DOCTOR OF PHILOSOPHY

The Application of Physiological Metrics in Validating User Experience Evaluation on Automotive Human Machine Interface Systems

Furness, Jane

Award date:
2016

Awarding institution:
Coventry University

The Application of Physiological Metrics in Validating User Experience Evaluation on Automotive Human Machine Interface Systems

By

Jane Furness

PhD

May 2016

Coventry University

A thesis submitted in partial fulfilment of the University's requirements for the Degree of Doctor of Philosophy /Master of Philosophy/Master of Research

The Application of Physiological Metrics in Validating User Experience Evaluation on Automotive Human Machine Interface Systems

By

Jane Furness

May 2016

A thesis submitted in partial fulfilment of the University's requirements for the Degree of Doctor of Philosophy /Master of Philosophy/Master of Research

The following ethics request has been approved by Sophie Krumins. All the relevant documentation will be available for you to download within the next 24 hours. Please log back into Ethics and select the request from your listing. Select the Downloads tab to retrieve the documentation.

Please proceed with good ethics.

Ref:	P23519
Project Title:	To Assess User Engagement when Interacting with Touch Screen Automotive Interfaces
Applicant:	Jane Furness
Supervisor:	Graham Shelton-Rayner
Module Code:	
Module Leader:	

Go to ethics.coventry.ac.uk to view this request in more detail.

The following ethics request has been approved by Elaine Cartmill. Please proceed with good ethics.

Ref:	P7949
Project Title:	Investigation into the use of physiological and psychological data to validate or correct subjective testing of Human-Machine Inferface Systems (HMIs)
Applicant:	Jane Furness
Supervisor:	Helen Maddock
Module Code:	
Module Leader:	

The following ethics request has been approved by Elaine Cartmill. All the relevant documentation will be available for you to download within the next 24 hours. Please log back into Ethics and select the request from your listing. Select the Downloads tab to retrieve the documentation.

Please proceed with good ethics.

Ref:	P13879
Project Title:	Exploring Physiological Markers for Human Machine Interface Evaluation
Applicant:	Jane Furness
Supervisor:	Helen Maddock
Module Code:	HLSR006
Module Leader:	

Go to ethics.coventry.ac.uk to view this request in more detail.

THIS MESSAGE HAS BEEN GENERATED AUTOMATICALLY - PLEASE DON'T REPLY TO THIS MESSAGE

Section 3 Submission Declaration		
	Yes	No
Have materials contained in your thesis/submission been used for any other submission for an academic award?	☐	X☐
If you have answered Yes to above please state award and awarding body and list the material:		
I am aware of no health reasons that will prevent me from undertaking and completing assessment and will undertake to notifying my Director of Studies and PGRSU - Registry as soon as any change in these circumstances occurs.	☐X	☐
Ethical Declaration: I declare that my research has full University Ethical approval and evidence of this has been included within my thesis/submission. Please also insert ethics reference number below Project Reference: **P7949** **P13879** **P23519**	X☐	☐

Freedom of Information:

Freedom of Information Act 2000 (FOIA) ensures access to any information held by Coventry University, including theses, unless an exception or exceptional circumstances apply.

In the interest of scholarship, theses of the University are normally made freely available online in CURVE, the Institutions Repository, immediately on deposit. You may wish to restrict access to your thesis for a period of three years. Reasons for restricting access to the electronic thesis should be derived from exemptions under FOIA. (Please also refer to the University Regulations Section 8.12.5)

Do you wish to restrict access to thesis/submission: Choose an item.
No
If Yes please specify reason for restriction:

Does any organisation, other than Coventry University, have an interest in the Intellectual Property Rights to your work?
Yes
Choose an item.
If Yes please specify Organisation:
Jaguar Land Rover
Please specify the nature of their interest:
Industrial Sponsors of the PhD project

Candidates Signature: [redacted]	**Date:** 10th May 2017

Acknowledgements

My warmest thanks to Professor Helen Maddock, for her encouragement, care and support during the last few years. To Dr Dale Richards, my Director of Studies, I give my sincere thanks and appreciation for all the help, support and guidance you have given to me, thanks to Dr Graham Shelton-Rayner who has always made time to advise and discuss matters with me. I would also like to thank the staff from Coventry University and students who have helped and given their time to support me in this work. I am very grateful that I have met you all, during these interesting years.

I would like to thank my sponsors, Jaguar Land Rover and my industrial supervisor Dr Laura Millen, who I always look forward to meeting, in addition to Laura, I would also like to convey my thanks to Lee Skrypchuk, Jim Braithwaite and Sebastian Paszkowicz, for their support, it has been a pleasure to be involved with your team.

Finally, thanks to my family for their support, help and encouragement, to Nigel and my three wonderful girls who make my life very special.

Abstract

Automotive in-vehicle information systems have seen an era of continuous development within the industry and are recognised as a key differentiator for prospective customers. This presents a significant challenge for designers and engineers in producing effective next generation systems which are helpful, novel, exciting, safe and easy to use. The usability of any new human machine interface (HMI) has an implicit cost in terms of the perceived aesthetic perception and associated user experience. Achieving the next engaging automotive interface, not only has to address the user requirements but also has to incorporate established safety standards whilst considering new interaction technologies.

An automotive (HMI) evaluation may combine a triad of physiological, subjective and performance-based measurements which are employed to provide relevant and valuable data for product evaluation. However, there is also a growing interest and appreciation that determining real-time quantitative metrics to drivers' affective responses provide valuable user affective feedback.

The aim of this research was to explore to what extent physiological metrics such as heart rate variability could be used to quantify or validate subjective testing of automotive HMIs. This research employed both objective and subjective metrics to assess user engagement during interactions with an automotive infotainment system. The mapping of both physiological and self-report scales was examined over a series of studies in order to provide a greater understanding of users' responses. By analysing the data collected it may provide guidance within the

early stages of in-vehicle design evaluation in terms of usability and user satisfaction. This research explored these metrics as an objective, quantitative, diagnostic measure of affective response, in the assessment of HMIs. Development of a robust methodology was constructed for the application and understanding of these metrics.

Findings from the three studies point towards the value of using a combination of methods when examining user interaction with an in-car HMI. For the next generation of interface systems, physiological measures, such as heart rate variability may offer an additional dimension of validity when examining the complexities of the driving task that drivers perform every day.

There appears to be no boundaries on technology advancements and with this, comes extra pressure for car manufacturers to produce similar interactive and connective devices to those that are already in use in homes. A successful in-car HMI system will be intuitive to use, aesthetically pleasing and possess an element of pleasure however, the design components that are needed for a highly usable HMI have to be considered within the context of the constraints of the manufacturing process and the risks associated with interacting with an in-car HMI whilst driving. The findings from the studies conducted in this research are discussed in relation to the usability and benefits of incorporating physiological measures that can assist in our understanding of driver interaction with different automotive HMIs.

Contents

List of Abbreviations

ADAS	Advanced Driver Assistance System
ANET	Affective Norms for English
ANEW	Affective Norms for English Words
ANS	Autonomic Nervous System
AV	Atrioventricular
CNS	Central Nervous System
EDA	Electrodermal Activity
ECG	Electrocardiograph
EEG	Electroencephalogram
EMG	Electromyogram
HCI	Human-computer Interaction
HMI	Human-machine Interface
HR	Heart Rate
HRV	Heart Rate Variability
HUD	Head-up Display
ISO	the International Organisation for Standardisation
IADS	International Affective Digital System
IAPS	International Affective Picture System

IVIS	In-vehicle Information System
LF	Low Frequency
HF	High Frequency
LF/HF	Low Frequency / High Frequency Ratio
MM	Mental Models
MRT	Multiple Resource Theory
NPY	Neuropeptide Y
OAA	Open Automotive Alliance
PASP	Peripheral Autonomic Surface Potential
PNS	Parasympathetic Nervous System
SA	Situational Awareness
SAM	Self-Assessment Manikin
SNS	Sympathetic Nervous System
UX	User Experience
VDU	Visual Display Unit

1 Chapter One: General Introduction

1.1 Overview

For automotive manufacturers to succeed in this highly competitive market, companies must be innovative and dynamic in nature to face the economic and environmental challenges that are part of this global economy. In 2015 The United Kingdom (UK) manufactured over 1.5 million cars, which was an increase of 3.9 per cent on 2014 (The Society of Motor Manufacturers and Traders, (SMMT) 2016). This rise has also been seen in new car sales, with the highest recorded number of over 2.6 million (SMMT 2015), Figure 1. A recent report speculates that UK car production levels by 2020 will be at a record of two million vehicles (Henry 2015). The UK automotive industry contributes significantly to the overall UK export market and accounts for over 11 per cent of the total UK exports (SMMT 2015).

With a large industry of this nature and the important ramifications that are connected to it, such as employment and secondary support industries, it is imperative for the longevity of success that the UK automotive industry continues to produce vehicles that match the expectation of today’s drivers. Companies are actively developing differentiators that will influence potential clients. In-vehicle information systems (IVIS) and Advanced Driver Assistance Systems (ADASs) are two areas that have attracted new developments, which have seen rapid advancements (Amditis and Polychronopoulos 2004, Mitsopoulos-Rubens, Trotter and Lenne 2011).

Over the last decade this increase in such technologies has been found to be both desirable for consumers and seen as influential in their decision making when choosing a specific car (Amditis *et al.* 2010). More recent advances have even seen development in getting closer to connected and autonomous vehicle technology. In March 2015 the UK Government announced that £200 million would be designated to these two important areas of automotive development.

Information systems and ADASs now deploy a wide range of applications such as climatic control, satellite navigation systems, adaptive cruise control, intelligent speed adaption, telecommunications and entertainment systems (Shirley *et al.* 2009). The challenge for the industry is to design systems which are novel, exciting and user friendly. This challenge is met by a collaborative partnership across different departments within the automotive companies. Of primary importance in these designs is user acceptance of new in-vehicle interfaces (Carsten and Nilsson 2001). It is therefore essential that the design and implementation of information systems ensures that safety is not compromised by aesthetic and complex novel design features, or merely driven by the technologists without appreciating the user needs (Lavie *et al*. 2011).

