in terms of

whole-body accelerations and transients. Perception is further impaired by the lack of vibrotactile stimuli that would normally be associated with engine and road surface vibration.

Later studies were beset by recruitment issues, with multiple cancellations and no-shows leading to smaller sample sizes and/or time overrun. In part, this may be due to the lack of a defined incentive to compensate participants for their time. A more reliable source of participants would be highly beneficial to future studies.

13.4. Simulator Revisions

Based on the experiences highlighted above, a series of upgrades to the driving simulator were performed. These changes were made following collection of the data for the case studies, ensuring consistency of the evaluation environment throughout the project, but providing improved capabilities for future research. Key among these was the relocation of the simulator to a dedicated space within the University's Department of Psychology, as part of a new collaborative venture titled the Interdisciplinary Driver Behaviour Research Group (IDBRG). This involved the renovation of a room within the Psychology Extension building to specifications defined by the author based on learning from visits to other simulator facilities, and described in detail in submission 3, chapter 7.

The relocation allowed the issues identified above to be addressed. Interior walls were painted matt black to reduce light reflection, complemented with matt black suspended ceiling tiles, blackout curtains and dark floor tiles. The location of the room also significantly reduced external noise issues. As with the original space, air conditioning was fitted to ensure participant comfort. Figure 45 shows the revised simulator installed in its new location.

Figure 45 - Revised simulator installed in new location

The original display framework was re-installed with an upgraded visualisation package from Holovis. This included additional display panels to the rear of the vehicle utilising short-throw projectors, providing discrete views to each of the rear-view mirrors and significantly improving the level of immersion experienced by the driver, as shown in Figure 46.

The additional display channels required updated hardware, consisting of new i7-based PCs with NVIDIA GTX780 graphics cards. These were significantly quieter than the original visual servers while offering improved graphics performance. The visual configuration was updated in SCANeR Studio to encompass the new displays, as described in submission 3, section 7.2. In addition, a new supervisor PC was procured providing improved processing performance and an update of SCANeR Studio to version 1.4 was installed, providing an increase in visual rendering performance.

Figure 46 - Driver's-eye view of revised simulator showing mirror views

Issues with the placement and feel of the vehicle controls were also addressed. Steering feel was improved through mechanical adjustment and re-configuration, with subjective comparison to a real Jaguar XJ. The original brake and throttle pedals were modified to install sensors and connect them to the existing USB interface before replacing them into the car, providing the driver with accurate ergonomics. The pedal inputs were calibrated in software to set correct pre-travel and endpoint.

Finally, two changes were made to address the speed perception issue. Firstly, a vibration shaker was installed to the underside of the chassis below the driver's seat. This was driven from the LFE channel of the simulator audio, emulating engine and road vibration. Secondly, additional static objects (trees) were added to the virtual environment to give additional motion cues. This is shown in Figure 47 below, providing a comparison of the original and modified environment.

Original Modified

Figure 47 - Comparison of virtual environment before and after modification

13.5. Protocol Revisions

Learning from the case studies was also applied to the evaluation protocol. Recruitment issues experienced highlighted the need for a more reliable source of participants. Notably, all three of the driving research facilities visited during the international placement (UMTRI, NADS and MIT AgeLab) had an established participant database and recruitment process that was constantly maintained. This allowed them to quickly access groups of participants that met specific study criteria. The collaboration with the Department of Psychology opened up access to the SONA participant database, jointly maintained by Psychology and Warwick Business School. Future academic studies will be able to benefit from this access, however it is expected that participants will be incentivised to participate in studies using SONA, hence this cost must be included in the study plan.

The evaluation protocol uses a single-task baseline condition as a basis for establishing the effect of the technologies under test on driving performance. This requires the assumption that the driver is at the top of the 'learning curve' for the driving task before undertaking the dual-task drive; given the simplicity of the driving task this is not an onerous assumption. However, a check could be performed by including a single-task segment at the beginning and end of each dual-task drive, providing a longitudinal view of driving performance and balancing any learning effects.

Case studies 1 and 3 featured the application of eye glance measures. Eye glance duration and frequency is a core acceptance measure for the NHTSA Visual-Manual distraction guidelines and should therefore be included in technology evaluation studies wherever possible. The manual coding approach employed in case study 1 was effective and low-cost but labour-intensive. For future studies, coding could be outsourced to temporary employees (following the MIT AgeLab approach), although this would have a cost implication for the project. An alternative would be to invest in an eye-tracking system that could be integrated into the simulator; this would have the benefit of allowing eye glance data to be synchronised with the simulator data (subject to software licensing).

Another area of learning from MIT AgeLab was their standardisation of experimental protocol. This was achieved through the use of an experiment control/data acquisition system that scripted and automated many of the instructions to the participant, thus minimising potential bias from the experimenter. While this project goes some way towards creating a standardised experimental approach, further consideration could be given to scripting instructions to participants prior to trials to ensure that consistency is maintained.

13.6. Summary

The evaluation toolset was applied to three distinct technologies across three case studies. The outcomes demonstrated that the evaluation toolset could be applied effectively and generate valuable, reliable and consistent data. Quantitative data provided clear indications of technology performance and user acceptance, while qualitative data identified key UX factors that enriched the quantitative findings.

The three case studies were conducted with slightly different goals in mind. Case study 1 demonstrated the acceptability of the system based on industry guidelines. Case study 2 compared three competing technologies based on their UX. Case study 3a provided evidence of user acceptance of the technology, while studies 3b and 3c addressed more fundamental aspects of user interaction. The latter case both served as an indication of the boundaries of the capability of the toolset, and a demonstration of the need to adopt an iterative, user-centred technology development approach.

Learning from the case studies was applied to improve the evaluation toolset. Changes to the simulator installation, hardware, software and virtual environment address shortcomings identified by users and offer improve immersion. Recommended protocol changes, based on experience from the case studies and learning from the international placement, will assist participant recruitment, simplify data analysis and reduce experimenter bias.

14. Application of Toolset to Technology Development

The toolset is designed to enable evaluation of a range of technologies through established methods, providing a 'level playing field' across which data sets can be compared. In order for the evaluation toolset to add value to JLR, it needs to be capable of providing clear and relevant information that can be used to support decision making in the PD process. This chapter explains how the toolset fits into the development process and it can be used to support data-led decision making.

14.1. Applicability to Technology Readiness Levels (TRLs)

As defined in section 7.2, the toolset was designed to be applied to technologies that feature a minimum level of interactive functionality and are evaluated by users in-context. For case study 1, the technology under test was the interface itself, with the software concept developed to a point at which user interaction was possible, but the appearance and workflow were not production-ready. Conversely, in case studies 2 and 3a it was the display hardware that was under test, utilising a sample interface to enable evaluation.

Following an interactive UCD approach, users should be engaged from the earliest stages of development, allowing UX issues to be identified and corrected at relatively low cost. Towards the end of the development cycle and as the product nears production, user interface elements should be well developed and the emphasis should shift to the validation and tuning of the production-intent design. By this point, it is difficult and expensive to make significant changes to the usability, workload and UX of a system (see Figure 7, pg. 11). Hence, there is a 'window' in the development timeline within which the evaluation toolset will be most effective.

Figure 48 illustrates this with reference to the Technology Readiness Levels, using the definition provided by the Automotive Council UK (2011) and included in Appendix 1. At TRL1-2, the technology is not sufficiently mature to support interactive evaluation - at this early stage, analytical evaluation may be applied to a virtual concept. At TRL 3, experimental studies and simulations are feasible and the toolset is suitable for use. At TRL 7 the technology is considered ready to apply to specific vehicle programmes, with performance validated at TRL 8 and vehicle launch at TRL 9.

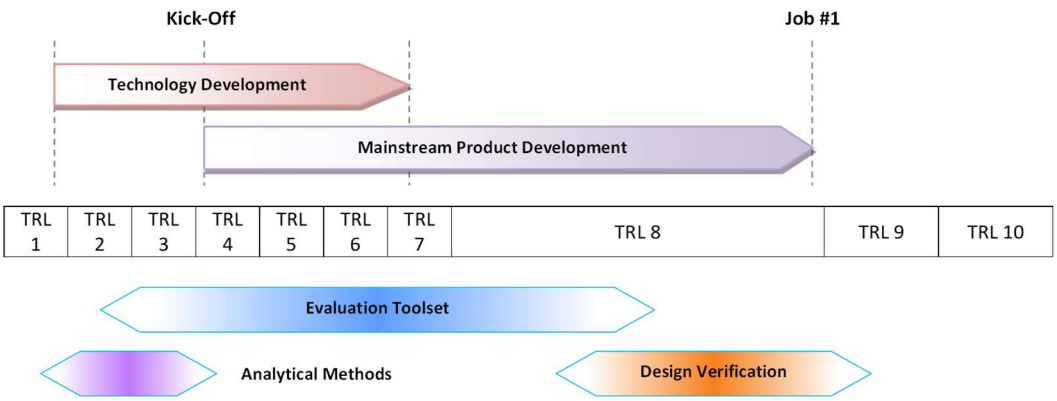


Figure 48 - Applicability of the Evaluation Toolset to TRLs

The apparent lengthening of TRL 8 in the diagram represents the transition from the technology development to product development process. The TRLs relate to the maturity of the specific technology under development; this will continue to mature as it is integrated into the target vehicle, but the timescale for doing so is now determined by the overall product maturity of the vehicle as a whole. This encompasses the work required to integrate the technology with other vehicle systems.

14.2. Determination of Pass/Fail Criteria

One of the key issues identified with evaluation approaches defined in the literature (see chapter 5) was the lack of specific pass/fail criteria against which technologies could be compared and verified. The Toolset addresses this issue by identifying pass/fail thresholds for each of the key performance metrics, described below. In addition, the toolset adopts the 'Traffic light' grading system applied by Jander, Borgval & Ramberg (2012):

- Green: Sufficient; does not lead to redesign recommendations
- Yellow: Further investigation is required; some re-design efforts should be performed
- Red: Severe; redesign efforts must be performed.

For each metric, a 'Green' grading indicates that the system meets the defined performance criteria in its current state of development. 'Yellow' criteria are less stringent and represent an inferior level of performance; this adjustment takes into account the early development stage and anticipates a 'glidepath' toward improved final performance, with incremental improvements as the technology matures. 'Red' criteria identify potentially significant issues with usability, workload or liking that raise questions over the suitability of the technology for automotive use. The criteria are summarised in Table 9 below, and explained in the following sections.

Table 9 – Pass/Fail criteria for Evaluation Toolset

Metric	Criteria	Grading bands		
		Green	Yellow	Red
Eye glance	% of glances > 2.0s duration	≤ 15	← →	≥ 18
Eye glance	Mean glance duration, s	≤ 2.0s	← →	> 2.4s
Eye glance	Total eyes off road time, s	≤ 12.0	← →	> 14.4s
Driving Performance	SDLP (absolute), m	≤ 0.30	← →	> 0.40
Driving Performance	SDLP (relative), % change from baseline	≤ +30	← →	> +50
Driving Performance	Mean Headway (relative), % change from baseline	≤ +30	← →	> +50
Workload	NASA-RTLX (absolute)	≤ 35	← →	> 45
Workload	NASA-RTLX (relative), % change from baseline	≤ +50	← →	> +150
Usability	Mean overall SUS score	≥ 70	← →	< 50
Hedonic rating	Median hedonic rating	≥ 6	← →	< 4

14.2.1. Eye Glance Metrics

Eye glance criteria are derived from the NHTSA guidelines (NHTSA, 2013). For a given secondary task:

a) No more than 15% of glance durations may be greater than 2.0 seconds
b) Mean single glance duration must not be greater than 2.0s
c) Total eyes-off-road time must not exceed 12.0 seconds

These criteria form the 'Green' thresholds for eye glance metrics. NHTSA's criteria are based on a series of studies that evaluated manual radio tuning, a common task with known accident risk. Criterion c) represents the 85th percentile of total glance time from simulator and road studies. At the 95th percentile, total glance time and mean glance duration are increased by approximately 20%, with the percentage of long glances (>2s) reaching 29%. Adopting this for the toolset criteria, the 'Red' grading is set at the above values +20%.

Note that these metrics are calculated per-task, whereas the other metrics are evaluated across the whole interaction. Hence, the number of tasks that sit in each pass/fail category must also be reported. This will help to differentiate between a system which is inherently distracting (i.e. all tasks fail to meet the criteria) and one in which a single task does not pass.

14.2.2. Workload metrics

Table 10 shows the subjective workload reported for case studies 1 and 2. Baseline and maximum figures are given, along with the percentage increase from the baseline.

Table 10 – Percentage increase from baseline for workload metrics

	Case study		
Metric	**1**	**2**	**3**
Baseline	20.4	16.0	-
Maximum RTLX	29.4	48.9	-
Change from Baseline %	+44.3%	+206.3%	-

The value for case study 1 was taken from the 'correction' condition, in which the voice dictation interface was evaluated. Eye glance measures from this study showed that the interface complied with NHTSA guidelines indicating acceptable workload level. Case study 2 included some tasks with long completion times that would be locked out under the NHTSA guidelines – this level of workload is likely to be unacceptable. In addition to the case study data, Reimer et al. (2013) found that a manual radio tuning task attracted a score of approximately 3.2 out of 10 on a self-reported workload scale, analogous to a TLX score of 32. Grier (2015) identified that the median TLX score for driving-related activities was 41.5, with the 75th percentile at 51.7. Based on these figures, the 'Green' threshold is set at 35, and the 'Red' threshold set at 45.