Traditionally automotive manufacturers ensure a degree of user feedback is factored into the design process of new HMIs, although, this tends to be subjective assessment (and in some cases embodies a lack of standardised approaches to usability assessment). There are several indices of human performance that are used when assessing driver interaction; predominantly these would focus on performance-based factors, subjective and physiological measurements (O'Donnell and Eggemeier 1986).

Although measurements around these factors have produced important and valuable data, there is a growing interest in automotive research for a diagnostic tool that can provide real-time quantitative metrics to drivers' affective responses when interacting with HMI systems (Picard, Vyzas and Healey 2001).

With the prospect of fully autonomous vehicles now becoming a reality, rather than a designer's futuristic concept, there is a real need for designers and engineers to use technologies to provide detailed information relating to events from the external environment, to ensure that journeys are safe in a plethora of different road environments. For many automotive designers, engineers and automotive companies advanced technologies that monitor the vehicle instrumentations and outside environments are already in place, providing the driver with useful assistance and providing, in some instances information to the driver in relation to what the car is doing. However, perhaps what is lagging behind the development of this technology, is the ability to monitor and provide key information on the driver's physiological state that could be relayed back to the car in order to achieve an efficient partnership between human and system. An example of this may be during the 'handover' stage from autonomous driving back to the driver. During this stage it would be crucial that the driver is 'prepared' to accept control of the car.

Beyond the use of physiological monitoring to assist the driver in operating the car, another key area where these objective metrics may be used is in gaining a holistic picture of the driver from a consumer perspective. Non-invasive physiological monitoring of the driver is an exciting area for development and with the prospect of drivers spending more time interacting with applications through their infotainment system, there is a real need to explore and gain an understanding of the driver's user experience (UX) and associated affective states, (such as frustration and enjoyment). By understanding this user state, we can then examine how an automotive HMI can influence the decisions road users make when driving their vehicles.

1.2 The importance of considering the driver's cognitive processing

Today's vehicles use complex technologies and for the driver this may involve performing over 1,600 individual tasks in a traffic and road environment (Walker, Stanton and Salmon 2015). This is perhaps overlooked due to the ubiquitous nature of cars and the relative ease of driving shown by so many individuals (Walker, Stanton and Salmon 2015).

Driving a modern vehicle involves participating in multiple secondary tasks, (such as interacting with the radio), which often contributes in distraction from the primary task (driving the vehicle) and may lead to road traffic accidents (Gable *et al.* 2013). To assist the engineering and design team in minimising visual distraction from the road a number of cognitive constructs should be considered. In automotive driving research, the act of driving has often been used as a model when theories are constructed due to the multi-tasking features employed when driving a vehicle (Walker, Stanton and Salmon 2015).

When designers and engineers are developing automotive in-vehicle interface systems, of prime importance is how the user will interact with the system and the level of distraction that interaction has on the user's primary task of driving the vehicle. With today's user interface systems being highly complex it is often a difficult balance for designers and engineers to develop and produce systems that encompass new technologies which are helpful to the driver but at the same time relatively simplistic in concept to minimise visual distraction from the road. To achieve this delicate but essential balance it is imperative that designers and engineers draw knowledge from theoretical models and constructs of cognitive processes that influence the way in which an individual interacts with an in-vehicle HMI.

In the evaluation process of HMIs, it is important that our understanding of these cognitive processes is not neglected during the design of user trials and certainly considered when analysing the results. By considering cognitive processes, meaningful and insightful findings can be fed back into the design process, thus facilitating improvements that may be implemented before the final design is determined.

The main cognitive processes that have a major impact on the way a user interacts with the HMIs are; Perceptual Processing of Visual Information, The Multiple Resource Theory, Workload, Situation Awareness and Mental Models. Each of the above concepts will be discussed further with its relevance to the design and evaluation process of automotive HMIs.

1.2.1 Perceptual Processing of Visual Information

How do we see, what we see? This small but fundamental question has been the root of multidisciplinary research for centuries. From discovering the physiological structures that transmit the visual images, to constructing theories and models that explain the interpretation of these ‘seen’ images. The research fields of neurosciences and psychology are perhaps the most dominant partners in this research, with a general consensus that for individuals to perceive meaning from their visual environment a set of complex operations has to be undertaken by the visual cortices of the brain. Our visual processing not only recognises objects in the environment but enables us to interact successfully within it. For these two distinct features of visual processing to occur there are separate neural pathways that interact and run parallel with each other. Although parallel pathways exist, one of the most difficult questions to answer is how the information from the different pathways converge together to produce the visual images that we see and recognise throughout our lives. This recognised cognition conundrum is termed the ‘binding problem’ (Treisman 1996).

Historically, knowledge on these different pathways has been determined by a number of studies carried out on primates and later with humans as developing technologies emerged (Gilbert 2013: 556-576). Functional magnetic resonance imaging (fMRI) is an example of one of these technologies that has enabled researchers to establish similarities between the visual sites of the brain. Based on neural pathway studies it is now known that two hierarchical pathways exist (Gilbert 2013: 556-576). The pathway that determines object recognition is the ventral pathway and the dorsal pathway uses visual information for guiding movements (Gilbert 2013: 556-576).

Our eyes focus the visual image onto the retina, which is a specialised layer of cells containing the photoreceptor cells which convert light photons to electrical signals. The electrical signals are then sent, via the optic nerve to the higher centres in the brain for further perceptual processing (Meister and Tessier-Lavigne 2013: 577-600).

Visual perception is the process of interpreting incoming visual information from the external environment and creating an internal representation of that information. The process of how our brain constructs whole pictures from the massive amount of incoming signals has been debated for many years with researchers designing studies to provide evidence for the many different theories that have arisen. Important contributors to perceptual research included Gibson (1966) who proposed a bottom-up processing theory where the capacity to interact with our environment was due to direct perception as an individual receives sufficient information from the environment to form a sense of the world that they live in. Gibson spent a considerable time examining motion and observed the problems pilots had as they attempted to take-off and land their aircraft. This led him to believe in optic flow patterns where the focussed point of interest remains still but the surrounding environment moves in an outwardly direction.

In contrast to Gibson's theory was the view that perception was a top-down process where memories and expectations enable us to interpret sensory data from the external environment. This top-down process and view of perception was proposed by Gregory (1970). Gregory believed that when we view a scene or object we construct a perceptual hypothesis from prior knowledge and experiences that we have encountered before. Gregory estimated that 90 per cent of the information that was 'seen' was lost by the time it reached the brain. He also discussed that although our perceptual hypotheses were almost always correct there were some instances where hypotheses were incorrect which led to mistakes in our perception of certain objects and could be explained by visual illusions.

Work carried out by the British neuroscientist and psychologist David Marr (1982) culminated in a framework for the process of vision where he explained that the vision process is modular. In his framework he describes three major stages in forming perception; the first is termed Primal Sketch where main features determined by intensity changes are processed, such as object boundaries. The second stage is a viewer-centred representation, termed 2.5D where properties of the visible surfaces are processed; this would include orientation, contours and depth. The final stage is the 3-D model which is an object-centred representation of three dimensional objects', this representation allows for the recognition of whole objects.

Hubel and Weisel's (1959) work on neural activity in the primary visual cortex led to a greater understanding of information processing and in particular the importance of orientation of visual images (see Hubel and Wiesel 1959). The seminal work of Anne Treisman (1982) focussed on how we could group visual features together in order to form objects, and how that information is selected and integrated to form objects that we recognise. In attentional studies she proposed the Feature Integration Theory of Attention (Treisman and Gelade 1980) and suggested that features, such as colour and shape were first to be perceived. These features are registered early and automatically, namely pre-attention, while objects are identified separately at a later stage. This later stage requires focused attention. The Feature Integration Theory suggests that there are two ways to become aware of objects. The first requires focused attention which is directed serially to different locations in order to integrate separate features into one concept. The second way that we become aware of objects is when focused attention is perhaps overloaded or the exposure of the object was too short. In this way top-down processing occurs and familiar objects can be predicted although in this process errors may be produced when the environment is not familiar.

Pioneers in attention research emerged during the 1950s and 1960s with two different theories; the first being Broadbent's Filter theory, which was often referred to as the early selection theory. Broadbent hypothesised that all stimuli reaching the sensory system would be processed until the point where physical descriptions were analysed and represented (e.g. loudness, pitch of auditory stimuli), at this point the processes that identified the stimuli (e.g. recognition and understanding of a word) were only capable of processing one stimulus at a time. Broadbent drew the conclusion that there was a selective filtering mechanism, that determined which stimuli would be processed further. The word 'early' relates to where the selection occurs in the stages of processing (Pashler 1999). Figure 1 represents the Early Selection Theory. The first diagram demonstrates the time course when processing two stimuli, where Stimulus 1 (S1) is to be deliberately attended. According to Broadbent's early selection theory the processing system would analyse the physical descriptions of both S1 and S2 immediately, followed by identification of S1 but not S2. The second diagram illustrates the sequence of processing when an individual attempt to recognise two stimuli (S1, S2). This process is more complex and involves processing in parallel until physical descriptions are gained, then one stimulus will be identified first followed by identification of the second.

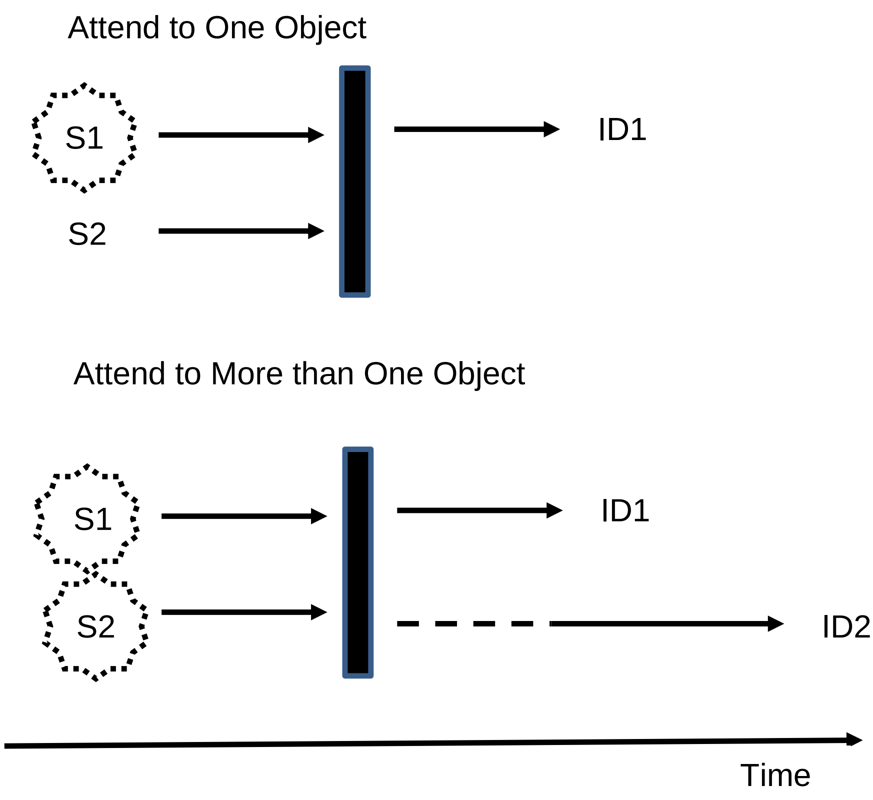


Figure 1. Time course of processing, Early Selection Theory
when either one or two stimuli are being attended (Adapted from Pashler 1999).

The second theory of attention was named the Late Selection Theory and was proposed by Deutsch and Deutsch in 1963, a more recent version of this theory was also discussed by Duncan in 1980 (Pashler 1999). In this theory recognition of stimuli proceeds unselectively with no capacity limitations. Selective processing in this theory only begins once analysis has been completed. The selected stimulus then gains access to a system which is required for awareness, memory and response. Figure 2 illustrates the Late Selection Theory, where S1 and S2 are presented at the same time, both would be processed in parallel to the point of identification, whether only one was being attended or both, shown in the second diagram.

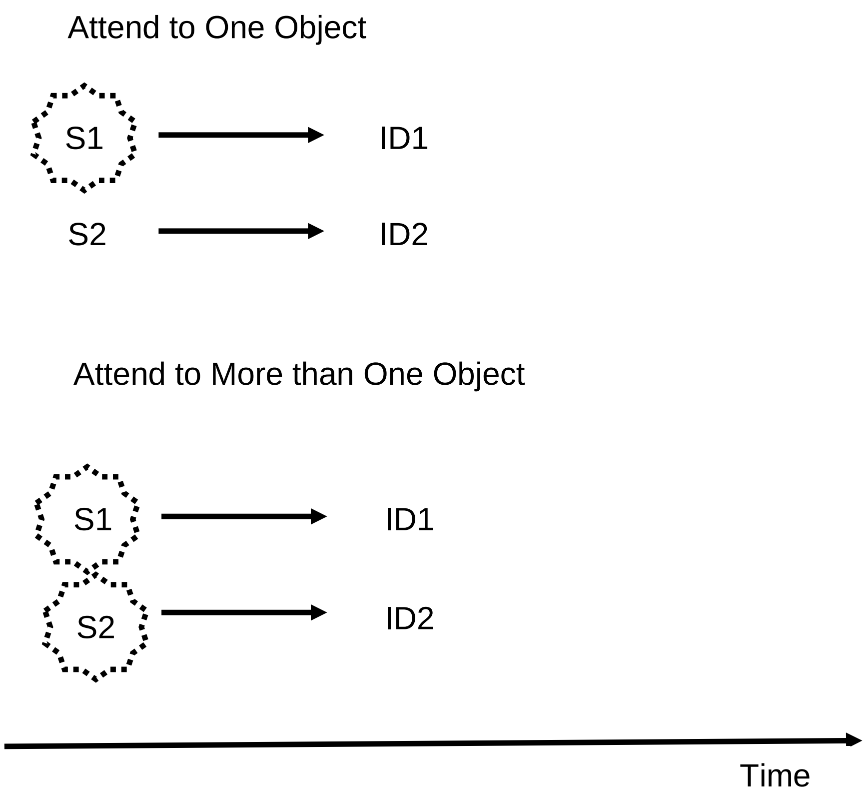


Figure 2. Time course of processing Late Selection Theory
(adapted from Pashler 1999)

Lavie's (1995) model of selective attention, termed the Perceptual Load Theory arose from the contradictory evidence from the 'early' and 'late' theorists and has provided a hybrid model. The Load Theory model proposes that individuals have a limited amount of attentional resources for processing incoming information or stimuli, (which is similar to early theorists' views), and this processing will proceed automatically until all the resources have been fully utilised (late theorists' views). The stage in the processing pathway for relevant information to be selected is therefore dependent on the level of perceptual load of the task. When the perceptual load is high with a task, the perceptual capacity is full and irrelevant information to the task is not perceived. Conversely, when the perceptual load is low and the task does not demand all the perceptual resources, irrelevant information can be processed (Lavie 1995).

Knowledge on how individuals process multiple stimuli and perform multiple tasks has a very real practical implication for automotive designers and engineers. Driving research continues to evaluate attention processing and how its relevance impacts on the design concepts of HMI systems. Driving often involves responding rapidly and accurately to a multitude of incoming stimuli from both the internal and external environment (Walker, Stanton and Salmon 2015). Internal distractors include the in-car infotainment system and perhaps passengers that expose the driver to additional stimuli that requires processing. The external environment that drivers have to consider include; traffic demands, weather condition, road and surface type and the route they have planned (Walker, Stanton and Salmon 2015).

As multi-tasking in vehicles increases, with complex state-of-the art infotainment HMI systems offering a variety of services, research into the benefits of multimodality are actively pursued. The theoretical concepts linked to multi-modality are MRT (see 1.2.2) or parallel processing capabilities. This was shown when a driver’s performance increased due to the processing demands being distributed via different modalities (Swette *et al.* 2013).

Automotive research actively explores the different modalities that information can be expressed by the in-car HMI system to the driver and methods by which the driver can communicate and interact with the system. By designing HMI systems that utilise non-visual interactions (such as voice activation, movement and gesturing by the driver) the driver may be able to attend to several different incoming stimuli without overloading the processing capacity and ultimately increase safety (Riener *et al*. 2013, Burnett *et al*. 2013).

A range of different modalities are being tested in studies to assess the different benefits, including driver distraction, workload and driver error. Modalities including; visual-haptic, speech dialog system (SDS) and gesturing appear to be the most promising for in-vehicle HMI interactions. Speech dialog systems have been shown to reduce driver's eyes-off-road time as they do not interfere with processing resources required for the primary task of driving as much as the visual-haptic interfaces (Wickens 2002). The SDS has also been regarded as a more intuitive method of interacting with a system as the driver can communicate directly with the system rather than relaying his communication into a menu-driven interface system (Hackenberg *et al.* 2013).

Gesturing by finger swiping a touchpad has been identified as a potential method of operating an eyes-free interface, which would reduce off-road glances and make a positive contribution to road safety. When testing an eyes-free interface, which was combined with a SDS, it was found that after initial practise, users were as accurate at gesturing as a comparable visual-haptic interface system (Zhao *et al.* 2007).

Visual-haptic in-vehicle interfaces have become commonplace in today's vehicles with some touchscreens now measuring over 42 centimetres (Rumelin and Butz 2013). With these screens being flat there is little or no haptic form to them, and drivers must use their eyes to locate the various functions that they offer (Rumelin and Butz 2013). Unless these large screens can combine a different modality there is a higher risk factor as the driver is taking their focus from the road ahead onto the screen (Rumelin and Butz 2013).

It is worth mentioning however, that by using a combination of modalities for interface interactions, driver performance does not always equate to a reduction in driver errors. Driver performance can be attributed to many different factors, and not always related to a driver taking their eyes off the road. As discussed above, weather, experience and other external factors can all contribute to driver performance.

1.2.2 The Multiple Resource Theory

Multiple Resource Theory (MRT) was a concept developed by Wickens (2002), whereby he expanded on Kahneman's work (1973). In his book Kahneman explore the different fields of thought on attention and how an individual's internal information processing system deals with incoming stimuli. Wickens (2002) devised a 4-D multiple resource model that can be applied to predict multitask workload overload. The 4-D model is based on the separate resources of information processing. Dimensions that are included in this model are: The Stages of Processing (Perception, Cognition and Responding); The Codes of Processing (Spatial and Verbal); The Modalities (Auditory and Visual) and the Visual Channels (Focal and Ambient). The MRT model can help in predicting the level of performance when two or more time shared tasks are being performed, making it a relevant model when an operator is in a high demand multitasking environment. Many of the trials that Wickens conducted involved vehicle drivers, as it was recognised that this group of individuals were frequently required to perform multiple tasks whilst driving a vehicle. The Multiple Resource Theory is not only applicable for research studies but can be applied in a practical way to assist designers and engineers in predicating differences in multitasking over different modalities.

1.2.3 Mental Workload

Mental workload is a related concept to the Multiple Resource Theory and although they may overlap they are distinctly different. Moray (1979) describes mental workload as the demand that tasks have on an individuals limited mental resources. This concept can be viewed from the perspective where the task demand is less than the capacity of resources available or that the task demands overload the capacity of resources. In the latter position the performance of the individual will be affected. If overload had been reached and the individual was performing multiple tasks, this would be an example where the MRT could contribute to mental workload and predict how much the performance will fail.

Despite research studies exploring workload for over 40 years, there is, as yet, no definitive universal accepted definition of this concept. Until the 1970s workload was not a commonly used term (Huey and Wickens, 1993). Features of workload can be generally categorised into three factors; the first factor relates to the quantity of work and the amount of tasks that require attention, the second factor focusses on the importance of time and the third factor is concerned with the psychological experiences that the operator may have. Workload is both multidimensional and multifaceted and due to this combination of different demands it is difficult to define uniquely. Although a consensus for a universal definition is not available, studies into measuring workload over a wide range of domains continues to provide interesting and relevant data on the mental cost of performing tasks under different situations. Data from workload studies can assist with predicting operator and system performance. By gathering data from workload studies improvements for the operator can be introduced, which may include; better working conditions, improved design of interface systems and more effective procedures.