To account for differences in baseline workload perception, and acknowledging that all secondary tasks will see an increase in workload, a relative criterion is also applied: the 'Green' threshold is at ≤ +50% from baseline, while the 'Red' threshold is set to > +150%.

14.2.3. Usability metrics

The distribution of SUS scores is shown to be negatively-skewed, with a greater proportion of scores above 50 (the halfway point of the scale). Figure 49 (Bangor et al., 2009) shows the numeric scores compared to adjective ratings, grade scales and acceptability ranges, derived from multiple studies. The SUS scores applied for the toolset are based on these, with the 'Red' threshold set at 50 (the lower end of acceptability and just below 'OK') and the 'Green' threshold set at 70, approximating to a usability rating of 'Good'.

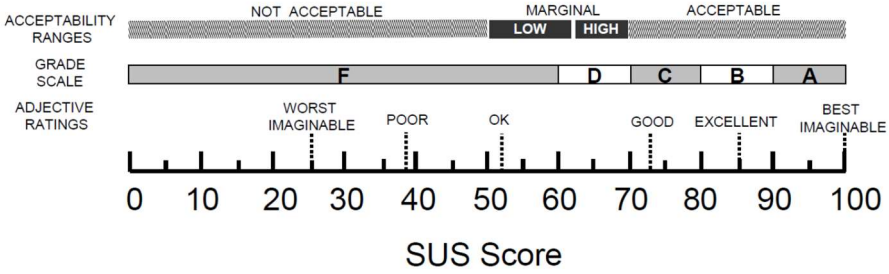


Figure 49 - Comparison of the adjective ratings, acceptability scores, and school grading scales, in relation to the average SUS score (Bangor et al., 2009)

14.2.4. Hedonic Rating

As a discrete scale, the hedonic rating produces ordinal data. It is therefore appropriate to consider the median (rather than the mean) as the measure of central tendency. Any technology introduced into a production vehicle should add to the user experience; i.e. have positive affect. The fail criterion is therefore set to a median score of less than 4 (based on a 0-8 scale) where 4 is the middle of the scale ('neither like nor dislike'), indicating a negative response. A score of ≥ 6 ('like moderately') signifies the 'Green' threshold.

14.2.5. Driving Metrics

As discussed in submission 2 section 9.3.1, lateral and longitudinal vehicle control are affected by visual and cognitive workload. In this way, variation in these parameters can be used to assess the level of workload imposed by a device. Table 11 shows the percentage increase from baseline for SDLP and mean distance headway, for each of the case studies. The largest effect on driving behaviour was observed in case study 2, which involved visual-manual tasks using a display mounted away from the driver's eye line.

Table 11 – Percentage increase from baseline for driving metrics

	Case study		
Metric	**1**	**2**	**3**
Baseline SDLP, m	0.20	0.21	0.26
Max SDLP, m	0.23	0.42	0.27
Change from Baseline, %	+9.5	+95.5	+4.5
Baseline mean Headway, m	94.6	67.0	71.1
Max mean Headway, m	112.8	118.0	88.5
Increase from baseline, %	+19.3	+76.1	+24.4

Based on the dimensions of the vehicle and roadway, and assuming that lateral deviations are normally distributed, the SDLP measured in case study 2 would see the car exceeding lane limits at least 5% of the time, i.e. for 50 meters from each kilometre. This is a clearly unacceptable level, therefore the 'Green' threshold is set at 0.30m and red at 0.40m. To account for variations in the baseline relative thresholds are set at +30% and +50%.

No absolute target for headway is set, due to potential differences in starting position relative to the lead car. Unlike lateral variation, there is an inverse relationship between accident risk and distance headway, as increasing the distance to the car ahead gives the driver more time to react. However, given that headway is indicative of cognitive distraction (Engstrom et al., 2005), relative thresholds are set at +30% and +50%.

14.3. Technology Scorecard

To support clear communication of findings, data should be presented in a way that minimises ambiguity while maximising information content. Cooper at al. (2002b) discuss the use of a 'project scorecard' to determine the merits of the project based on key criteria:

- Strategic alignment and importance
- Product advantage
- Market attractiveness
- Synergies with core competencies
- Technical feasibility
- Risk vs. return

Within this list, certain criteria, such as strategic alignment and likelihood of technical feasibility are considered 'Must meet'; that is, the project will be stopped if it fails to meet the required standard. For a given technology development project, strategic alignment and synergies are determined at a higher level by the business and technology strategy. Technical feasibility and product advantage can be illustrated through the technology evaluation process.

A scorecard is a visual data display used in a strategic performance measurement system to compare performance against targets and thresholds (Eckerson, 2010). It provides snapshots of performance at a given time through combinations of images and text (Kerzner, 2011). This allows managers to view all data relevant to project performance in one place, and to chart progress over time.

This concept can be adapted to the evaluation process to create a 'technology scorecard' that collates the data from a technology evaluation and puts it into a clear, concise and common format that can be easily interpreted by project stakeholders. Keyes (2011) identifies that

within a successful project management measurement system, *"Performance measurement systems must provide intelligence for decision makers, not just compile data"*, with performance measures relating directly to core strategic objectives. With the Technology Scorecard, performance is assessed against the pre-determined pass/fail criteria discussed above, providing a direct representation relative to global measures.

The completed Technology Scorecard is shown in Figure 50 below. This version is based on the data from case study 2, with the addition of dummy data for glance time, based on task completion data. The Scorecard is designed to be printed onto an A3 sheet; a reduced version is included in appendix 2.

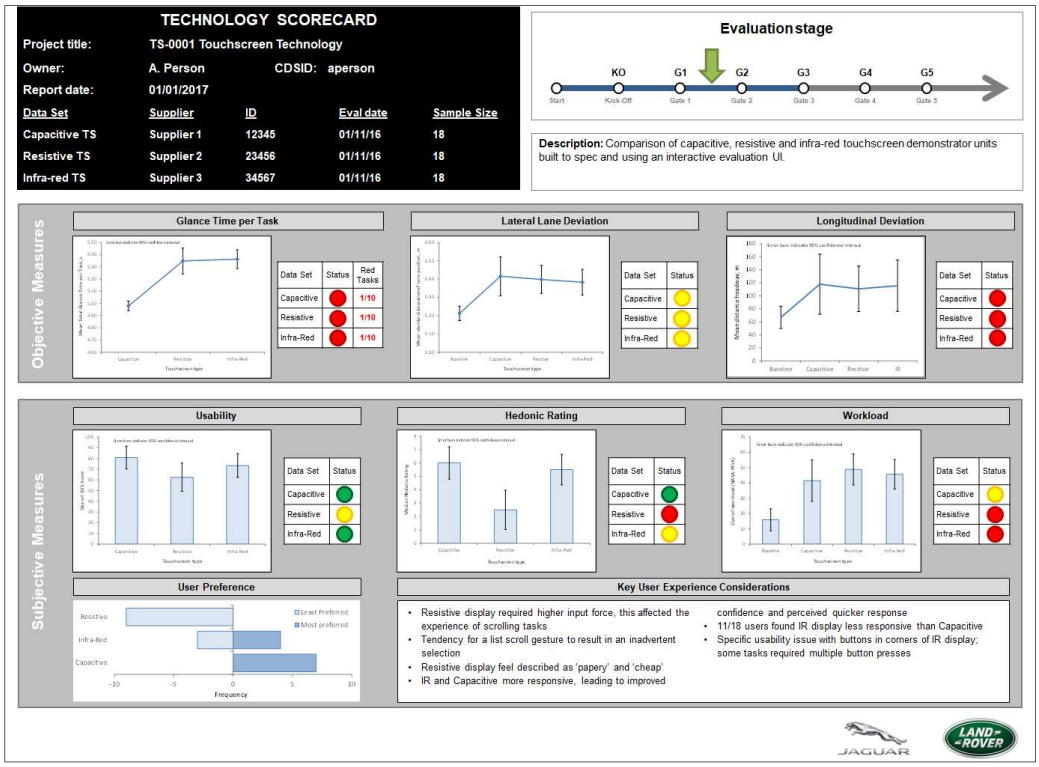


Figure 50 – Technology Scorecard

The Scorecard is designed using Gestalt principles of similarity and proximity, along with pre-attentive processing attributes of colour and enclosure to group the different elements of the data, allowing the user to quickly identify and read the relevant data (Eckerson, 2010). The tables to the right of each graphical data element contain the Green/Yellow/Red performance indicators, providing an instant indication of the performance of the technology for each metric.

The top section of the scorecard consists of three elements: the project information box, the evaluation stage indicator and the project description. The information box (Figure 51) contains relevant data about the project including name, owner and report date. In addition, it contains information about each of the data sets included in the report: the name and ID of the

technology under test, the evaluation date and the sample size for the data. This makes it clear to the user what information is displayed when it was acquired.

TECHNOLOGY SCORECARD				
Project title:	TS-0001 Touchscreen Technology			
Owner:	A. Person	CDSID:	aperson	
Report date:	01/01/2017			
Data Set	Supplier	ID	Eval date	Sample Size
Capacitive TS	Supplier 1	12345	01/11/16	18
Resistive TS	Supplier 2	23456	01/11/16	18
Infra-red TS	Supplier 3	34567	01/11/16	18

Figure 51 – Project information box

The evaluation stage indicator gives a visual indication of the point along the development path for the project at which the report was generated, giving a snapshot of the maturity of the technology. The description box contains text providing details of the project.

The middle section of the Scorecard presents the objective data from the study. Data is presented in graphical form, accompanied by a table containing the performance indicators. The glance time section is shown in Figure 52 below; the driving metrics, lateral and longitudinal deviation, are also presented in this section.

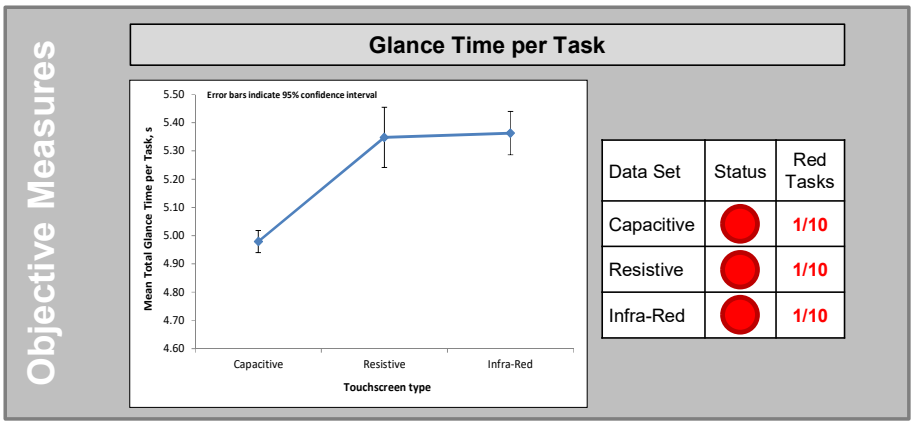


Figure 52 - Objective data element from Scorecard

The graphical data is presented at the device level to allow comparison between the technologies. However, 10 different tasks were performed in this study with differing levels of workload. To account for this, the status table on the right also includes a count of the 'Red' tasks, i.e. those that failed to meet the minimum threshold. A technology is graded 'Red' if any of the tasks fail.

The third section contains the subjective data, including graphical representations of usability, hedonic rating and workload with their respective status tables. In addition, this section features a ranking of user preference, along with a text box for key UX considerations. This

contains observations and/or verbatims that reflect issues identified by the users. This data provides context and explanation to the graphical data, so it is important that all appropriate comments are included.

14.4. Knowledge Capture

In order to achieve maximum benefit from the Evaluation Toolset approach, outputs from each study must be captured for future reference and comparison to other technologies. Cooper et al. (2002a) advocate the application of a 'capture and handling' system to manage ideas that pass the initial screen in a stage-gate process. Data from technology evaluations should be logged to a central 'Knowledge Store' that can be accessed by managers and engineers. The availability of such a database of results will help to provide justifications for development decisions by providing evidence of the relative performance of technologies, and by demonstrating improvement in performance through the PD process. Figure 53 below illustrates the information flow for a hypothetical knowledge capture system.