1.2.4 Situational Awareness

This term was described by Endsley (1995) with regard to aircraft pilots and relates to an individual being able to identify, process and comprehend information as it is received. Situational awareness (SA) is linked to processing an accurate understanding of what is happening in close proximity as well as having a good understanding regarding the near future. Endsley (1988) defines SA as;

> *'The perception of the elements in the environment within a volume of time and space, the comprehension of their meaning and the projection of their status in the near future'.*

Situational awareness is a term which is applicable across different domains; an example of this would be an aircraft pilot and an anaesthesiologist. Although the job roles of these two professionals are very different, what they have in common is the quantity of incoming data that they are constantly being presented with, which has the potential to change rapidly. The aircraft pilot and anaesthesiologist must perceive and comprehend large quantities of data first before making correct decisions in 'real time' as well as anticipating future events that may occur. Figure 2 represents the different elements of Endsley's model of SA in dynamic decision making. Another important component for SA is the associated temporal aspects. Due to the nature of different professions, for example: an aircraft pilot and an anaesthetist, their situation can change constantly, requiring individual skills to be flexible, when decisions change quickly from one event to the next. In any complex, dynamic environment SA is a vital factor of an individual's performance (Endsley 1995). SA is a critical factor when assessing human performance in relation to driving. Nowadays drivers have an array of in-vehicle devices that they interact with via the HMI system as well as managing the external events that occur in the driving environment. Situational awareness is critical not only for the driver and the passenger's safety but also for other road users and pedestrians. An example of this would be a driver receiving an incoming-telephone call, as for a short period of time they are distracted from their primary task of driving, which results in missing some vital information relating to the position of a vehicle in front, this delay in receiving information may result in a collision occurring.

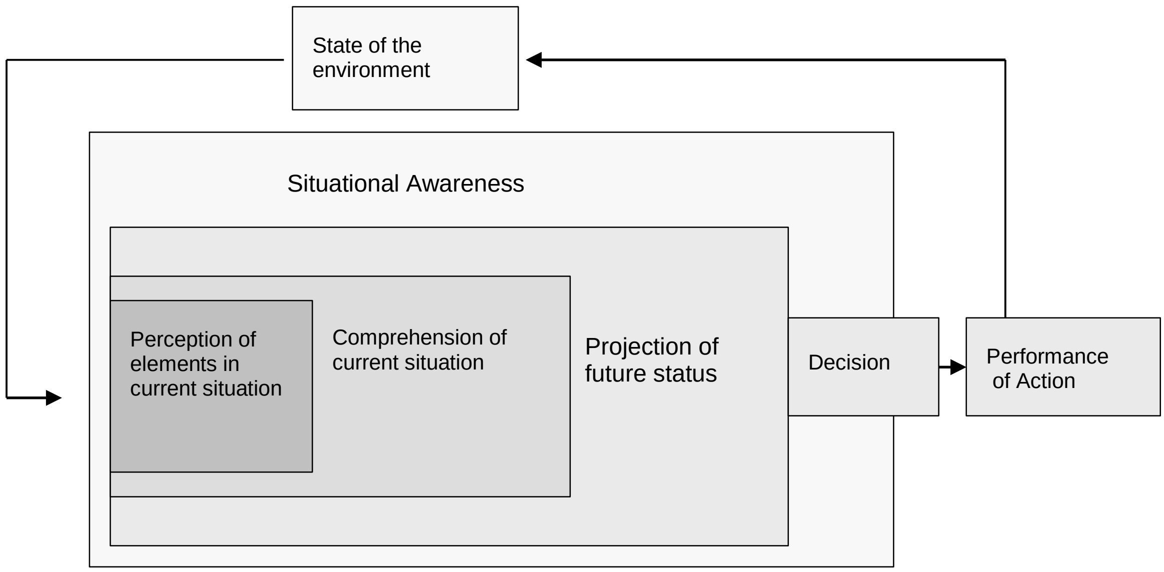


Figure 3. Illustration based on Endsley's model of Situation
(adapted from Endsley 1995)

1.2.5 Mental Models

In discussing the cognitive processes such as mental workload, SA, levels of sensory perception and information processing, it is important to understand the mental constructs that we create that allow us to perform different tasks. Mental models are internal representations that mirror aspects of the external environment that we live in and are used daily to interact with it (Craik 1943) Mental models are unique to each individual and are used to reason, anticipate and make decisions. By making mental models new information can be filtered and stored and when required retrieved at a later date (Johnson-Laird 2013, Underwood *et al.* 2002).

In the 1930s Charles Sanders Peirce, an American logician linked reasoning to ‘moving pictures of thought’ (Peirce, 1932) and Scottish psychologist and physiologist Kenneth Craik suggested that individual minds may carry a ‘small-scale’ model of how the external environment functions and these models could help to reason, prepare and anticipate better for present and future events (Craik, 1943). In the 1980s psychologist Johnson-Laird developed Craik’s original idea of a mental model to his own research area of human reasoning. Johnson-Laird (1983) proposed that the ability to reason was not due to mental logic that prescribed formal rules of inference but to the ability of humans to create a representation of the world from the meaning and knowledge that they have for the world.

The concept of mental models is highly relevant in HMI system designs, as first-time users to a new system construct mental models from prior background knowledge and previous experiences they have had with other HMI systems. By retrieving a mental model, users can navigate their way through a new system with a degree of familiarity which provides the user with ease of operation and a confidence with the system. Mental models are constantly evolving as the user interacts with the system. It is important for designers and engineers to have an understanding of their users’ mental models to enable improvements to the made to the interface design. In addition to mental models, which reside in the minds of individuals, fuzzy cognitive maps are physical constructs and extensions of cognitive maps. Designers and engineers are using these latest techniques to gain insightful predications of how their users interact with the system (Gray, Zanre and Gray 2014). In terms of usability and user experience, the more simplistic an individual’s mental model is of a task then the greater their performance will increase and produce a greater user experience with the system (Norman 2004).

1.2.6 Processing of Pictures and Words

In Studies One and Two the application of emotional words (Study One and Two) and pictures (Study Two) were presented to the participants, to assess the physiological impact that they evoked. Research and studies propose that the way in which individuals process words and pictures are different, and therefore the way in which an individual recognises and recalls these items is also different (Paivio 2007). It has further been demonstrated that when emotional stimuli are processed, results differ from neutral emotional stimuli (Stenberg 2006). With this in mind, it was important to review the relevant contributors to gain a fuller understanding of the way in which words and imagery are processed, as in a practical context relating to the automotive industry, the method of presenting information through the in-vehicle HMI system is a topical and relevant debate.

Cognitive processing according to Paivio's Dual Coding Theory (DCT) (Paivio 2007), involves two separate subsystems. One subsystem processes incoming verbal information whilst the second subsystem is specialised in dealing with picture representations (Paivio 2007). Both these systems are composed of internals units termed logogens and imagens that are activated with incoming information. These internal units can function either independently or together.

Paivio (2007) believes that cognition is an interplay of these two systems, and in some situations the verbal system will dominate, whilst in other situations the non-verbal or imagery system will dominate. This theory is different from other processing coding theories as it is modality specific. Initial DCT research focused on the improvements to memory by using imagery. and proposed that recognition and recall was enhanced by the interplay of the two systems and weakened if only one channel is used to process the information (Clark and Paivio 1991). DCT research has also focused on emotions and demonstrated that people would select pictures faster than they would select words. This picture superiority over words has been reported in numerous studies (Amrhein, McDaniel and Waddill 2002, Stenberg 2006).

Another important contributor to cognition processing is Baddeley, who's Working Memory Theory (WMT) is still being developed 40 years after conception (Baddeley 2012). As with the DCT, Baddeley's WMT or framework comprises of many different compartments processing different types of information. As research progresses, Baddeley is confident that more detailed models of the different components of the WMT will be developed and practical applications of the WMT outside the laboratory will occur (Baddeley 2012).

1.2.7 Assessment of Driver Interactions

The potential for driving errors and accidents during a journey are large due to the number of different elements that contribute to the driving task, these include frustration, complexity and workload (Walker, Stanton and Salmon 2015). Researchers have traditionally focused on performance, subjective and physiological measurements to assess the cognitive performance during driver studies (O'Donnell and Eggemeier 1986).

Performance-based data can be recorded and analysed to assess driver error on a wide range of parameters, either in a 'real' environment or a simulated environment. These measurements are usually taken from three main areas; pedal responses, driver steering responses and lateral control measures (Ostlund *et al*. 2005). Within these three domains there can be over 50 measurements that can recorded and analysed (Green 2013). With this large number of possible measurements, there has been difficulty in producing consistent data across research papers as detailed information on the measurements have not been clearly defined and terminology relating to the different measurements has been inconsistent. With this lack of clarity, comparing driving studies can often be difficult (Green 2013). In addition to this, the actual recording of performance measures can be difficult to set-up and achieve, as in the 'real' environment there is a risk factor to other road users, and in a simulated environment a driver may respond differently from a 'real' environment. Both of these issues can have an effect on the data that is being produced (Green 2013).

Subjective assessment tools are often used to assess a driver's workload, this is important in automotive HMI design as there are many different features in a car that can increase the driver's workload (e.g. communication and entertainment). These distractors can lead to the primary task of driving being affected (Schneegass et al. 2013). Subjective assessments are also used in usability studies to gain views and opinions of drivers to in-vehicle interface systems and the interactions that they have with the system. Other subjective assessments may be used to gain information on driver preferences to interior layout and materials.