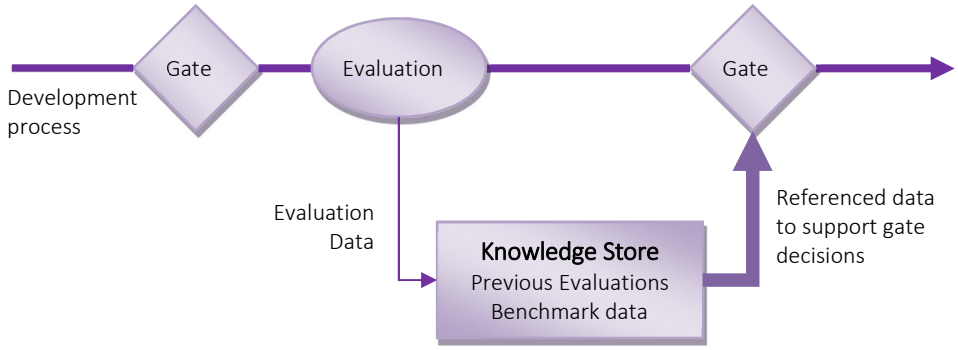


Figure 53 - Knowledge Capture of Evaluation Data

14.5. Recommendations for Toolset Implementation

The previous sections identified how the Evaluation Toolset might fit into the established development workflow. This section identifies specific provisions that are recommended for implementation of the Toolset methodology.

As discussed in section 13.5, the driving research facilities visited during the international placement all had an established participant database and recruitment process that was constantly maintained. Having experienced some recruitment issues in the earlier case studies, access to the University's SONA participant recruitment system was obtained. It is unlikely however that long-term access to this resource will be available, especially for non-academic projects; also industrial studies may have more specific requirements for participants. It is therefore recommended that a participant recruitment system is established, preferably with a database of participants from a non-engineering background. Harvey et al (2011c) note that

the availability of participants can influence designers' decision to conduct evaluations; hence a participant database will support evaluation activity.

As discussed in section 9.1.4, NHTSA guidelines recommend a minimum sample size of 24 for technology evaluations. Data from case study 2 showed large variance which was due in part to the reduced sample size. A sample size of 30 is therefore recommended which, along with improvements in participant instructions detailed below, will help to ensure that the measures applied have optimum sensitivity. The minimum sample size should be set to 24 in line with the NHTSA guidelines. Issues experienced with participant recruitment would be alleviated with the application of a participant recruitment system, as described above.

Providing participants with clearer instructions and additional practice time may help to reduce the variability of the data. While the instructions currently provided largely meet with ISO guidance (ISO, 2010b), there are some areas which could be improved. Firstly, participants should be instructed to regain their nominal follow distance to the lead vehicle after the completion of each task. This could assisted in the simulation scenario by moderating the speed of the lead vehicle until an appropriate follow position is established. Secondly, although no significant difference was seen in baseline driving performance across the studies, establishing acceptance criteria for the baseline drive would help to ensure a minimum level of performance and may help to identify unsuitable participants. Participants should be instructed to balance the performance of the primary and secondary tasks to avoid excessive focus on one or the other. Finally, participants should be given a separate dual-task practice scenario to prepare them for the task drives; this would of course increase the duration of the study which may have practical implications for the intended scope.

The evaluation protocol was subject to approval by BSREC, involving a detailed examination of the ethical implications of the study. The application process was lengthy but rigorous, ensuring that academic standards for ethics were maintained. In translating the toolset to an industrial environment, the risks to participants remain the same, but the academic framework is removed. Ideally, a group should be established within JLR to oversee research ethics and all studies should be approved before commencing, using academic ethical approval processes as a model. This will ensure that both participants and experimenters are aware of potential risks and minimise any disconnect between industrial and academic partners.

The toolset is designed to produce repeatable, reliable measurements across technologies and experimenters. Even so, it is important that experimenters receive appropriate training in the application of the methods, such that variability is minimised. Training should also cover research integrity and ethics to ensure that new experimenters are familiar with the established practice.

As the toolset incorporates objective and subjective metrics, it is necessary to consider the relative importance of these in different scenarios. Firstly, certain HMI features may be safety-critical (e.g. warnings, ADAS functions), whereby interaction occurs over a short time period. The objective eye glance and driving metrics reflect the safe operation of the vehicle, while the subjective usability and hedonic rating metrics give some measure of the User Experience. For safety-critical systems, the design goal may be oriented towards safe operation of the vehicle, with less consideration given to the creation of an engaging user experience. In these cases, priority may be given to the objective data in the overall evaluation. However, in situations where interaction is more sustained are there is greater opportunity for an 'experience' to be formed, the subjective measures become more relevant, reflecting the design goal of the system. Clearly, the performance measures still apply, effectively determining the limits of acceptable deviation from the baseline.

While it would normally be expected to see a positive correlation between hedonic rating and usability (as seen in section 13.2.2), it is possible that these measures may disagree. In such cases, the qualitative data may provide insights into the reasons for this. Ultimately, the decision on the value of a given technology rests with the judgement of the engineer; the evaluation toolset provides a framework for capturing and presenting the relevant information to support that judgement.

Conducting user-focused research clearly has a direct cost implication for the project and budget may not be initially available. This is offset however by reduced costs in development as issues are identified earlier. A shift of emphasis (and resource) to up-front activities requires high-level management support, both in terms of allocating budget and implementing findings. In presenting clear evidence of the performance of technology concepts, the toolset supports data-led decision making which will help to justify research expenditure.

Finally, it is important that the pass/fail criteria are reviewed periodically, to ensure alignment with the latest industry guidelines and to reflect customer expectations. Improvements in UX throughout the development process should also be monitored to confirm that the anticipated improvements (‘glidepath’) are realised, and criteria thresholds adjusted as necessary.

14.6. Toolset Strengths and Limitations

The Evaluation Toolset has been shown to be effective in the context of the case studies conducted under this project; however, as discussed in section 14.1, the toolset has limitations that must be acknowledged. Strengths and limitations are compared in Table 12 below.

With respect to reference tasks: the Alliance of Automobile Manufacturers (AAM) specifies a radio tuning task as reference for an 'acceptable' activity to be conducted while driving (Alliance of Automobile Manufacturers, 2006). It is unclear as to whether performance of this task would meet the pass/fail criteria established above.

Resource required for data analysis is identified as a limitation but could be a) delegated to a pool of trained part-time assessors (as per the approach adopted by MIT AgeLab); or b) automated, through the use of eye tracking systems and analysis scripts. Examples of such tools were developed in the course of the project including pro-forma Excel workbooks and MATLAB scripts; these help to reduce variability between analysts and minimise error but care must be taken to ensure that automation does not lose oversight of data quality. There is a cost implication for either solution that must be weighed against potential time savings.

Table 12 – Strengths and Limitations of Evaluation Toolset

Strengths	Limitations
Toolset provides rich information derived from users on their experience of the technology concept in context, highlighting UX issues that would not be identified by an analytical approach.	User-focused testing is more costly and time consuming than analytical approaches; however analytical approaches are not appropriate in some cases (e.g. evaluating display technology)
Standardised approach minimises set up time for studies and ensures a 'level playing field' on which technologies can be compared. Comparisons can be between technologies in a single study, between different technologies across studies, and between a single technology evaluated at different stages of development.	As identified in case study 3, there are situations in which more in-depth investigation of fundamental factors is required, in which case the standardised approach is not appropriate. In such cases, the experimental protocol must be tailored to the specific case.
Provides direct comparison to pass/fail criteria derived from industry guidelines and study data, allowing a quick indication of the potential performance of the technology from both a safety and UX perspective.	Toolset does not currently specify a default reference task that might serve as a baseline comparison for driving performance, eye glance behaviour and workload.

Utilises simple motorway driving scenario, reducing participant acclimatisation time. Low driving workload creates ideal conditions for users to allocate attention to the technology concept. Driving task designed to be sensitive to both lateral and longitudinal variations.	The scenario may not be suitable for specific features such as auto parking, reverse sensor/camera systems, pedestrian protection, or off-road technology. Flexibility of the simulator hardware allows new environments to be created.
Toolset generates data across a range of metrics to provide a complete picture of performance. Triangulation of data sources increases confidence in findings.	Toolset creates a considerable amount of data, requiring resource for analysis, with associated cost/time implications. Could be remedied by a) using trained assessors; b) automating process (e.g. using eye-tracking kit, analysis scripts)
Toolset employs simple, easy to understand subjective measures that can be completed by novice users and minimise time required compared to complex questionnaires.	Hedonic rating scale provides a limited measure of product emotion. Multi-dimensional approaches such as AttrakDiff (Hassenzahl, 2004a, Hassenzahl et al., 2015) or PED (Edmunds and Dorn, 2012) may provide a more comprehensive indication of the emotions experienced by the user; however these all require longer responses from participants and would impair the usability of the toolset itself.
Technology scorecard provides clear and concise representation of evaluation outcomes to support development decision making. Data is presented it in a standardised format with clear indications of performance relative to pass/fail criteria.	Subjective ratings applied post-task are global and cannot discriminate between specific tasks. Tasks can be differentiated using eye glance/task completion data.
Evaluation data is collected using a Knowledge Capture process and stored for future reference, allowing comparisons against benchmark data or longitudinal references.	

14.7. Summary

This chapter discussed how the Evaluation Toolset fits into the technology development workflow. The toolset is suitable for use on concepts with a Technology Readiness Level between TRL 3 and TRL 8, signifying the range between interactive concept and production-intent.

One of the key benefits of the Toolset is that it provides direct indications of the acceptability of a technology. Industry guidance and case study outcomes were used to define pass/fail criteria against which technology performance would be graded, using a 'traffic light' system. This allows engineers and managers to quickly determine the suitability of a concept for further development.

Outputs from evaluations are collated and presented in a Technology Scorecard: a one-page summary of key findings that includes both objective and subjective information. Results are displayed clearly in a logical and structured manner, supporting data-led decision making. The application of a knowledge capture system allows direct comparisons between technologies, studies and baseline data to provide clear indications of relative performance.

Finally, while the Toolset has many strengths there are also limitations, some of which can be resolved in application through the provisions recommended, which are based on learning from the implementation of the case studies and the international placement.

15. Impact and Added Value

The previous chapters set out the background, rationale and approach to the project, detailing the development of the Evaluation Toolset and application to the case studies. Findings from the studies were discussed with reference to application to the technology development process. This chapter outlines the impact created by the project, both in the industrial and academic sectors.

15.1. Industrial Impact

The industrial impact of the project is discussed in three sections: Direct impact, based on application of the Toolset to the case studies; Process impact, based on the derivation of the evaluation protocol and Technology Scorecard; and Capability impact, based on the development of the evaluation environment and opportunities for future application.

15.1.1. Direct Impact

The project had direct impact into three JLR technology projects (as featured in the case studies), providing detailed, relevant and reliable data on objective performance and user perception of novel technology concepts. This data was reported into JLR and provided data-led evidence to support the development decision-making process. Specifically:

- Case study 1 confirmed the potential of the prototype voice dictation interface, highlighting potential usability improvements. As an early-phase software concept changes carry little cost and risk, providing maximum benefit and reducing the number of changes further down the development path.
- Case study 2 demonstrated strong user preference for capacitive touchscreens over the incumbent resistive technology and confirmed the link from users' experiences of touchscreen mobile devices to their expectations for in-car touchscreens. While there appeared to be little objective difference between the capacitive and infra-red displays, qualitative data highlighted some specific usability issues with the latter; this finding supports the selection of the capacitive display for future vehicles.
- Case study 3 determined feasibility for the use of 3D displays in the instrument cluster (Figure 54), highlighting potential issues relating to image quality and visual discomfort with longer-term use that will require further investigation during product development. Left uninvestigated, this issue has the potential to cause major issues once the product is in the market; however early identification allows for robust testing and design activity to alleviate the problem.

Figure 54 - Evaluating the 3D display

"Working with WMG on the 3D Displays project provided valuable information to support the development process. Using the driving simulator allowed us to test the 3D displays with users and assess their response to a novel technology in a realistic context."

Elvir Hasedžić, JLR Research Lead Engineer

15.1.2. Process Impact

The Evaluation Toolset protocol is designed to provide a standardised approach to technology evaluation that effectively ensures a 'level playing field' on which technology concepts can be evaluated and compared. The process reduces variability between studies and experimenters; also minimises setup cost through standardisation of test procedure. As shown above, when applied to early-stage concepts, the approach yields valuable data to support further development.

The Technology Scorecard (Figure 55) collates evaluation data and presents it in a clear and easy-to-understand format to support the product development decision-making process. Technologies can be compared within a test set, against a benchmark or longitudinally to illustrate development over time. As a tool, this helps to illustrate the relative performance of candidate technologies and ensure that all relevant metrics are taken into consideration.

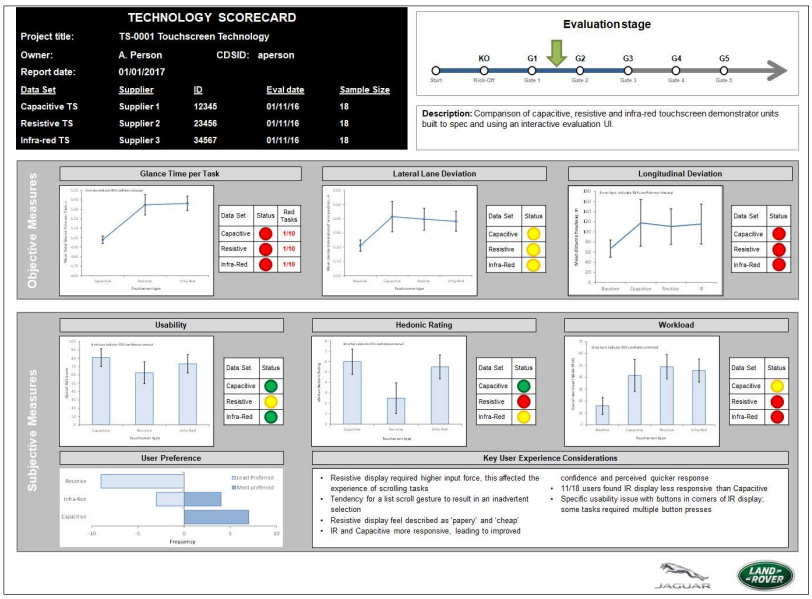


Figure 55 – Technology Scorecard

It is difficult to directly assess the financial value of the Evaluation Toolset until a product that has been through it reaches the market. However, while there is an inherent cost is undertaking customer-focused studies such as those discussed, there is also cost associated with releasing products that do not meet customer expectations. Figure 56 shows the 'Prevention, Appraisal, Failure' (PAF) cost of quality model, illustrating the 'Cost of Good Quality' and the 'Cost of Poor Quality' (Farr, 2011). A high-quality product requires high prevention and inspection costs to ensure that the product reaches market without defects. Conversely, a low quality product has low prevention and inspection costs, but incurs high costs due to failure. These costs may be internal, such as scrappage or rework; or external, such as warranty cost and damage to brand value.

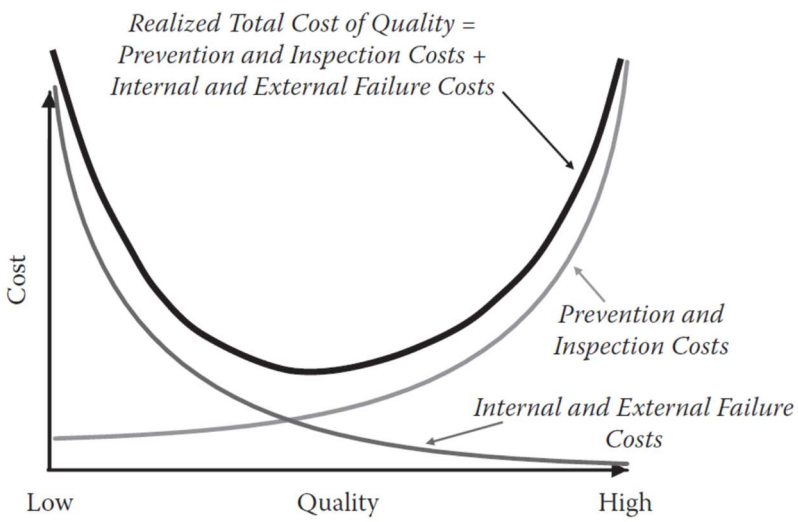


Figure 56 - Cost of Quality (Farr, 2011)

The quality literature indicates that costs incurred though the process will be outweighed by savings achieved through identifying issues up-front. The '1-10-100 rule' (Omachonu and Ross, 2004, Ross, 1999, Labovitz et al., 1992) states that cost escalates by a factor of ten with each step in product readiness: a problem that requires correction will cost ten times more than one prevented at the design stage, whereas a product failure in the market will cost ten times more to rectify than a corrective action. The implication is that there are considerable benefits to identifying and rectifying issues early in the design phase.

The UX of automotive technologies can have a significant impact on the brand value of an OEM if serious issues become apparent once the product is in the market. Submission 1 featured the example of MyFord Touch, which launched to market while suffering significant usability and reliability issues. This caused major damage to the Ford brand, with a drop from 5th to 27th in the JD Power IQS ranking in just two years, a position that has taken a further 5 years and a completely new system architecture (SYNC 3, launched in 2016) to rectify (Martinez, 2017). Currently, Ford still face a class action law suit over the issues with MyFord Touch[7]. The cost of rectification and the damage to the value of the brand represent a 'cost of poor quality' as defined by the PAF model above, which may have been avoided by early identification and correction of usability and UX issues. Figure 57 illustrates this with reference to the application of the Evaluation Toolset.

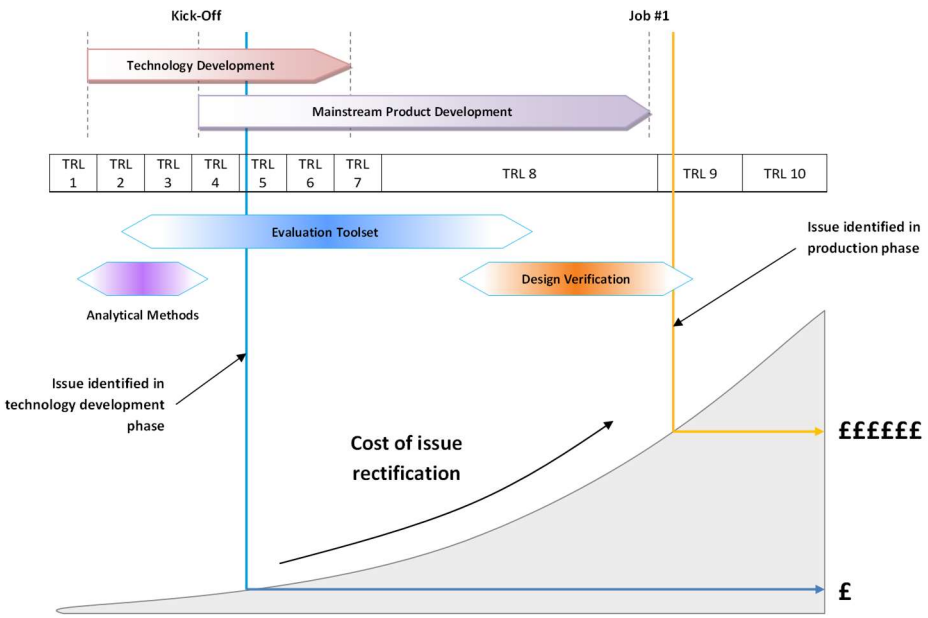


Figure 57 - Cost implications for identification of issues during development

[7] https://www.hbsslaw.com/cases/ford---myford-touch

15.1.3. Capability Impact

The WMG driving simulator developed under this project provides JLR with an additional research partner for conducting simulator-based research. The simulator was specifically designed with flexibility in mind, supporting the integration of both hardware and software concepts encompassing commonly-used HMI development tools, as demonstrated by the case studies. This makes it suitable for use in a wide range of in-vehicle technology evaluations.

Collaborative research ensures that academic standards of research ethics and data integrity are retained, while the proximity of the simulator facility and team to JLR's Research & Technology functions supports close co-ordination of research activity. In addition, the simulator team are able to support demonstrations and scoping studies for JLR and partners.

The collaboration with the Department of Psychology (see section 15.2.2) and relocation of the simulator has led to improvements in the simulator fidelity and experience. In addition, this relationship broadens research capability through access to additional expertise (including physiological measurement, visual attention and risk behaviour analysis), adding expert insight into driver behaviour that will benefit JLR in future research activity.

> *"The simulator that is proposed for relocation within Psychology is an important part of our portfolio of facilities… I welcome the collaboration with the Psychology Department"*
>
> Letter of support (extract) from Lee Skrypchuk,
> JLR HMI Technical Specialist

15.2. Academic Impact

The project also generated academic impact through publications, research collaborations and funding calls. These are detailed below:

15.2.1. Publications

Three peer-reviewed conference and journal papers were published during the course of the project, two as lead author and one as co-author.

- Comparing the User Experience of Touchscreen Technologies in an Automotive Application (Pitts et al., 2014); conference paper based on case study 2 and presented at Automotive UI 2014 in Seattle, WA, USA.
- Adding Depth: Establishing 3D Display Fundamentals for Automotive Applications (Pitts et al., 2015); SAE Technical Paper based on Case Study 3 and presented at SAE World Congress 2015 in Detroit, MI, USA.

- Cross-cultural differences in automotive HMI design: a comparative study between UK and Indian users' design preferences (Khan et al., 2016); journal paper as co-author. My contribution related to UX, Kano model, experimental methods, statistics and editing.

15.2.2. Collaboration with Department of Psychology

A key outcome from the project has been the formation of a new interdisciplinary collaborative relationship with the Department of Psychology, which saw the simulator re-installed in a dedicated space in the Psychology building. The collaboration opens access to wider expertise for both parties, with potential to broaden the range of research opportunities.

As an example: a detailed research proposal was created for a project looking at driver fatigue, in collaboration between WMG and the Sleep Lab at the Department of Psychology. This was enabled by the formation of the partnership and the installation of the simulator.

The simulator is currently in use on a JLR-supported PhD project involving Psychology and Computer Science; this would not have been possible without the facility in place. In addition, the simulator supports undergraduate projects, with the first to use the facility attracting the departmental prize for best 3rd year project.

> *"The WMG/Psychology driving simulator collaboration has enabled us to engage in road safety related Behavioral Science research in a way that was not previously possible. The facility has proved to be a great asset to the Department and has helped to generate interest from current and prospective students as well as from external partners and the general public. Working with WMG has also provided unique and extremely valuable access to technical expertise and automotive experience that complements our own research expertise and broadens greatly the reach of our research capabilities."*
>
> Derrick Watson, Deputy Head, Department of Psychology,
> University of Warwick

15.2.3. Funding

The simulator facility has enabled and supported a number of funding applications over the course of the project. Aside from the initial budget for hardware & software under LCVTP, the following funding opportunities have been realised:

- EPSRC Robotics and Autonomous Systems grant (£50k): visualisation & software upgrades
- University of Warwick Central Capital Expenditure grant (£85k): development of Psychology facility
- Warwick Behavioural Science GRP award (£2.6k): funding for pilot study into risky driving behaviour
- EPSRC Towards Autonomy - Smart and Connected Control PhD studentship (£220k): in collaboration with JLR, Psychology and Computer Science
- Innovate UK Better Interactions Between People and Machines feasibility study (£50k total award): Gesture Recognition for Automotive Applications (GRAAppl)

In addition to the above, additional research bids were compiled for a study into driver fatigue in collaboration with the Psychology Sleep Lab, and a mobile phone distraction project for a UK-based insurance company. Additional expressions of interest have been received from Highways England, Dorset & Gloucestershire Police forces and Optis, a developer of lighting simulation software.

15.3. Summary

The project has succeeded in delivering both industrial and academic impact. The case study findings directly supported three JLR technology projects, while the Evaluation Toolset provides a basis for evaluations going forward. The development of the driving simulator and collaboration with the Department of Psychology provides JLR with an academic research capability co-located with the Research & Technology team, allowing close co-ordination of research activity.

The Psychology collaboration also creates new opportunities for interdisciplinary research. Creating a bespoke facility for the simulator provided the opportunity to improve its capability and integrate new functionality, leading to improved research outcomes for future studies. The simulator facility has been instrumental in supporting a number of successful funding bids, with additional interest suggesting the potential for future funded studies.

16. Conclusions

The development of new automotive HMI technologies offers manufacturers the opportunity to gain commercial advantage in a competitive and crowded marketplace. User response to technology influences the perception of the brand as a whole, so it is important that in-vehicle systems provide a high-quality user experience. However, the demands of usability and user experience must be balanced against the requirement for driver safety. It is critical therefore that manufacturers are able to obtain clear and relevant information on user experience and safety when developing new technology products.

The aim of the project was to develop an Evaluation Toolset that would support the development of automotive HMI concepts through a User Experience-based approach. Following a thorough review of the literature, industry guidelines and research practice, an evaluation protocol was devised to provide a standardised evaluation approach for early-stage concepts. The protocol encompasses key objective measures of driving and task performance, along with subjective ratings of workload, usability and hedonic rating. These quantitative measures are supported by qualitative data, providing detail and context to the evaluation outcomes.

The second element of the Toolset is the evaluation environment: the driving simulator. This was developed to support the integration of hardware and software concepts, allowing early-phase prototypes to be evaluated in a realistic driving context. A basic driving environment was designed to minimise participant training time and allow users to engage with the technology concepts under test. Further developments of the simulator offer improved performance and experience, providing an excellent platform for future research projects.

The Evaluation Toolset was applied to three case study technologies, representing both hardware and software concepts at differing technology readiness levels. The outcomes from the case studies confirmed its effectiveness, producing valuable data on usability and User Experience. However, the specific requirements of case study 3 highlighted the limitations of the approach. Recommendations for revisions to the toolset were made based on the case study findings.

In order for the Toolset to fully support development decisions, clear and concise data must be presented. A Technology Scorecard was developed which collates all of the key metrics defined under the evaluation protocol with reference to pre-defined pass/fail criteria, allowing at-a-glance indication of technology performance. A proposed knowledge capture system would collate study findings and create a database of results for future reference.

16.1. Achievement of Objectives

Five objectives were defined under the project methodology. These are discussed below:

Determine the key objective performance indicators for driver behaviour – a review of the literature and current HMI guidelines identified a number of objective metrics to characterise driver behaviour. In assessing primary vehicle control, lateral and longitudinal position variation are key measures. Lateral variation is measured by Standard Deviation of Lane Positon (SDLP), which determines deviation from an arbitrary mean lane position. Longitudinal variation is measured relative to a lead vehicle as Headway, the spatial or temporal distance between the two vehicles. Visual distraction tasks will generally see an increase in lateral variation as the driver diverts their attention from lane tracking. Cognitive distraction will generally see an increase in headway as the driver subconsciously creates more space as a coping strategy.