Questionnaires are the most popular method of assessment. To assess driver workload, the NASA - Task Load Index (TLX) is popular technique and was designed as a multidimensional construct based on ratings on six subscales (mental demand, physical demand, temporal demand, performance, effort and frustration) (Hart and Staveland 1988). Although originally designed for use in aviation it has been applied to a wide range of studies, including automotive drivers. Pauzie (2008) adapted the NASA-TLX specifically for the automotive domain and named it The Driver Activity Load Index. Although subjective questionnaires can provide rich information, often in the driving context they are completed at the end of the task and therefore have a temporary delay, which may reflect perceived workload rather than actual workload (Schneegass *et al.* 2013). A driver distraction study, using secondary tasks was conducted using a post-drive questionnaire, and found that there were large discrepancies between the drivers' understanding of the secondary tasks. The researchers also commented that although the questionnaire was handed out immediately after the driving task, many accounts were reconstructions rather than recollections of the secondary task interactions (Petzoldt and Utesch 2016).

Physiological workload measurements are based on the assumption that physiological responses will vary as a function of workload (Embrey et al. 2006) and have been used to assess driver cognitive performance in a range of driving studies. Although there is no 'gold standard' for physiological measurements, in automotive studies, traditionally workload studies have commonly used heart rate and skin conductance activity (see Chapter 2).

1.2.8 Relationship of Arousal and Performance

As we have seen, there are a number of cognitive elements that are important in considering human interaction with technology. The fundamental law that links performance to arousal is the Yerkes-Dodson law, which states that the quality of performance on any task is an inverted U-shaped function of arousal, and that the range over which performance improves with increasing arousal varies with task complexity (Yerkes and Dodson, 1908). This law was derived from studies on animal discrimination learning. Yerkes and Dodson experimented on mice where they would deliver a shock to the mice which facilitated the learning of a brightness discrimination. By increasing the intensity of the shock the mice were able to discriminate better up to a certain point. After that point if the intensity of the shock increased, learning decreased. This would suggest that there was a strong connection between physical and psychological effects when confronted with an external stimulus.

In driving studies, the link between arousal levels, performance and driving error can easily be seen. If a driver is interacting with too many different devices, and at the same time driving fast in difficult road conditions (eg.fog), the driver may be subject to increased stress with an associated high level of arousal. This accumulation of effects can trigger the selectivity of attention and result in a ‘tunnelling’ phenomenon where the driver may miss an external hazard that may result in an incident (Memon and Young 1997).

Coughlin, Reimer and Mehler (2009) adapted the Yerkes-Dodson Law to fit a driving concept where both curves of the inverted ‘U’ were areas for concern. These areas included cognitive underload and fatigue as well as overload and high distraction. A driver who is under aroused, through tiredness, or over aroused may produce driving errors that could have serious consequences. The above mentioned researchers were developing a vehicle safety system which would alert the driver to their arousal state. The driver would then have the opportunity to adjust their behaviour, (e.g. by taking a rest break, if tired).

The act of driving in today’s cities and towns can be a stressful environment, with road congestion and road works being unavoidable at certain times of the day. In addition to these normal driving stressors, additional stress can be placed on a driver for a number of reasons (Walker, Stanton and Salmon 2015).   The driver may enter the vehicle already having had several stress events in the home environment to contend with. As they enter their vehicle and begin the journey to work they may discover that several of the in-car devices are not working in the manner in which they expect; thus feedback to the driver may be insufficient and result in driver frustration. This frustration can lead to inappropriate decisions being made and may increase the risk of having an accident. It is very important that a driver’s internal emotional states are considered during vehicle design as they have a major impact on the decisions drivers make which can affect other individuals or the impact that stress can have on their own physical well-being. This relates to both safety events whilst driving and consumer decisions that could impact the livelihood of the automotive manufacturer.

1.2.9 Emotion and Human-Computer Interaction

Before the revolution in personal computers, they were recognised more as tools to perform specific processing tasks and were very much in the domain of the computer specialist, whom was perhaps viewed quizzically by the rest of the workforce. Advances in computer technology have facilitated the important role that computers now play in communication throughout the world. In our daily lives we now rely heavily on interactions between user and computer systems, whether it is in work or social sectors. This interaction has become a key area of research.

Researchers developing improvements in human-computer interaction (HCI) are increasingly aware that the way in which machines can recognise and respond to humans is a necessary component in adaptive computer systems (Picard, Vyzas and Healey 2001). Terms that describe this interaction are affective computing and machine emotional intelligence (Picard, Vyzas and Healey 2001, Picard 2003). These two factors are thought to be increasingly important in adaptive computer systems where user affective feedback can improve the overall user experience between human-computer interactions (Yang, Ji, Chen and Fu 2014). Recognising that emotional states have an influencing factor in the way people think, behave and respond has renewed interest in this area.

Emotional or affective research has notable challenges mainly due to a lack of consensus between scientific theorists regarding what emotions are, who can experience them and even the methods employed to study emotions (Gross and Feldman-Barrett 2011).

In computer disciplines the definition introduced by Picard in the 1990s has been widely accepted and is defined as '*measuring observations of motor system behaviour that correspond with high probability to an underlying emotion or combination of emotions'.* Picard also suggested that in the context of computer science the terms *'emotional state', 'affective state' and 'sentic state'* could be used interchangeably (Picard, Vyzas and Healey 2001).

Kleinginna and Kleinginna (1981) tried to collate definitions of emotion and their findings illustrated the problem as over ninety definitions were recorded. However, there was general agreement that emotions are a group of psychological states that include subjective experience, expressive behaviour and physiological responses (Gross and Feldman-Barrett 2011, LeDoux and Damasio 2013). Many experts in affective research agree that emotions have a limited set of components and characteristics (Izard 2006). Emotions are complex, multifaceted and have an infrastructure that includes neural systems allocated to emotional processes which stimulate cognition and action and initiate response systems (Birbaumer and Ohman 1993, Oatley and Jenkins 1996).

Emotional research generally follows two paths, the first focussing on the central nervous system's involvement in emotions with the second highlighting the importance of the autonomic nervous system's role. When discussing the first path, it is the anatomical function of the human brain that has been of key interest and trying to determine which parts of this complex structure are responsible for different emotions. To gain knowledge and obtain a greater understanding of the interplay of the CNS a variety of studies have been undertaken. By using neuroimaging, activation of certain locations of the brain can be seen during an emotional event. The ANS displays a complex interplay of the sympathetic nervous system and the parasympathetic nervous system when an emotional response is elicited. A review of the CNS and ANS by Hagemann, Waldstein and Thayer (2003) stated that only a few studies focus on both of these systems. The authors suggest that by having a more integrated approach, further knowledge and understanding of the interrelations between the two systems would be achieved.

The link between the autonomic nervous system and emotion was discussed as early as 1915 by physiologist Walter Cannon, (Brown and Fee 2002) and although there have been numerous publications on the autonomic nervous system and emotion there is still scientific debate surrounding this relationship (Kreibig 2010). A review carried out by Mauss and Robinson (2009) on the measures of emotion, discusses the difficulties with body systems in recognising emotions as well as issues relating to measuring emotional responses. The different systems that can recognise emotions are grouped into four main categories.

1. Body behaviour and facial expressions
2. Vocal characteristics
3. Physiological biosignals
4. Subjective experience (self-report)

The review conducted in 2010 by Kreibig examined 134 publications on experimental studies of emotional effects on the autonomic nervous system and although there were many different variables considered, the findings from the review suggest that there is a considerable autonomic nervous system response to different emotional states (Kreibig 2010). Of the cardiovascular measures that were reviewed heart rate was the most frequently applied measure with heart rate variability being the fourth most commonly applied metric in emotional research.

Difficulties in describing personal emotions and assessing these emotions appear to have stemmed from a lack of differentiating between them, as they are often seen as connected experiences rather than separate entities (Russell and Fehr 1994). Osgood's (1952) seminal work in semantic differential, where the variance in emotional assessments is attributed to three dimensions (Valence, Arousal and Dominance), has given guidance for studying emotions on dimensional models (Mehrabian and Russell 1974). The most recognised model is a two-dimensional representation as displayed in Figure 3. In this model the horizontal axis represents the valence dimension and the vertical axis represents the arousal dimension. Different researchers have often attached different labels to describe these two dimensions, such as the positive and negative affect (Tellegen, Watson and Clark*.* 1999), the tension and energy dimensions (Thayer, 1989) and the approach and withdrawal dimensions (Lang, Bradley and Cuthbert 1998).

Lang and Bradley (1997) also discussed a third dimension, the dominance dimension and created a three dimensional model, where valence represents the positive or negative feeling of an emotion and arousal is the strength or intensity of the emotion (excitement / calmness). Although dominance is not classified as a primary dimension it represents the control / uncontrolled element of an emotion.

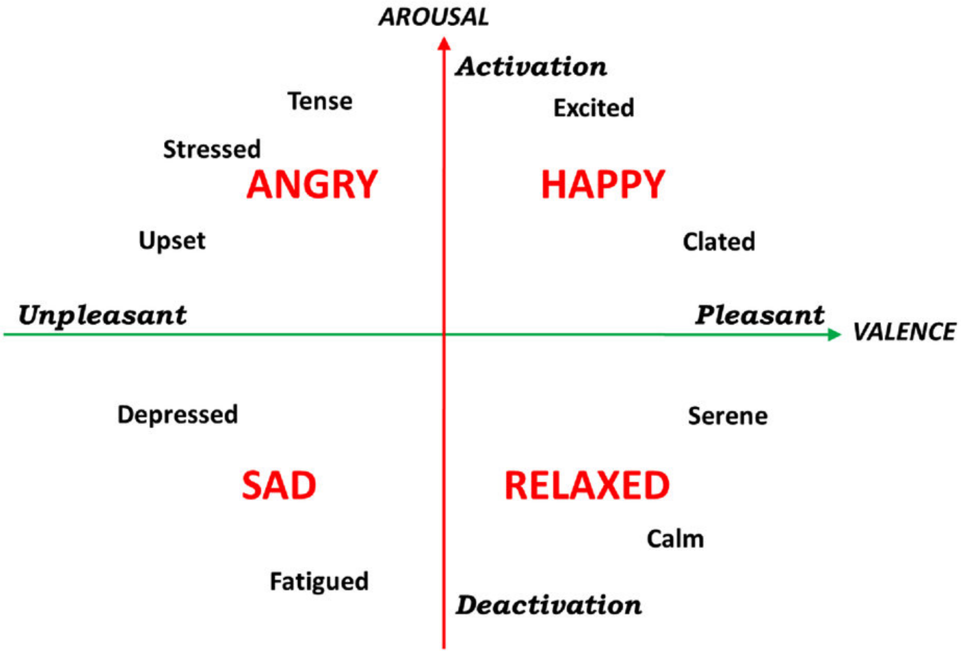


Figure 4. Representation of the circumplex model of affect (Valenza *et al.* 2014)

To assist in emotion and psychology research Lang, Bradley and Cuthbert (2008) developed a range of emotional material that can be accessed and used as stimuli sets to promote a level of standardisation within the research community. These stimuli consist of images (International Affective Picture System (IAPS), sounds, (International Affective Digital Sounds (IADS) (Bradley and Lang 1999), words (Affective Norms for English Words (ANEW) (Bradley and Lang 1999) and text (Affective Norms for English Text (ANET) (Bradley and Lang 2007).