In addition to the vehicle control metrics, eye glance behaviour is an important indicator of distraction. Diverting visual attention from the forward roadway is known to be correlated with accident risk, with individual glances of over 2 seconds linked to a significant increase in risk. Visual distraction will see an increase in glances away from the forward roadway; however cognitive distractions will see an increased focus on the forward roadway, albeit with a decrease in glances to mirrors and a reduction in peripheral awareness. This 'tunnel vision' effect decreases the driver's ability to identify potential hazards in the environment.

Develop a context-rich environment to enable evaluation of novel HMI technologies – the usability and UX literature both highlighted the importance of context of use on the user's overall perception of the product. The context of use must therefore be incorporated into the evaluation process to ensure that data collected accurately reflects how the product is likely to be received by the user. In addition, the evaluation environment must support the range of metrics required to produce a complete picture of the interaction experience. Given the specific nature of the in-car environment, the decision was taken to develop a driving simulator to serve as the evaluation environment. The simulated driving task emulates the workload characteristics of real-world driving, while the vehicle cabin maintains the ergonomics and ambiance of the real vehicle.

The case studies served to demonstrate that the simulator could successfully be used to test a range of in-vehicle technologies, including both hardware and software concepts. The driving metrics captured from the simulator data during the studies identified the influence of secondary task completion on driver behavior, serving as an indicator of driver distraction.

While indicators of behavioural validity observed during the case studies were encouraging, a number of issues were highlighted that reduced the fidelity of the simulator. A programme of improvements was undertaken which included moving the simulator to a new, bespoke space, reinstating the original pedal set and adding rear displays. These changes enhance the capability of the simulator for future research.

Identify and evaluate methods for capturing the User Experience of in-car technology – Review of the literature identified key models of UX that describe how the user's perception of a product is formed. Utility and usability is at the core of UX, in that it is fundamental that a product can effectively complete its design function in order for higher level experiences to be created. Hence, usability measures form a core aspect of assessing UX. Models also indicate that perceived quality has a hedonic aspect, a key into the emotional response of the user. The toolset therefore employs the hedonic rating scale, adopted from the sensory sciences, as a measure of overall liking of the product under test. Finally, key qualitative observations relating to product preference are captured using a semi-structured interview technique, yielding valuable insights into product performance and user behavior that can be used to shape future generations of product.

Create and test a protocol for the evaluation of technologies from the perspectives of UX and driver safety – The wide range of product evaluation methods identified in the literature can be applied at different stages of the development programme to technologies of different technology readiness level. The toolset described here seeks to reduce the variability in the application of these methods by a) restricting candidate technologies to those that have some interactive functionality, and b) by specifying a standardised evaluation approach that can be consistently applied across technologies and by different experimenters, ensuring a 'level playing field' when making judgements about the relative merit of different concepts. The toolset incorporates both objective and subjective quantitative measures, alongside qualitative data relating to the quality of the user experience. In this way, a full picture of the impact of the technology can be obtained, encompassing UX and driver safety.

Determine how the evaluation protocol can be integrated into the up-front NPD process for in-vehicle technology – The integration of customer data into the NPD process is identified as a critical success factor for new products. Applying user-centred evaluations early in the NPD process can help to identify and rectify issues at a point where it is relatively inexpensive and simple to make changes to the product, thus reducing the risk inherent in new technology projects. The key aim of the Evaluation Toolset is to support the strategic decision-making process by providing clear and accurate data on product performance. To achieve this, a Technology Scorecard is created for each project, outlining the relative performance of each

of the technologies under test and key drivers of UX. This can form part of a larger project scorecard, incorporating factors such as financials and strategic fit.

16.2. Limitations of the Research and Further Work

The main limitation of the research is that the application of the Evaluation Toolset has not been monitored in full from the beginning to the end of a technology development programme. While improvements to the protocol were made throughout the project, there was no structured review process undertaken with stakeholders. Undertaking a 'Build-Test-Feedback-Revise' approach analogous to the 'spiral' development phases discussed in section 0 may have helped to ensure that the toolset was fully aligned with the requirements of the OEM, ready for application to live development projects. The design of the Technology Scorecard would also benefit from this iterative approach, with improvements made based on stakeholder feedback. Using the methodology in a full programme would provide direct evidence of its effectiveness and provide data to evaluate the choice of metrics and validate the pass/fail criteria. Application to a full programme would also allow the long-term monitoring of the technology as it moves through the development process, providing an indication of value added.

The application of software tools would help to improve the Toolset. Electronic questionnaires could be developed for use in the evaluations to reduce data entry workload and minimise paper use. Database and reporting software could also automate the creation of the Technology Scorecard, allowing rapid comparisons across evaluation studies.

There are a number of ways in which the driving simulator could be developed to increase its functionality. As discussed, the capture of eye glance behaviour provides an important measure of driver distraction. The integration of an eye tracking system into the simulator would allow automated capture of eye glance data, minimising analysis time. In addition, a Detection-Response Task (DRT) system would provide an additional measure of driver attention while engaged in the driving and technology tasks. The estimated cost for this capability is approximately £4,000 for the DRT system and up to £40,000 for a multi-camera research-grade eye tracker.

The simple driving environment developed allows users to easily learn and perform the driving task. While effective, it is missing some of the cues that drivers may find familiar while driving; adding additional objects such as signage and bridges may help to improve the realism of the environment. It is also noted that the motorway-based, steady-state nature of the driving task is not suited to certain types of technology, such as reverse parking aids or pedestrian avoidance systems. The simulation software provides the flexibility to create new

environments tailored to specific requirements; an appropriate environment for this type of test (e.g. an urban environment) could be developed if required.

Improvements could also be made to the audio and vibration systems in the simulator. Separation of internal (engine) and external (wind, noise, tyres, traffic) noise sources would help to improve the level of immersion. Modelling of vibration signals based on road conditions would also provide a more realistic stimulus to the driver.

References

ABI RESEARCH. 2011. *97% of All Smartphones Will Have Touchscreens by 2016* [Online]. Available: https://www.abiresearch.com/press/97-of-all-smartphones-will-have-touchscreens-by-20 [Accessed 4th May 2014].

ADAMS, A. & COX, A. L. 2008. Questionnaires, in-depth interviews and focus groups. *Research methods for human-computer interaction.* Cambridge University Press, pp. 17-34.

ALLIANCE OF AUTOMOBILE MANUFACTURERS 2006. Statement of Principles, Criteria and Verification Procedures on Driver Interactions with Advanced In-Vehicle Information and Communication Systems.

ASSOCIATION DES CONSTRUCTEURS EUROPÉENS D'AUTOMOBILES 2018. The Automobile Industry Pocket Guide 2018/2019.

AUTOMOTIVE COUNCIL UK 2011. Automotive Technology and Manufacturing Readiness Levels: A guide to recognised stages of development within the Automotive Industry.

BACH, K. M., JÆGER, M. G., SKOV, M. B. & THOMASSEN, N. G. 2009. Interacting with in-vehicle systems: understanding, measuring, and evaluating attention. Proceedings of the 23rd British HCI Group Annual Conference on People and Computers: Celebrating People and Technology, Cambridge, United Kingdom, 2009, 1671070: British Computer Society, 453-462.

BANGOR, A., KORTUM, P. & MILLER, J. 2009. Determining What Individual SUS Scores Mean: Adding an Adjective Rating Scale. *Journal of Usability Studies,* 4 (3), 114-123.

BANGOR, A., KORTUM, P. T. & MILLER, J. T. 2008. An Empirical Evaluation of the System Usability Scale. *International Journal of Human-Computer Interaction,* 24 (6), 574-594.

BENGLER, K. 2007. Subject Testing for Evaluation of Driver Information Systems and Driver Assistance Systems — Learning Effects and Methodological Solutions. *In:* CACCIABUE, P. C. (ed.) *Modelling Driver Behaviour in Automotive Environments: Critical Issues in Driver Interactions with Intelligent Transport Systems.* London: Springer, pp. 123-134.

BEVAN, N. 1995. Usability is Quality of Use. 6th International Conference on Human Computer Interaction, Yokohama, Japan, July, 1995: Elsevier.

BHISE, V., DOWD, J. & SMID, E. 2003. A Comprehensive HMI Evaluation Process for Automotive Cockpit Design. *Human Factors in Driving, Seating, & Vision (SP-1772).* SAE International. Technical Paper 2003-01-0126.

BHUIYAN, N. 2011. A framework for successful new product development. *Journal of Industrial Engineering and Management,* 4 (4), 746-770.

BLANDFORD, A. & GREEN, T. 2008. Methodological development. *In:* CAIRNS, P. & COX, A. L. (eds.) *Research Methods for Human-Computer Interaction.* Cambridge University Press, pp. 158-174.

BMW. 2019. *Natural and fully multimodal interaction with the vehicle and its surroundings: BMW Group presents BMW Natural Interaction for the first time at Mobile World Congress 2019* [Online]. BMW. Available: https://www.press.bmwgroup.com/global/article/attachment/T0292196EN/427624 [Accessed 10th June 2019].

BOUCHARD, S., ROBILLARD, G. & RENAUD, P. 2007. Revising the factor structure of the Simulator Sickness Questionnaire. *Annual Review of CyberTherapy and Telemedicine,* 5, 117-122.

BROOKE, J. 1996. SUS: A 'quick and dirty' usability scale. *In:* JORDAN, P. W., THOMAS, B. A., WEERDMEESTER, B. & MCCLELLAND, I. L. (eds.) *Usability evaluation in industry.* London, England: Taylor & Francis, pp. 189–194.

BROOKE, J. 2013. SUS: A Retrospective. *Journal of Usability Studies,* 8 (2), 29-40.

BROOKS, J. O., GOODENOUGH, R. R., CRISLER, M. C., KLEIN, N. D., ALLEY, R. L., KOON, B. L., LOGAN JR, W. C., OGLE, J. H., TYRRELL, R. A. & WILLS, R. F. 2010. Simulator sickness during driving simulation studies. *Accident Analysis & Prevention,* 42 (3), 788-796.

BROSTRÖM, R., BENGTSSON, P. & AXELSSON, J. 2011. Correlation between safety assessments in the driver–car interaction design process. *Applied Ergonomics,* 42 (4), 575-582.

BURNETT, G. 2008. Designing and evaluating in-car user-interfaces. *In:* LUMSDEN, J. (ed.) *Handbook of Research on User-Interface Design and Evaluation for Mobile Technology.* IGI Global, pp. 218-236.

BYERS, J. C., BITTNER, A. C. & HILL, S. G. 1989. Traditional and raw task load index (TLX) correlations: Are paired comparisons necessary? *Advances in Indusrial Erogonomics and Safety*, 481-485.

CACCIABUE, P. C. & CARSTEN, O. 2010. A simple model of driver behaviour to sustain design and safety assessment of automated systems in automotive environments. *Applied Ergonomics,* 41, 187-197.

CANALYS. 2018. *Android Auto and Apple CarPlay in half of vehicles sold in US and Europe in Q1* [Online]. Available: https://www.canalys.com/newsroom/android-auto-and-apple-carplay-in-half-of-vehicles-sold-in-us-and-europe-in-q1?campaignname=IV2018 [Accessed 10th June 2019].

COBB, S. V. G., NICHOLS, S., RAMSEY, A. & WILSON, J. R. 1999. Virtual Reality-Induced Symptoms and Effects (VRISE). *Presence: Teleoperators and Virtual Environments,* 8 (2), 169-186.

COLLET, C., GUILLOT, A. & PETIT, C. 2010. Phoning while driving II: a review of driving conditions influence. *Ergonomics,* 53 (5), 602-616.

COOPER, R. G. 1988. Predevelopment activities determine new product success. *Industrial Marketing Management,* 17 (3), 237-247.

COOPER, R. G. 1996. Overhauling the new product process. *Industrial Marketing Management,* 25 (6), 465-482.

COOPER, R. G. 1998. Benchmarking new product performance: Results of the best practices study. *European Management Journal,* 16 (1), 1-17.

COOPER, R. G. 2011. *Winning at New Products : Creating Value Through Innovation,* New York, NY, Basic Books.

COOPER, R. G. 2014. What's Next? After Stage-Gate. *Research-Technology Management,* 57 (1), 20-31.

COOPER, R. G. 2016. Agile-Stage-Gate Hybrids The Next Stage for Product Development. *Research-Technology Management,* 59 (1), 21-29.

COOPER, R. G. 2017. We've Come a Long Way Baby. *Journal of Product Innovation Management,* 34 (3), 387-391.

COOPER, R. G., EDGETT, S. J. & KLEINSCHMIDT, E. J. 2002a. Optimizing the stage-gate process: What best-practice companies do - I. *Research-Technology Management,* 45 (5), 21-27.

COOPER, R. G., EDGETT, S. J. & KLEINSCHMIDT, E. J. 2002b. Optimizing the stage-gate process: What best-practice companies do - II. *Research-Technology Management,* 45 (6), 43-49.

CORTINA, J. M. 1993. What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology,* 78 (1), 98-104.

CQM 1993. A special issue on Kano's Methods for Understanding Customer-defined Quality. *Centre for Quality Management Journal,* 2 (4), 3-36.

CRESWELL, J. W. 2006. *Research design: qualitative, quantitative, and mixed methods approaches*, SAGE.

CUNNINGHAM, M. & REGAN, M. Autonomous Vehicles: Human Factors Challenges and Opportunities. Australasian Road Safety Conference, Gold Coast, Australia, 14 - 16 October, 2015: Australian College of Road Safety.

DARVEIGHT, C. 2010. *Better Cars, Smaller Profits. The Automotive Industry at the Crossroad* [Online]. autoevolution.com. Available: http://www.autoevolution.com/news/better-cars-smaller-profits-the-automotive-industry-at-the-crossroad-26697.html [Accessed 23rd October 2013].

DAVIS, F. D. 1989. Perceived Usefulness, Perceived Ease Of Use, And User Acceptance. *MIS Quarterly,* 13 (3), 319-340.

DAVIS, F. D., BAGOZZI, R. P. & WARSHAW, P. R. 1989. User Acceptance Of Computer Technology: A Comparison Of Two Theoretical Models. *Management Science,* 35 (8), 982-982.

DE WAARD, D. 1996. *The Measurement of Drivers' Mental Workload.* PhD Thesis, University of Groningen.

DEBERNARD, S., CHAUVIN, C., POKAM, R. & LANGLOIS, S. 2016. Designing Human-Machine Interface for Autonomous Vehicles. *IFAC-PapersOnLine,* 49 (19), 609-614.

DESMET, P. M. A. & HEKKERT, P. 2007. Framework of product experience. *International Journal of Design,* 1 (1), 57-66.

DINGUS, T. A., HANOWSKI, R. J. & KLAUER, S. G. 2011. Estimating Crash Risk. *Ergonomics in Design,* 19 (4), 8-12.

DINGUS, T. A., KLAUER, S.G., NEALE, V. L., PETERSEN, A., LEE, S. E.,, SUDWEEKS, J., PEREZ, M. A., HANKEY, J., RAMSEY, D., GUPTA, S., & BUCHER, C., DOERZAPH, Z. R., JERMELAND, J., AND KNIPLING, R.R. 2006. The 100-Car Naturalistic Driving Study: Phase II – Results of the 100-Car Field Experiment. National Highway Traffic Safety Administration.

DÖRING, T., KERN, D., MARSHALL, P., PFEIFFER, M., SCHÖNING, J., GRUHN, V. & SCHMIDT, A. 2011. Gestural interaction on the steering wheel: reducing the visual demand. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Vancouver, BC, Canada, 2011, 1979010: ACM, 483-492.

ECKERSON, W. W. 2010. *Performance Dashboards : Measuring, Monitoring, and Managing Your Business (2nd edn.)*, John Wiley & Sons.

EDMUNDS, R. & DORN, L. 2012. Developing a Multidimensional Assessment of In-Vehicle Technology. *In:* DORN, L. (ed.) *Driver Behaviour and Training Volume V.* Farnham: Ashgate, pp. 531-547.

ENGSTROM, J., ABERG, N., JOHANSSON, E. & HAMMARBACK, J. Comparison between Visual and Tactile Signal Detection Tasks Applied to the Safety Assessment of In-Vehicle Information Systems. Proceedings of the 3rd International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design, Rockport, Maine, June 27-30, 2005.

EUROPEAN COMMISSION 1999. Statement of Principles on Human Machine Interface (HMI) for In-Vehicle Information and Communication Systems European Commision.

EUROPEAN COMMISSION 2007. Commission Recommendation on Safe and Efficient In-Vehicle Information and Communication Systems: Update of the European Statement of Principles on Human Machine Interface. European Commision.

FARR, J. 2011. *Systems Life Cycle Costing,* Boca Raton, CRC Press.

FRAWLEY, P. 2012. *The Ten Most Advanced Cars Of Their Decades* [Online]. Jalopnik.com. Available: http://jalopnik.com/5888941/the-ten-most-advanced-cars-of-their-decades/ [Accessed 20th May 2013].

FREDHEIM, H. 2011. *Why User Experience Cannot Be Designed* [Online]. Smashing Magazine. Available: http://www.smashingmagazine.com/2011/03/15/why-user-experience-cannot-be-designed/] [Accessed March 10th 2015].

GASPAR, J., FONTUL, M., HENRIQUES, E. & SILVA, A. 2014. User satisfaction modeling framework for automotive audio interfaces. *International Journal of Industrial Ergonomics,* 44 (5), 662-674.

GODLEY, S. T., TRIGGS, T.J., FILDES, B.N. 2001. Driving simulator validation for speed research. 34, 589-600.

GRIER, R. A. 2015. How High is High? A Meta-Analysis of NASA-TLX Global Workload Scores. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting,* 59 (1), 1727-1731.

GÜTTNER, A. & SOMMER-DITTRICH, T. 2008. Current Issues at OEMs and Suppliers. *In:* PARRY, G. & GRAVES, A. (eds.) *Build To Order.* Springer London, pp. 55-73.

HART, S. G. 2008. NASA Task Load Index (NASA-TLX); 20 years later. NASA-Ames Research Centre.

HART, S. G. & STAVELAND, L. E. 1988. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. *In:* HANCOCK, P. A. & MESHKATI, N. (eds.) *Human Mental Workload.* pp. 239-250.

HARVEY, C. & STANTON, N. A. 2012. Trade-off between context and objectivity in an analytic approach to the evaluation of in-vehicle interfaces. *IET Intelligent Transport Systems,* 6 (3), 243-258.

HARVEY, C., STANTON, N. A., PICKERING, C. A., MCDONALD, M. & ZHENG, P. 2011a. In-Vehicle Information Systems to Meet the Needs of Drivers. *International Journal of Human–Computer Interaction,* 27 (6), 505-522.

HARVEY, C., STANTON, N. A., PICKERING, C. A., MCDONALD, M. & ZHENG, P. 2011b. To twist or poke? A method for identifying usability issues with the rotary controller and touch screen for control of in-vehicle information systems. *Ergonomics,* 54 (7), 609-625.

HARVEY, C., STANTON, N. A., PICKERING, C. A., MCDONALD, M. & ZHENG, P. 2011c. A usability evaluation toolkit for In-Vehicle Information Systems (IVISs). *Applied Ergonomics,* 42 (4), 563-574.

HASSENZAHL, M. 2001. The Effect of Perceived Hedonic Quality on Product Appealingness. *International Journal of Human-Computer Interaction,* 13 (4), 481 - 499.

HASSENZAHL, M. 2004a. The Interplay of Beauty, Goodness, and Usability in Interactive Products. *Human–Computer Interaction,* 19 (4), 319-349.

HASSENZAHL, M. 2004b. The thing and I: understanding the relationship between user and product. *In:* MARK A. BLYTHE, A. F. M., KEES OVERBEEKE AND PETER C. WRIGHT (ed.) *Funology: From Usability to Enjoyment.* Dordrecht, Netherlands: Kluwer Academic Publishers, pp. 31-42.

HASSENZAHL, M., WIKLUND-ENGBLOM, A., BENGS, A., HÄGGLUND, S. & DIEFENBACH, S. 2015. Experience-Oriented and Product-Oriented Evaluation: Psychological Need Fulfillment, Positive Affect, and Product Perception. *International Journal of Human–Computer Interaction,* 31 (8), 530-544.

HAUSEN, D. 2012. Peripheral interaction: facilitating interaction with secondary tasks. Proceedings of the Sixth International Conference on Tangible, Embedded and Embodied Interaction, Kingston, Ontario, Canada, 2012, 2148227: ACM, 387-388.

HEAPS, R. 2010. *7 amazing advances in new car technology* [Online]. Bankrate.com. Available: http://www.bankrate.com/finance/auto/7-amazing-advances-in-new-car-technology-1.aspx [Accessed 17th May 2013].

HIBBERT, L. 2013. Driven To Success. *Professional Engineering.* Caspian Media.

HORNBÆK, K. 2006. Current practice in measuring usability: Challenges to usability studies and research. *International journal of human-computer studies,* 64 (2), 79-102.

HOYES, T. W., STANTON, N.A., TAYLOR, R.G. 1997. Risk Taking In Simulated Environments: Evidence Relevant To Risk Homeostasis Theory. *In:* ROTHENGATTER, J. & CARBONELL-VAYA, E. (eds.) *Traffic and Transport Psychology: Theory and Application.* Amsterdam, The Netherlands: Elsevier, pp. 203-208.

IDC. 2014. *Worldwide Smartphone Shipments Top One Billion Units for the First Time, According to IDC* [Online]. Available: http://www.idc.com/getdoc.jsp?containerId=prUS24645514 [Accessed 4th May 2014].

IGBARIA, M., SCHIFFMAN, S. J. & WIECKOWSKI, T. J. 1994. The respective roles of perceived usefulness and perceived fun in the acceptance of microcomputer technology. *Behaviour & Information Technology,* 13 (6), 349-361.

IHS MARKIT. 2016. *Consumers Crave Advanced Technology in New Vehicle Purchases, IHS Markit Survey Finds* [Online]. Available: https://news.ihsmarkit.com/press-release/automotive/consumers-crave-advanced-technology-new-vehicle-purchases-ihs-markit-survey [Accessed 5th June 2019].

IHS MARKIT. 2017. *Automotive Touch Screen Shipments to Top 50 Million Units in 2017, IHS Markit Says* [Online]. IHS Markit. Available: https://technology.ihs.com/592534 [Accessed 6th September 2017].

IJSSELSTEIJN, W. A., SEUNTIENS, P. J. H. & MEESTERS, L. M. J. 2005. Human factors of 3D displays. *In:* SCHREER, O., KAUFF, P. & SIKORA, T. (eds.) *3D videocommunication:*

algorithms, concepts and real-time systems in human centred communication. New Jersey: John Wiley & Sons, pp. 219-233.

INTERNATIONAL ENGINEERING CONSORTIUM. 2007. *ProForums Tutorial: The Human-Machine Interface (HMI)* [Online]. Web ProForums. Available: http://www.iec.org/online/tutorials/hmi/index.asp [Accessed 9th February 2009.

INTERNATIONAL TELECOMMUNICATION UNION 2012. Methodology for the subjective assessment of the quality of television pictures.

ISO 1998. 9241-11:1998 - Ergonomic requirements for office work with visual display terminals (VDTs) - Guidance on usability.

ISO 2006. 9241-110:2006 - Ergonomics of human-system interaction -- Part 110: Dialogue principles.

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

ISO 2010a. 9241-210:2010 - Ergonomics of human-system interaction -- Part 210: Human-centred design for interactive systems.

ISO 2010b. 26022:2010 - Road vehicles — Ergonomic aspects of transport information and control systems — Simulated lane change test to assess in-vehicle secondary task demand.

JAHN, T. & BINZ, H. Early phase project maturity controlling of innovative systems in the automotive industry. Proceedings of ICED 2007, the 16th International Conference on Engineering Design, 2007.

JANDER, H., BORGVALL, J. & RAMBERG, R. 2012. Towards a Methodological Framework for HMI Readiness Evaluation. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting,* 56 (1), 2349-2353.

JAPAN AUTOMOBILE MANUFACTURERS ASSOCIATION 2004. Guidelines for In-Vehicle Display Systems version 3.0.

JOHNSON, R. B. & ONWUEGBUZIE, A. J. 2006. Mixed Methods Research: A Research Paradigm Whose Time Has Come. *In:* BRYMAN, A. (ed.) *Mixed methods volume II.* SAGE.

JORDAN, P. W. 2000. *Designing pleasurable products : an introduction to the new human factors,* London, Taylor & Francis.

KANO, N., SERAKU, N., TAKAHASHI, F. & TSUJI, S.-I. 1984. Attractive Quality and Must-Be Quality. *Journal of the Japanese Society for Quality Control,* 14 (2), 147-156.

KAPTEIN, N. A., THEEUWES, J. & HORST, R. V. D. 1996. Driving Simulator Validity: Some Considerations *Transportation Research Record: Journal of the Transportation Research Board,* 1550, 30-36.

KENNEDY, R. S., LANE, N. E., BERBAUM, K. S. & LILIENTHAL, M. G. 1993a. Simulator Sickness Questionnaire: An Enhanced Method for Quantifying Simulator Sickness. *The International Journal of Aviation Psychology* 3(3), 203-220.

KENNEDY, R. S., LANE, N. E., BERBAUM, K. S. & LILIENTHAL, M. G. 1993b. Simulator Sickness Questionnaire: An Enhanced Method for Quantifying Simulator Sickness. *The International Journal of Aviation Psychology,* 3 (3), 203-220.

KENNEDY, R. S., STANNEY, K. M. & DUNLAP, W. P. 2000. Duration and Exposure to Virtual Environments: Sickness Curves During and Across Sessions. *Presence: Teleoperators and Virtual Environments,* 9 (5), 463-472.

KERZNER, H. 2011. *Project Management Metrics, KPIs and Dashboards*, John Wiles & Sons.

KEYES, J. 2011. *Implementing the Project Management Balanced Scorecard*, CRC Press.