Individuals experience emotions every day, and they are crucial in day to day decision making. This was proposed by Damasio (1994) during his work with individuals that had suffered brain injuries, specifically prefrontal cortex damage. On recovery these individuals were unable to make rational decisions and had also lost the ability to process emotions. Emotion and cognition are closely linked, cognition interprets and enables understanding while emotion usually occurs first and allows for quick reactions and decisions to be made.

In a driving context this can be important, if the driver is not in a positive and calm state, the quick decisions he or she makes could have serious consequences. One area that can dictate positive / negative emotional state is the interaction the driver has with the in-car interface system. If the driver is struggling to interact with the interface and becomes increasingly frustrated, this will prevent them concentrating and learning the different steps that are required to interact successfully with the interface. This driver may completely disengage with the interface, which could be dangerous as the driver would not be receiving any feedback from the vehicle (Norman 2013). Designers of in-car HMI systems must recognise the importance that emotions have on their customers and produce HMI systems that will provoke positive emotions and by doing so increase the user experience and likelihood of a positive driving outcome when they interact with their vehicle.

1.3 User Experience and Usability

As we begin to appreciate the cognitive and physiological bases of what underlies human performance in relation to driving and interacting with automotive systems, it is important to understand the nature of interaction between the user and the system. The term ‘user experience’ (UX) has become synonymous with designers and engineers who produce products or systems for their customers, the automotive industry is an excellent example, where the attention now focusses on the overall hedonic experience that the driver has with their vehicle rather than just one component of it. It is not only the designers and engineers that are utilising this terminology, as during an advertising campaign for a certain brand of car, it is often the perceived experience that the owner will have with this particular brand of car that is the focus of the campaign. (see Norman 2013).

Early UX research is in agreement with Norman's views stating that too often usability research was directed on task efficiency and goal achievements (Hassenzahl and Tractinsky 2006). Norman believes that 'UX' combines all the different components of an individual's experience with a system or product. This concept of UX has generally been accepted and recent papers now propose that the UX starts sometime before the actual interaction with the product or system and continues after the user finishes with the product or system (Karapanos *et al.* 2019, Roth, Vorderer and Klimt 2009). The many different features that contribute to the user's experience before and after the interaction with the product include advertising, prior experience with similar or related products and systems.

Although user experience found favour with the Human-Computer Interaction (HCI) community it was quickly criticised as being a transient and ambiguous term (Forlizzi and Battarbee 2004). Perhaps this all-encompassing phrase has encumbered 'UX' research as papers frequently comment on the difficulty in producing a single concise definition for 'UX' (Law *et al.* 2009).

The definition most frequently quoted is from the International Organisation for Standardisation (ISO) which defines user experience as 'a person's perceptions and responses resulting from the use and / or anticipated use of a product, system or service' (ISO 9241-210:2010). Nielsen and Norman (2016) emphasise the importance of including all aspects of the end-user's interaction with the company, its services and its products. They follow on by saying that it is also important to separate the overall user experience from that of the user interface. Although the user interface is a crucial part of the design, to achieve a high-quality user experience for the customer there must be a successful multidisciplinary collaboration between industry, engineer, interface and graphic designers. For companies that design products such as smart phones, video games and internet sites it is crucial that their users' UX is positive and desirable as their continued success in the market depends on the loyalty of their customers.

Similarities in attributes may also be drawn between the aviation and automotive industries, as they both involve the pilot or driver interacting with multiple systems simultaneously and these complex interactions can have a large impact on the users overall UX. Both of these industries have actively been involved in UX research and although in previous years the focus was a more task and goal oriented view, more recent UX research examines and recognises the importance of the hedonic experience that users have with interactive products and systems. This hedonic experience is recognised as an important consideration for the longevity of using the system or product (Bargas-Avilla and Hornbaek 2011).

As discussed above, the car driver has many different interfaces that they have to interact with, and it will be the overall experience with the wider system that will determine whether the user has had a positive or negative experience. All designers wish to produce features of products that give the user a positive experience but unfortunately often the designers intended experiences are not replicated by the user and the actual experiences that the individuals have are very different from the intended experience of the designer. This is especially true of industries that have manufacturing constraints on design elements. For example, a designer within an automotive manufacturer may present an elaborate (and pleasing) visualisation to the engineer, to be told that the exact nature of the design cannot be achieved using the materials at their disposal, or the layout available based on other system requirements. Thus an automotive designer may find their initial design is slowly changed through manufacturing iteration, rather than user friendly iterative design. This could then lead to a poor user experience, which could present itself as user frustration and disengagement with the device or feature that the individual is interacting with (Norman 2013). Interface aesthetics are strongly linked with usability and in the automotive industry this is an area where creative designers are able to establish differences between different car manufacturers. By increasing the aesthetic quality and appeal to drivers, their acceptance and engagement with the in-vehicle interface may increase.

A user experience is subjective and varies between individuals. The differences between how individuals interact with the same system is due to many complex features including the experiences that they bring to the system and the different mental models that they have already constructed (Johnson-Laird 1983, Gentner and Stevens 1983). The UX will also change over time as the user spends more time with the system or product and gains information, experience and confidence with the product (Just and Carpenter 1985).

Perhaps this area is more challenging when we begin to consider the imminent introduction of autonomous cars now being seen as a valid progression in the automotive industry. The HMI's role will become crucial to the driver as they interact with the automotive system (Richards and Stedmon 2016). In this instance it is even more important that the driver feels, not only an integral part of the system, but will need to possess a high level of acceptance and trust in using this technology. Over past years the automotive manufacturer has been more focussed on the UX of the HMI in terms of infotainment systems of potential customers. The interaction the driver has with the infotainment system must result in a positive experience every time they use the system.

For the next generation of in-vehicle interfaces, designers and engineers may be presented with many innovative new features that will need to be incorporated into the system. However, this will only be successfully achieved if drivers believe that the interface they are interacting with provides them with a pleasurable and positive experience whilst simultaneously delivering them novel and innovative features that improve their overall holistic experience as they spend time on the road.

1.4 The Automotive User Interface

Vehicle owners have seen a plethora of technology, being introduced into vehicles over many years, designed to assist and enhance the journey that drivers undertake each day. Without question, the technology available for in-vehicle HMI systems has changed considerably over the last decade (Smith, Vardhan and Malet 2014).

With the use of mobile devices and touchscreen technology, being a part of daily life, individuals entering their vehicles desire these technologies, and the capabilities they offer, to be available within their cars. Although the technology is available to produce these devices in vehicles, several problems arise due to the differences between the driving environment and the stationary environment. Any device that requires interaction with the driver, can affect the driver's focus and attention, with serious risk consequences. With safety being the key consideration in automotive design and engineering, it is perhaps not surprising that the industry has taken a 'following' approach with in-vehicle HMI technology, rather than a 'lead' approach (Ustwo Auto 2016).

The future vision of a connected car will influence designers and engineers in their HMI concepts. As the world's forefront technology leaders, Apple, Google and Microsoft enter the car market, producing products such as Apple's CarPlay (2014) and Google's android platform collaborators, (Open Automotive Alliance (OAA) 2015) the next generation of in-vehicle HMI systems may see a dynamic shift in the design of in-vehicle HMI systems.

Figure 5, represents the past and present vehicle features, which as illustrated have not significantly changed throughout the decades.

Figure 5. HMI Past and Present (Ustwo Auto 2014)

With the unveiling of the new Tesla Model 3, on the 31st March 2016, the vehicle's interface was a large touchscreen graphical user interface (GUI). Although similar in concept to the Model S large 17" GUI, there were some differences. In the Model 3 the GUI was slightly smaller (15") and positioned on top of the central column rather than being embedded into the car's interior in the Model S. Another difference was that the GUI of the model 3 was presented in landscape rather than portrait (Figure 6). It will be interesting to see if any changes will be made to the interface before the car is in production in 2017.

Figure 6. Tesla Model 3 and Models S Interface (The Verge 2016)

Large touchscreens have gained preference from some car manufacturers and were originally adapted from the aeroplane's cockpit. Although the large screen has corrected some of the problems that arose with a small screen, such as driver difficulty in locating and selecting icons, which contributed to driver distraction from the primary task, the GUI still produces driver distraction, as all the feedback from the interface is visual (Burnett *et al.* 2013) (see chapter 1.2 for additional information on visual-haptic modality).

Voice command and recognition systems also have a role in HMIs, as they should provide a safety advantage over graphical touchscreens due to a reduced competition for visual processing between the primary task of driving and the secondary task of interacting with the interface (Owens, McLaughlin and Sudweeks 2011). However, with the nuances of language and different cultures, it may prove difficult to implement a single blueprint for every driver (Reagan and Kidd 2013), (see chapter 1.2 for additional information on SDS). Another problem for drivers using a voice recognition system, was the distraction from the primary task of driving when listening and responding to the voice command. Drivers were inclined to look towards the interface rather than keeping their eyes focussed on the road ahead (Reimer *et al.* 2013).

Not all car manufacturers have moved away from implementing physical dials and switches. A central rotary control is seen in the Mercedes, where the control knob directs the graphical screen (Figure 7).