KHAN, T., PITTS, M. & WILLIAMS, M. A. 2016. Cross-Cultural Differences in Automotive HMI Design: A Comparative Study Between UK and Indian Users' Design Preferences. *Journal of Usability Studies,* 11 (2), 45-65.

KLAUER, S. G., DINGUS, T. A., NEALE, V. L., SUDWEEKS, J. D. & RAMSEY, D. J. 2006. The Impact of Driver Inattention on Near-Crash/Crash Risk: An Analysis Using the 100-Car Naturalistic Driving Study Data. National Highway Traffic Safety Administration. DOT HS 810 594.

LABOVITZ, G., CHANG, Y. S. & ROSANSKY, V. 1992. *Making quality work: a leadership guide for the results-driven manager,* Essex Junction, VT, Wiley.

LAMBOOIJ, M., FORTUIN, M., IJSSELSTEIJN, W. A. & HEYNDERICKX, I. 2009. Visual Discomfort and Visual Fatigue of Stereoscopic Displays: A Review. *Journal of Imaging Science and Technology,* 55 (3), 1-14.

LEWIS, J. & SAURO, J. 2009. The Factor Structure of the System Usability Scale. *In:* KUROSU, M. (ed.) *Human Centered Design.* Springer Berlin Heidelberg, pp. 94-103.

LEWIS, J. R. 2012. Usability Testing. *In:* SALVENDY, G. (ed.) *Handbook of Human Factors and Ergonomics.* 4th ed., pp. 1267-1312.

LIM, J. 2011. Hedonic scaling: A review of methods and theory. *Food Quality and Preference,* 22 (8), 733-747.

MANES, D., GREEN, P. 1997. Evaluation of a driver interface: effects of control type (knob versus buttons) and menu structure (depth versus breadth). University of Michigan, Ann Arbor, Transportation Research Institute, Human Factors Division.

MANIAK, R., MIDLER, C., BEAUME, R. & VON PECHMANN, F. 2014. Featuring capability: How carmakers organize to deploy innovative features across products. *Journal of Product Innovation Management,* 31 (1), 114-127.

MARTINEZ, M. 2017. *How Ford pulled up its quality score* [Online]. Automotive News. Available: http://www.autonews.com/article/20170625/OEM06/170629840/ford-quality-jdpower-ranking [Accessed 20th August 2017].

MCCLELLAND, I. 1995. Product assessment and user trials. *In:* WILSON, J. R. & CORLETT, E. N. (eds.) *Evaluation of human work: A practical ergonmoics methodology.* 2nd ed.: CRC Press, pp. 249-284.

MCINTIRE, J. P., HAVIG, P. R. & GEISELMAN, E. E. 2014. Stereoscopic 3D displays and human performance: A comprehensive review. *Displays,* 35 (1), 18-26.

MEHLER, B., REIMER, B. & ZEC, M. 2012. Defining workload in the context of driver state detection and HMI evaluation. Proceedings of the 4th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, Portsmouth, New Hampshire, 2012: ACM, 187-191.

MERCEDES-BENZ. 2018. *World premiere at CES - MBUX: A completely new user experience for the new compact cars* [Online]. Available: https://media.daimler.com/marsMediaSite/ko/en/43019316 [Accessed 10th June 2019].

MILLER, S. 2001. Literature Review: Workload Measures. National Advanced Driving Simulator.

MORSE, L. C. & BABCOCK, D. L. 2010. *Managing engineering and technology : an introduction to management for engineers,* Upper Saddle River, NJ, Prentice Hall.

MURPHY, S. A. & KUMAR, V. 1997. The front end of new product development: a Canadian survey. *R&D Management,* 27 (1), 5-15.

NATIONAL HIGHWAY TRAFFIC SAFETY ADMINISTRATION 2013. Visual-Manual NHTSA Driver Distraction Guidelines for In-Vehicle Electronic Devices. US Department of Transport. 2013-09883.

NATIONAL HIGHWAY TRAFFIC SAFETY ADMINISTRATION 2016. Visual-Manual NHTSA Driver Distraction Guidelines for Portable and Aftermarket Devices. US Department of Transport. 2016-29051.

NAUJOKS, F., WIEDEMANN, K., SCHÖMIG, N., HERGETH, S. & KEINATH, A. 2019. Towards guidelines and verification methods for automated vehicle HMIs. *Transportation Research Part F: Traffic Psychology and Behaviour,* 60, 121-136.

NIELSEN, J. & NORMAN, D. n.d. *The Definition of User Experience* [Online]. Nielsen Norman Group. Available: http://www.nngroup.com/articles/definition-user-experience/ [Accessed 3rd March 2015].

NILSSON, R. 2010. *Usability vs. User Experience (UX)* [Online]. NeoSpot. Available: http://www.neospot.se/usability-vs-user-experience/ [Accessed 10th March 2015].

NUNNALLY, J. C. & BERNSTEIN, I. H. 1994. *Psychometric Theory,* New York, McGraw-Hill.

OLIVER WYMAN 2007. 2015 Car Innovation: A comprehensive study on innovation in the automotive industry.

OMACHONU, V. K. & ROSS, J. 2004. *Principles of Total Quality,* Boca Raton, CRC Press.

OWENS, J. D. & DAVIES, J. The importance of a new product development (NPD) process: getting started. 1st European Conference on Knowledge Management, Bled, Slovenia, 2000.

PARK, J., HAN, S. H., KIM, H. K., OH, S. & MOON, H. 2013. Modeling user experience: A case study on a mobile device. *International Journal of Industrial Ergonomics,* 43 (2), 187-196.

PERYAM, D. R. & PILGRIM, F. J. 1957. Hedonic scale method of measuring food preferences. *Food technology*.

PETTITT, M., BURNETT, G. & STEVENS, A. 2007. An Extended Keystroke Level Model (KLM) for the Visual Demand of In-Vehicle Information Predicting Systems. *Conference on Human Factors in Computing Systems, Vol 1 and 2*, 1515-1524.

PHAAL, R., FARRUKH, C. & PROBERT, D. R. 2013. Technology Management and Roadmapping at the Firm Level. *In:* MOEHRLE, G. M., ISENMANN, R. & PHAAL, R. (eds.) *Technology Roadmapping for Strategy and Innovation: Charting the Route to Success.* Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 13-29.

PITTS, M. J., BURNETT, G. E., SKRYPCHUK, L., WELLINGS, T., ATTRIDGE, A. & WILLIAMS, M. A. 2012a. Visual-Haptic Feedback Interaction in Automotive Touchscreens. *Displays,* 33 (1), 7-16.

PITTS, M. J., HASEDŽIĆ, E., SKRYPCHUK, L., ATTRIDGE, A. & WILLIAMS, M. A. 2015. Adding Depth: Establishing 3D Display Fundamentals for Automotive Applications. SAE Technical Paper 2015-01-0147.

PITTS, M. J., SKRYPCHUK, L., ATTRIDGE, A. & WILLIAMS, M. A. Comparing the User Experience of Touchscreen Technologies in an Automotive Application. Proceedings of the 6th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, Seattle, WA, USA, 2014: ACM, 1-8.

PITTS, M. J., SKRYPCHUK, L., WELLINGS, T., ATTRIDGE, A. & WILLIAMS, M. A. 2012b. Evaluating User Response to In-Car Haptic Feedback Touchscreens Using the Lane Change Test. *Advances in Human-Computer Interaction,* 2012, 13.

PROBERT, D. R., PHAAL, R. & FARRUKH, C. J. P. 2000. Development of a structured approach to assessing technology management practice. *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture,* 214 (4), 313-321.

RASKIN, J. 2000. *The humane interface : new directions for designing interactive systems,* Reading, MA, Addison-Wesley.

REED, M. P., GREEN, P.A. 1999. Comparison of driving performance on-road and in low-cost simulator using a concurrent telephone dialing task. *Ergonomics,* 42, 1015-1037.

REGAN, M. A., HALLETT, C. & GORDON, C. P. 2011. Driver distraction and driver inattention: Definition, relationship and taxonomy. *Accident Analysis & Prevention,* 43 (5), 1771-1781.

REIMER, B., MEHLER, B., DOBRES, J. & COUGHLIN, J. F. 2013. The Effects of a Production Level "Voice-Command" Interface on Driver Behavior: Summary Findings on Reported Workload, Physiology, Visual Attention, and Driving Performance. MIT AgeLab. Tech. rep. 2013-17A.

ROBBINS, J. 2012. *Consumer Technology Forces Carmakers To Satisfy Demand, Not Create It* [Online]. Industry Week. Available: http://www.industryweek.com/product-development/consumer-technology-forces-carmakers-satisfy-demand-not-create-it [Accessed 18th May 2013 2013].

ROSS, J. 1999. *Total Quality Management,* New York, Routledge.

RYDSTRÖM, A., BENGTSSON, P., GRANE, C., BROSTRÖM, R., AGARDH, J. & NILSSON, J. Multifunctional systems in vehicles: a usability evaluation. *In:* THATCHER, A., JAMES, J. & TODD, A., eds. CybErg 2005. The Fourth International Cyberspace Conference on Ergonomics, Johannesburg, SA, 2005: International Ergonomics Association Press.

SAE 2004. J2364 - Navigation and Route Guidance Function Accessibility While Driving. SAE International.

SAUER, J., SEIBEL, K. & RÜTTINGER, B. 2010. The influence of user expertise and prototype fidelity in usability tests. *Applied Ergonomics,* 41 (1), 130-140.

SAUER, J. & SONDEREGGER, A. 2009. The influence of prototype fidelity and aesthetics of design in usability tests: Effects on user behaviour, subjective evaluation and emotion. *Applied Ergonomics,* 40 (4), 670-677.

SAURO, J. & LEWIS, J. R. 2012. *Quantifying the User Experience,* Boston, Morgan Kaufmann.

SCHILLING, M. A. 2010. *Strategic Management of Technological Innovation,* New York, NY, McGraw-Hill.

SEARS, A., PLAISANT, C. & SHNEIDERMAN, B. 1992. A New Era for High Precision Touchscreens. *In:* H.R. HARTSON & D. HIX (eds.) *Advances in Human-Computer Interaction.* New Jersey: Ablex Publishing, pp. 1-33.

SHUTKO, J. & TIJERINA, L. 2011. Ford's Approach to Managing Driver Attention: SYNC and MyFord Touch. *Ergonomics in Design,* 19 (4), 13-16.

SPARKES, M. 2015. WhatsApp overtakes text messages. *The Telegraph*, 12th Jan.

STANTON, N. A., SALMON, P. M., WALKER, G. H., BABER, C. & JENKINS, D. P. 2013. *Human Factors Methods: A Practical Guide for Engineering And Design,* Farnham, England, Ashgate Publishing Ltd.

STONE, H. & SIDEL, J. L. 2004. *Sensory evaluation practices,* California, Elsevier Academic Press.

STUTTS, J. C., REINFURT, D. W., STAPLIN, L. & RODGMAN, E. A. 2001. The role of driver distraction in traffic crashes. AAA Foundation of Traffic Safety.

SUMIE, M., LI, C. & GREEN, P. 1998. Usability of menu-based interfaces for motor vehicle secondary functions. University of Michigan, Ann Arbor, Transportation Research Institute, Human Factors Division.

TAVAKOL, M. & DENNICK, R. 2011. Making sense of Cronbach's alpha *International Journal of Medical Education,* 2, 53-55.

TEDDLIE, C. & TASHAKKORI, A. 2009. *Foundations of mixed methods research : integrating quantitative and qualitative approaches in the social and behavioural sciences*, SAGE.

THOMKE, S. & FUJIMOTO, T. 2000. The Effect of "Front-Loading" Problem-Solving on Product Development Performance. *Journal of Product Innovation Management,* 17 (2), 128-142.

TIDWELL, J. 2006. *Designing Interfaces,* Sebastopol, O'Reilly.

TIJERINA, L. 1996. Executive Summary: Heavy vehicle driver workload assessment. National Highway Traffic Safety Administration. DOT HS 808 466.

TÖNNIS, M., BROY, V. & KLINKER, G. A Survey of Challenges Related to the Design of 3D User Interfaces for Car Drivers. IEEE Symposium on 3D User Interfaces, 25-29 March 2006, 2006, 127-134.

TROTT, P. 2012. *Innovation management and new product development,* Harlow, England, Financial Times/Prentice Hall.

TULLIS, T. & ALBERT, B. 2008. *Measuring the user experience : collecting, analyzing, and presenting usability metrics,* Amsterdam, Elsevier.

VAN KUIJK, J. I., CHRISTIAANS, H. H. C. M., HKANIS, H. & VAN EIJK, D. J. 2006. Usability in the Development of Consumer Electronics: Issues and Actors. IEA2006: the 16th world congress on ergonomics, Maastricht, The Netherlands, July 10-14, 2006.

VENKATESH, V. 2000. Determinants of Perceived Ease of Use: Integrating Control, Intrinsic Motivation, and Emotion into the Technology Acceptance Model. *Information Systems Research,* 11 (4), 342-365.