Figure 7. Mercedes AMG GTS console controller
(McKenzie 2015)

Earlier problems arose with this type of control knob, as seen with the BMW i-Drive controller (Figure 8). The difficulty with this control, was that the driver was using a rotatory device to operate a linear function on the graphical screen. This added complexity to the driver, increased their cognitive workload every time an action was required (Smith, Vardhan and Malet, 2014).

This material has been removed from this thesis due to Third Party Copyright. The unabridged version of the thesis can be viewed at the Lanchester Library, Coventry University.

Figure 8. BMW i-drive controller
(Autocar 2016)

On the Jaguar F-TYPE interface there is an 8-inch touch screen controlling navigation and infotainment features. This type of interface also interacts with the driver's smartphone, and with Jaguar InControl technology the driver can tailor specific apps to provide information on local areas of interest on route (Jaguar 2014), (Figure 9).

This material has been removed from this thesis due to Third Party Copyright. The unabridged version of the thesis can be viewed at the Lanchester Library, Coventry University.

Figure 9. Jaguar F Type Interface

(Furness 2015)

Research into swipe and gesturing has seen prototypes that appear to be promising and practical in the car environment (see chapter 1.2 for additional information on gesturing modality). A designer from Apple (Kern 2014) has produced a gesturing prototype design using a tablet, which utilises the whole screen as the control centre rather than separate areas for different functions. By touching the screen with different numbers of fingers and lengthening and shortening the proximity of these fingers, different control features could be triggered (Kern 2014). The advantage of this would be, that placement of fingers anywhere on the touchscreen would engage with the interface, which in a theoretical concept should enable drivers to keep their eyes on the road. Perhaps a caveat to this technology would be that the driver would not have an opportunity over time, to build a mental model of the different operations that could be performed.

Technologists from the gaming research industry are now seeing viable possibilities to integrate their innovative concepts to the in-car HMI systems. Two such companies include Leap Motion technology (2014), which uses touch free natural gestures to engage with an in-vehicle infotainment system, and AIREAL technology, which delivers a novel haptic sensation in free air, using air vortices. Although AIREAL technology is still in the developmental process, researchers believe that the potential for wider applications and adaptions of this technology are promising (Sodhi *et al.* 2013). Another area of development that has transferred from the aviation industry is Heads-up displays (HUDs), which are designed to help keep the driver's eyes focussed on the road ahead rather than glancing in the direction of the interface.

In 2013, Cisco Systems ran a customer experience report focussing on the buying and driving experience. The report gathered information from 1,500 customers from 10 different countries. Feedback from the customers indicated that they required a more personal experience when driving their car, while at the same time indicating that they would like to see more automation brought into the car (Cisco Customer Experience Research, Automotive Industry 2013).

With autonomous vehicles ‘knocking at the door’, the user experience has to be intuitive, dynamic and feature a range of modalities that are personalised to the driver. The potential for drivers to have sections of their journey as passengers, will open up many other opportunities for technologists to create engaging applications that drivers will enjoy. One of these new areas is the possibility of real time psychophysiological sensing equipment that could be embedded into the car furniture and relay information to the driver relating to their physiological states. This prospect could have important consequences when autonomous driving has been introduced, as these measures could provide an important marker for the ‘hand-over’ phase from autonomous to manual driving. Other interesting and novel applications of psychophysiological data would also be an attractive feature for an in-vehicle HMI.

We can see that the design and use of automotive HMIs, whilst occupying an important aspect of differentiating the manufacturer’s competitive edge, poses a challenging design issue. The act of driving is predominantly a visual task that requires focussed attention of the driver, who at times will need to interact with in-vehicle systems. The automotive environment that the driver finds themselves in, may now be considered as a contested design battlefield, whereby manufacturers seek to gain the upper hand by the HMI that exists between the driver and the system, and assessing this in terms of human performance and UX is a critical determinant of whether one HMI solution is better than another.

We have seen how complex elements can be important determinants of human performance with HMIs, and understanding of cognitive processes allows us to design and assess different HMIs. These methods can range from identifying areas of increased mental workload, enhancing SA, or ensuring that the driver has spare cognitive capacity to ensure they are not distracted from their primary visual task. Many ways in which cognitive performance may be assessed can involve a high degree of subjective measurements. These metrics, due to their subjective nature, do not always possess high validity or paint a true picture of human performance. Therefore, we turn our sights towards more objective measurement that can assist in determining human performance, especially in relation to investigating affective states that underpin UX.

2 Chapter Two

2.1 The importance of considering the driver's physiological state

Human physiology is a branch of biology that studies how the human body functions; this can be from a single cell to a whole body system (Widmaier, Raff and Strang 2006). To gain an insight into how the human body functions methods, from simple visual observations to complex neurological imaging assessments, are employed. Each method when appropriately and correctly applied will provide information regarding the state of the human body (Martini 2006).

Benefactors of physiological assessments have traditionally been physicians and scientists, where the information from the physiological assessments have aided in diagnosis, treatments and predications of prognosis (Fox 2011). Until recently medical equipment employed to record and measure physiological data was often large, immobile, costly and not user friendly. Other issues occurred relating to the actual procedure in obtaining the physiological data, as it was often invasive and frequently caused pain and distress to the individual (Goldstein 2010).

In the 1980s there was a surge in public participation in sporting activities as part of an active lifestyle with several professional bodies examining physiological data to gain additional information relating to their specific sport (Nickholson, Hoye and Houlihan 2010). Elite athletes such as runners and cyclists were frequently monitored and their physiological data was analysed to plan training sessions and plot their progress. Today, the professional international cycling teams' physiological metrics are vital in determining the cyclists form. It has been quoted many times by the top professional cyclists that 'It's all about the numbers' (Wiggins 2013). David Brailsford, the former sports director of the Great Britain Cycling team and current director of the Sky professional cycling team is renowned for optimising performances around physiological and psychological data by using the latest technology to gain a clearer and more informed picture of his cyclists physiological and mental states.

Technology engineers and designers recognised the potential of physiological monitoring equipment used in the sports industry as by the 1980s there was a broad range of mobile devices that were able to assess and produce reliable physiological data on a wide range of parameters, at a reasonable cost for consumers. The sport and leisure industry has embraced this new developing technology and non-invasive mobile devices are now able to give detailed physiological data on biological states. With access to the internet and the public's insatiable interest in health and fitness programmes participants now have the potential to become knowledgeable on their own physiology. Telecommunication technology (smart phones) has enabled users, supported by specific applications, to easily conduct physiological measurements. Non-invasive mobile equipment now offer coaches, athletes and sport and fitness enthusiasts the opportunity to train, using physiological data and when used correctly these devices can provide accurate data that can help to improve training sessions and ultimately the overall performance of the individual.

With the availability of these non-invasive mobile devices, other industries are now exploring the use of physiological metrics and how they can be usefully applied in their sector (Lenneman and Becks 2010, Mehler *et al.* 2009, Mehler, Reimer and Coughlin 2010). The automotive industry is one such example that recognises the application of physiological metrics as a potential source of valuable information when examining the different interactions that occur between the driver and the vehicle. One area in which significant research is being undertaken is assessing the interactions that occur when drivers are interacting with their in-vehicle infotainment systems (Reyes *et al.* 2009). Although the use of physiological data is in its relative infancy in commercial automotive engineering and design, the aviation industry, in particular the defence sector, have been utilising this data for many years and were among the first to use physiological data to measure the effects of workload on pilots and air traffic controllers (Hilburn *et al.* 1997, Wilson 2002). Studies indicate that by using physiological data, real-time events can be studied and used to give quantifiable results (Bonner and Wilson 2001).

The computing industry is another domain that has used physiological metrics. In this industry physiological measurements have gained popularity when assessing affective or emotional states. By using physiological measures, a greater understanding of emotional states, and how they impact on the decisions that people make, can be achieved (Vyzas and Picard 1999, Tosa and Nakatsu 1996).

The relationship between affective and physiology states, provides an opportunity for designers of human machine interface systems to obtain a clearer insight of their potential users (Picard, Vyzas and Healey 2001). This area of research is not without its difficulties, as determining ‘real time’ affective states of participants can be extremely complex. Two of the human physiological body systems’ that can be assessed and offer perhaps the greatest significance to this research, are the cardiovascular and the integumentary systems. Although both of these systems are highly complex and integral to many biological processes, it is their relationship with another physiological body system, namely the autonomic nervous system (ANS) and its rapid responsiveness in order to maintain homeostasis that is of particular interest to this research. With technological advancements cardiovascular and electrodermal activity metrics offer a clear and central contribution for use in physiological assessments in a wide range of applications.

2.1.1 Autonomic Nervous System

The autonomic nervous system (ANS) is primarily responsible for innervating smooth and cardiac muscle, in addition to glandular tissue (Martini 2006). The term ‘autonomic’ derives from two Greek words meaning ‘self-governing’ and ‘independent’ and therefore this system generally operates subconsciously (Cunningham and Klein 2007). As well as regulating smooth and cardiac muscle the ANS also assists in maintaining a constant body temperature, controls eating, drinking and sexual behaviour (Cunningham and Klein 2007). Although the ANS is regarded as involuntary, the behaviours controlled by it are closely linked with voluntary movements which are controlled by the somatic motor system, which innervates skeletal muscle and is responsible for body movement (Martini 2006). For example, the process of running is a voluntary action but the metabolic requirements and thermoregulatory events associated with running are automatic and due to physiological changes in the body brought about by the ANS (Scherer 2001). Such changes would be seen in the cardiorespiratory system and include alterations in ventilation, blood flow and cardiac output (Ernst 2014). Similarly, emotional behaviours have automatic characteristics that may be reflected in physiological changes. An example of this would be an individual delivering a speech to an audience for the first time. For many, this experience would produce physiological changes such as increased heart rate and sweating palms (Martini 2006).