VOLKSWAGEN. 2016. *World premiere in Wolfsburg: Volkswagen presents the new Golf* [Online]. Available: https://www.volkswagen-newsroom.com/en/press-releases/world-premiere-in-wolfsburg-volkswagen-presents-the-new-golf-1678 [Accessed 10th June 2019].

VREDENBURG, K., ISENSEE, S. & RIGHI, C. 2001. *User-Centered Design: An Integrated Approach*, Prentice Hall PTR.

WANOUS, J. P. & REICHERS, A. E. 1996. Estimating the reliability of a single-item measure. *Psychological Reports,* 78 (2), 631-634.

WELLINGS, T., PITTS, M. J. & WILLIAMS, M. A. 2012. Characterising the experience of interaction: an evaluation of automotive rotary dials. *Ergonomics,* 55 (11), 1298-1315.

WELLINGS, T., WILLIAMS, M. & TENNANT, C. 2010. Understanding customers' holistic perception of switches in automotive human-machine interfaces. *Applied Ergonomics,* 41 (1), 8-17.

WELLINGS, T., WILLIAMS, M. A. & PITTS, M. 2008. Customer perception of switch-feel in luxury sports utility vehicles. *Food Quality and Preference,* 19 (8), 737-746.

WICKENS, C. D. 2008. Multiple Resources and Mental Workload. *Human Factors,* 50 (3), 449-455.

WICKENS, C. D., HOLLANDS, J. G., BANBURY, S. & PARASURAMAN, R. 2013. *Engineering Psychology and Human Performance.*

WIERWILLE, W. W. 1993. Demands on driver resources associated with introducing advanced technology into the vehicle. *Transportation Research Part F: Traffic Psychology and Behaviour,* 1 (2), 133-142.

WOODCOCK, A., WELLINGS, T. & BINNERSLEY, J. 2012. A review of HMI issues experienced by early adopters of low carbon vehicles. *Advances in Human Aspects of Road and Rail Transportation.* CRC Press, pp. 20-29.

WU, F.-G., LIN, H. & YOU, M. 2011. Direct-touch vs. mouse input for navigation modes of the web map. *Displays,* 32 (5), 261-267.

YEE, S., NGUYEN, L., GREEN, P., OBERHOLTZER, J., MILLER, B. 2007. Visual, auditory, cognitive, and psychomotor demands of real in-vehicle tasks. University of Michigan, Ann Arbor, Transportation Research Institute.

YEH, T. M., PAI, F. Y. & YANG, C. C. 2010. Performance improvement in new product development with effective tools and techniques adoption for high-tech industries. *Quality & Quantity,* 44 (1), 131-152.

YOUNG, K., REGAN, M. & HAMMER, M. 2003. Driver Distraction: A Review of the Literature. Monash University Accident Research Centre. Rep. No. 206.

YOUNG, M. S., BROOKHUIS, K. A., WICKENS, C. D. & HANCOCK, P. A. 2015. State of science: mental workload in ergonomics. *Ergonomics,* 58 (1), 1-17.

YOUNG, M. S. & STANTON, N. A. 2005. Mental workload. *In:* STANTON, N. A., HEDGE, A., BROOKHUIS, K. A., SALAS, E. & HENDRICK, H. (eds.) *Handbook of Human Factors and Ergonomics Methods.* London: CRC Press, pp. 39-1-39-9.

ZAYCHIK, K. B. & CARDULLO, F. M. 2003. Simulator Sickness: the Problem Remains. AIAA Modeling and Simulation Technologies Conference and Exhibit, Austin, Texas, 11-14 August, 2003.

Appendix 1 Automotive Technology and Manufacturing Readiness Level Definitions

Source: "Automotive Technology and Manufacturing Readiness Levels: A guide to recognised stages of development within the Automotive Industry", Automotive Council UK, January 2011.

TRL	Technology Readiness	MRL	Manufacturing Readiness
1	• Basic Principles have been observed and reported • Scientific research undertaken. • Scientific research is beginning to be translated into applied research and development. • Paper studies and scientific experiments have taken place. • Performance has been predicted		
2	• Speculative applications have been identified. • Exploration into key principles is ongoing. • Application specific simulations or experiments have been undertaken. • Performance predictions have been refined		• A high level assessment of manufacturing opportunities has been made.
3	• Analytical and experimental assessments have identified critical functionality and/or characteristics. • Analytical and laboratory studies have physically validated predictions of separate elements of the technology or components that are not yet integrated or representative. • Performance investigation using analytical experimentation and/or simulations is underway	1	• Basic Manufacturing Implications have been identified. • Materials for manufacturing have been characterised and assessed.

4	• The technology component and/or basic subsystem have been validated in the laboratory or test house environment. • The basic concept has been observed in other industry sectors (e.g. Space, Aerospace). • Requirements and interactions with relevant vehicle systems have been determined	**2**	• Manufacturing concepts and feasibility have been determined and processes have been identified. • Producibility assessments are underway and include advanced design for manufacturing considerations.
5	• The technology component and/or basic subsystem have been validated in relevant environment, potentially through a mule or adapted current production vehicle. • Basic technological components are integrated with reasonably realistic supporting elements so that the technology can be tested with equipment that can simulate and validate all system specifications within a laboratory, test house or test track setting with integrated components • Design rules have been established. • Performance results demonstrate the viability of the technology and confidence to select it for new vehicle programme consideration.	**3**	• A manufacturing proof-of-concept has been developed • Analytical or laboratory experiments validate paper studies. • Experimental hardware or processes have been created, but are not yet integrated or representative. • Materials and/or processes have been characterised for manufacturability and availability. • Initial manufacturing cost projections have been made. • Supply chain requirements have been determined.
6	• A model or prototype of the technology system or subsystem has been demonstrated as part of a vehicle that can simulate and validate all system specifications within a test house, test track or similar operational environment. • Performance results validate the technology's viability for a specific vehicle class.	**4**	• Capability exists to produce the technology in a laboratory or prototype environment. • Series production requirements, such as in manufacturing technology development, have been identified. • Processes to ensure manufacturability, producibility and quality are in place and are sufficient to produce demonstrators. • Manufacturing risks have been identified for prototype build. • Cost drivers have been confirmed. • Design concepts have been optimised for production. • APQP processes have been scoped and are initiated.

7	• Multiple prototypes have been demonstrated in an operational, on-vehicle environment. • The technology performs as required. • Limit testing and ultimate performance characteristics are now determined. • The technology is suitable to be incorporated into specific vehicle platform development programmes.	**5**	• Capability exists to produce prototype components in a production relevant environment. • Critical technologies and components have been identified. • Prototype materials, tooling and test equipment, as well as personnel skills have been demonstrated with components in a production relevant environment. • FMEA and DFMA have been initiated.
8	• Test and demonstration phases have been completed to customer's satisfaction. • The technology has been proven to work in its final form and under expected conditions. • Performance has been validated, and confirmed	**6**	• Capability exists to produce integrated system or subsystem in a production relevant environment. • The majority of manufacturing processes have been defined and characterised. • Preliminary design of critical components has been completed. • Prototype materials, tooling and test equipment, as well as personnel skills have been demonstrated on subsystems/ systems in a production relevant environment. • Detailed cost analyses include design trades. • Cost targets are allocated and approved as viable. • Producibility considerations are shaping system development plans. • Long lead and key supply chain elements have been identified.

9	• The actual technology system has been qualified through operational experience. • The technology has been applied in its final form and under real-world conditions. • Real-world performance of the technology is a success. • The vehicle or product has been launched into the market place. • Scaled up/down technology is in development for other classes of vehicle.	7	• Capability exists to produce systems, subsystems or components in a production representative environment. • Material specifications are approved. • Materials are available to meet planned pilot line build schedule. • Pilot line capability has been demonstrated including run at rate capability. • Unit cost reduction efforts are underway. • Supply chain and supplier Quality Assurances have been assessed. • Long lead procurement plans are in place. • Production tooling and test equipment design & development has been initiated • FMEA and DFMA have been completed.
		8	• Initial production is underway • Manufacturing and quality processes and procedures have been proven in production environment. • An early supply chain is established and stable. • Manufacturing processes have been validated.
		9	• Full/volume rate production capability has been demonstrated. • Major system design features are stable and proven in test and evaluation. • Materials are available to meet planned rate production schedules. • Manufacturing processes and procedures are established and controlled to three-sigma or some other appropriate quality level to meet design characteristic tolerances in a low rate production environment. • Manufacturing control processes are validated. • Actual cost model has been developed for full rate production.

10	• The technology is successfully in service in multiple application forms, vehicle platforms and geographic regions. In-service and life-time warranty data is available, confirming actual market life, time performance and reliability	**10**	• Full Rate Production is demonstrated • Lean production practices are in place and continuous process improvements are on-going. • Engineering/design changes are limited to quality and cost improvements. • System, components or other items are in rate production and meet all engineering, performance, quality and reliability requirements. • All materials, manufacturing processes and procedures, inspection and test equipment are in production and controlled to six-sigma or some other appropriate quality level. • Unit costs are at target levels and are applicable to multiple markets. • The manufacturing capability is globally deployable

Appendix 2 Technology Scorecard

TECHNOLOGY SCORECARD

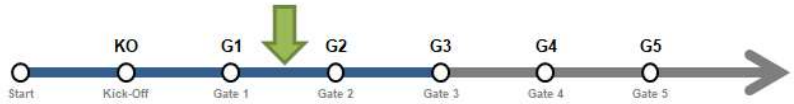
Project title: TS-0001 Touchscreen Technology

Owner: A. Person **CDSID:** aperson

Report date: 01/01/2017

Data Set	Supplier	ID	Eval date	Sample Size
Capacitive TS	Supplier 1	12345	01/11/16	18
Resistive TS	Supplier 2	23456	01/11/16	18
Infra-red TS	Supplier 3	34567	01/11/16	18

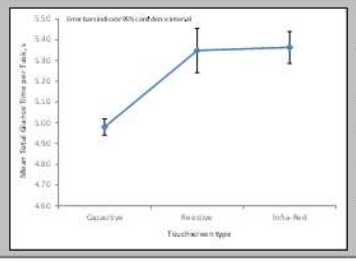
Evaluation stage



Description: Comparison of capacitive, resistive and infra-red touchscreen demonstrator units built to spec and using an interactive evaluation UI.

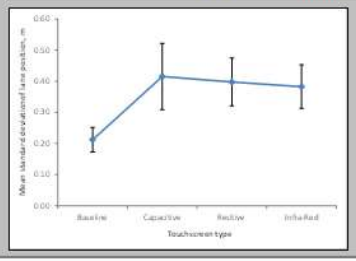
Objective Measures

Glance Time per Task



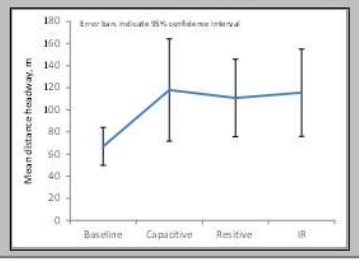
Data Set	Status	Red Tasks
Capacitive		1/10
Resistive		1/10
Infra-Red		1/10

Lateral Lane Deviation



Data Set	Status
Capacitive	
Resistive	
Infra-Red	

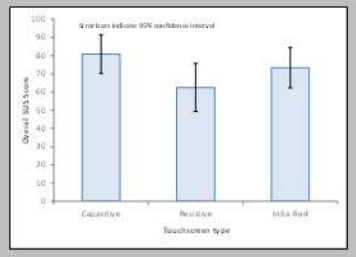
Longitudinal Deviation



Data Set	Status
Capacitive	
Resistive	
Infra-Red	

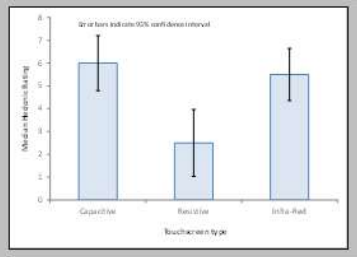
Subjective Measures

Usability



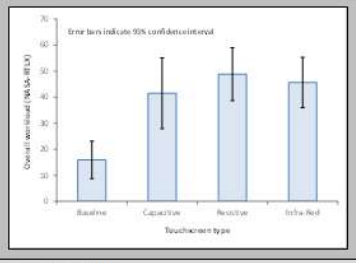
Data Set	Status
Capacitive	
Resistive	
Infra-Red	

Hedonic Rating



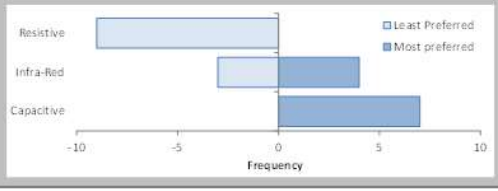
Data Set	Status
Capacitive	
Resistive	
Infra-Red	

Workload



Data Set	Status
Capacitive	
Resistive	
Infra-Red	

User Preference



Key User Experience Considerations

- Resistive display required higher input force, this affected the experience of scrolling tasks
- Tendency for a list scroll gesture to result in an inadvertent selection
- Resistive display feel described as 'papery' and 'cheap'
- IR and Capacitive more responsive, leading to improved confidence and perceived quicker response
- 11/18 users found IR display less responsive than Capacitive
- Specific usability issue with buttons in corners of IR display; some tasks required multiple button presses