The ANS is a regulatory system and is essential for maintaining body homeostasis (Cunningham and Klein 2007). The concept of homeostasis was first introduced by Walter B. Cannon in 1932 (Goldstein and Kopin 2007) who had developed the idea from Claude Bernard's *'le milieu interior',* meaning the internal environment (Horn and Swanson 2013). Bernard had observed that during a wide range of behavioural states and external conditions a human's internal environment remained relatively stable. This *'le milieu interior'* included body temperature, blood pressure and the extracellular fluid compartment (Horn and Swanson 2013). Cannon used the word homeostasis to describe the complex mechanisms that maintain temperature, blood pressure, body fluids and other physiological variables within a small physiological range (Goldstein and Kopin 2007, Horn and Swanson 2013). These homeostatic mechanisms are also very adaptable to varying behavioural conditions. This can be explained when a healthy individual runs. As a person runs their cardiac output may increase threefold whereas their blood pressure will increase in a far smaller range (Martini 2006). In the absence of homeostatic mechanisms, it would be reasonable to expect that the person's blood pressure would increase in direct proportion to cardiac output. If this was the case the end result would be that blood vessels may rupture due to the uncontrolled rise in blood pressure (Horn and Swanson 2013).

As with any complex system there are many components involved that all have an important role in homeostasis. The somatic nervous system and the neuroendocrine system are two systems that are interlinked with the ANS to maintain homeostasis (Horn and Swanson 2013). The somatic nervous system controls skeletal muscle, and therefore body movement depends on this system, whereas the autonomic system has control over cardiac muscle, smooth tissue and glands, but both work together to produce behaviours such as running and climbing (Horn and Swanson 2013).

In the human body the ANS has an influencing factor over the cardiovascular, respiratory, gastrointestinal, urinary and reproductive systems (Martini 2006). With respect to the cardiovascular system the ANS regulates the heart rate and blood pressure. Although, the ANS comprises of three different subsystems, for this research only two will be discussed, namely the sympathetic nervous system (SNS) and the parasympathetic nervous system (PNS) (Furness 2006). Although, in both systems their anatomical structure, synaptic communication via neurotransmitters and final innervation of the target organs and tissues may be different there is a constant interplay between the two systems ensuring that the internal environment of the human body is kept within the narrow boundaries for optimal health and well-being of the individual (Horn and Swanson 2013).

Organs and tissues that are innervated by the SNS include the heart, vasculature, pupils, lungs, sweat and salivary glands, whereas the PNS innervation includes the heart's pacemaker cells and atria, smooth muscles and glands (Horn and Swanson 2013). In the majority of situations, the SNS and the PNS have opposing effects, where one causes stimulation and the other causes inhibition (Fox 2011). Although this can be used as a general statement, as with any highly complex structure there are always exceptions. There are instances where each system will work independently and other cases where some structures of the body are only innervated by one system and others when both branches will work synergistically together (Fox 2011). In the cardiovascular system there is a crossover between the PNS and SNS, and both branches are in a constant of state of tonic innervation, and primed for activity whenever environmental or situational demands are required to be acted upon (Horn and Swanson 2013). A diagrammatic image of PNS and SNS, displaying fibres, neurotransmitters, receptors and target tissues, is illustrated in Figure 10.

When a person is frightened, which could be for many different and individual reasons, they will experience physiological changes. Their heart will speed up, their respiration will become rapid and shallow, palms become sweaty, mouths become dry and their muscles tense (Martini 2006). At the same time that these body changes are occurring a frightened individual may also have a strong urge to run or escape from the situation that is causing these uncomfortable symptoms (Martini 2006). The physiological body changes that are associated with fear, as described in the scenario above, are due to the SNS (Fox 2011) with the ANS being regulated by a central neuronal network of which the hypothalamus is the key to that network.

This material has been removed from this thesis due to Third Party Copyright. The unabridged version of the thesis can be viewed at the Lanchester Library, Coventry University.

Figure 10. Diagrammatic image

displaying Parasympathetic and Sympathetic fibre, Neurotransmitters ACh = acetylcholine, N = noradrenaline, NE = norepinephrine, D = Dopamine, Receptors M = muscarinic, N = nicotinic, D_1 = dopaminergic and Target tissues (Pharmacology-Online 2011)

The sympathetic nervous system (SNS) is involved with the ‘fight or flight’ response as it increases the body’s general arousal state and primes it in readiness for potential emergency situations (Cunningham and Klein 2007). This response is regarded as a critical function for animals and humans, as it prepares them to either stay and fight, or flee from the dangerous threat (Cunningham and Klein 2007). As the scenario depicted, stimulation of the sympathetic system causes an increase in a range of physiological responses, including a heightened mental alertness (Widmaier, Raff and Strang, 2006).

The sympathetic pathways arise from the cervical and lumbosacral areas of the spinal cord and synapse in the ganglia of the sympathetic chains which lie on each side of the spinal cord (Horn and Swanson 2013). Although these preganglionic fibres are myelinated, their conduction is relatively slow. Sympathetic preganglionic neurons discharge the neurotransmitter Acetylcholine (Ach) and are termed cholinergic, whereas adrenaline (A) is the neurotransmitter of the postganglionic sympathetic neurons and they are referred to as adrenergic (Martini 2006). Although cholinergic and adrenergic fibres are dominant in the complex interactions of the heart, there are other transmitters that are released at the same time as Ach and adrenaline, and these are termed co-transmitters. One such transmitter is Neuropeptide Y (NPY) which acts on adrenergic receptors and has an influence on arterial contractions (Ernst 2014).

When comparing the SNS with the PNS, in terms of transmitters, early research indicated that adrenaline was diffused into local tissues and transported in the blood to different parts of the body (Ernst 2014). It has now been established that the sympathetic neurons have the ability to independently activate and control specific targets. The parasympathetic nervous system activity is very quick, almost instantaneous (1 millisecond), (Malpas 2010) whereas the sympathetic nervous system's activity takes a longer length of time. This difference in time is mainly due to the fact that the SNS does not discharge its neurotransmitter exactly onto the effector organ, instead it diffuses the neurotransmitter in the approximate area where it is then 'taken-up' by the effector organ (Malpas 2010). This process can take up to 30 seconds to reach a ‘peak’ level and will remain at this level for a longer duration than in parasympathetic activity. The parasympathetic neurotransmitters are destroyed very quickly, by the enzyme esterase, as soon as the mediators have been released (Martini 2006). Due to the diffusing nature of the sympathetic neurotransmitters, effects can persist for a greater duration than during PNS activity (Ernst 2014).

Although sympathetic responses can be essential in an emergency situation, it is also accepted that extreme sympathetic responses over a prolonged period of time can lead to the pathological condition of *post-traumatic stress disorder,* which can bring about life-changing behaviours and have a major impact on a person’s life (Horn and Swanson 2013, Hyman and Cohen 2013). Figure 11 illustrates the neural synapses and transmitters of the PNS and SNS.

The parasympathetic nervous system (PNS) is sometimes referred to as the ‘restorative’ or ‘rest and digest’ system as the general nature of this system, is to conserve energy and encourage sedentary demands such as digestion and absorption of food (Widmaier, Raff and Strang, 2006). The PNS also has influences in the gastrointestinal tract, urinary tract and reproductive organs (Fox 2011). Preganglionic neurons in the PNS use Ach as a neurotransmitter and Ach is also the neurotransmitter at the end-organ where it activates the muscarinic receptors on the target cell (Ernst 2014). The response of this stimulation can either be excitatory or inhibitory (Ernst 2014). As with the sympathetic nervous system, other transmitters and co-transmitters can be triggered (Horn and Swanson 2013).

Figure 11. Sympathetic and parasympathetic neural synapses

2.1.2 Electrodermal activity in arousal, workload and emotion research

The skin is a highly complex and dynamic structure that has many features which protect the human body from different external threats. The primary role of the skin is to act as a barrier, and it is the first line of defence for invading pathogens or environmental threats such as chemicals or temperature. The skin also provides incoming information on different sense functions, such as nociceptors that convey information on pain, mechanoreceptors for touch and temperature. The skin not only prevents external threats from entering the body but allows internal fluids to exit the body, which is essential for thermoregulation. One method of expelling heat from the body and maintaining thermoregulation is by evaporation when the eccrine glands produce sweat or perspiration. The production of sweat is tightly regulated by the sympathetic fibres which are controlled by the hypothalamus. By keeping the electrolyte balance and the stratum corneum or outer layer of the epidermis moist, fine tactile skills and elasticity of palms and feet are maintained (Sato *et al.* 1989).

The eccrine or sweat glands arise from the hypodermis or dermis sections of the skin and exit through small pores directly on the skin's surface. On a human's body there are two types of skin, which are identified as ridged and polygonal skin. It is the ridged skin that is of interest in this study and is only located on the palms and soles, including the flexor side of the fingers and toes. Ridged skin is hairless and contains no sebaceous glands, unlike polygonal skin type which covers the majority of the body.

Eccrine sweat glands are classified as exocrine glands as the fluid that is secreted flows directly onto the skin's surface (Dawson, Schell and Filion 2007). Sweat glands are widely distributed throughout the body with only a few specific regions where the glands are absent, the inner ear canal and lips being two such areas (Pinkus 1970). The number of sweat glands found on an adult's body can vary greatly, with a range of 1.6 – 4.0 million (Sato *et al.* 1989). The density of sweat glands also varies with the palmar and plantar regions of the hands and feet being areas of high density along with the forehead, whereas the rest of the arms and legs are described as areas of low density (Kuno 1956). It is estimated that each palm may contain between 600 and 700 eccrine sweat glands per cm^2. As the number of eccrine sweat glands are fixed at birth it is normally assumed that children will have a greater density of sweat glands than adults due to a smaller total body surface area than adults (Montagna and Parakkal 1974).

Eccrine sweat glands are comprised of two main parts; the first being the coiled secretory component that arises from the hypodermal or dermal layer of the skin and drains into the second part which is a long narrow duct that travels through the epidermis and opens via a small pore onto the surface of the skin (Ellis 1967) (see Figure 12). The composition of the fluid or sweat from the eccrine glands is water (99%), sodium chloride, potassium, bicarbonate, lactic anion, urea, amines and vitamins (Sato *et al.* 1989). Sweat discharges out onto the surface of the skin in a pulsatory motion rather than a continuous flow (Sato *et al.* 1989).

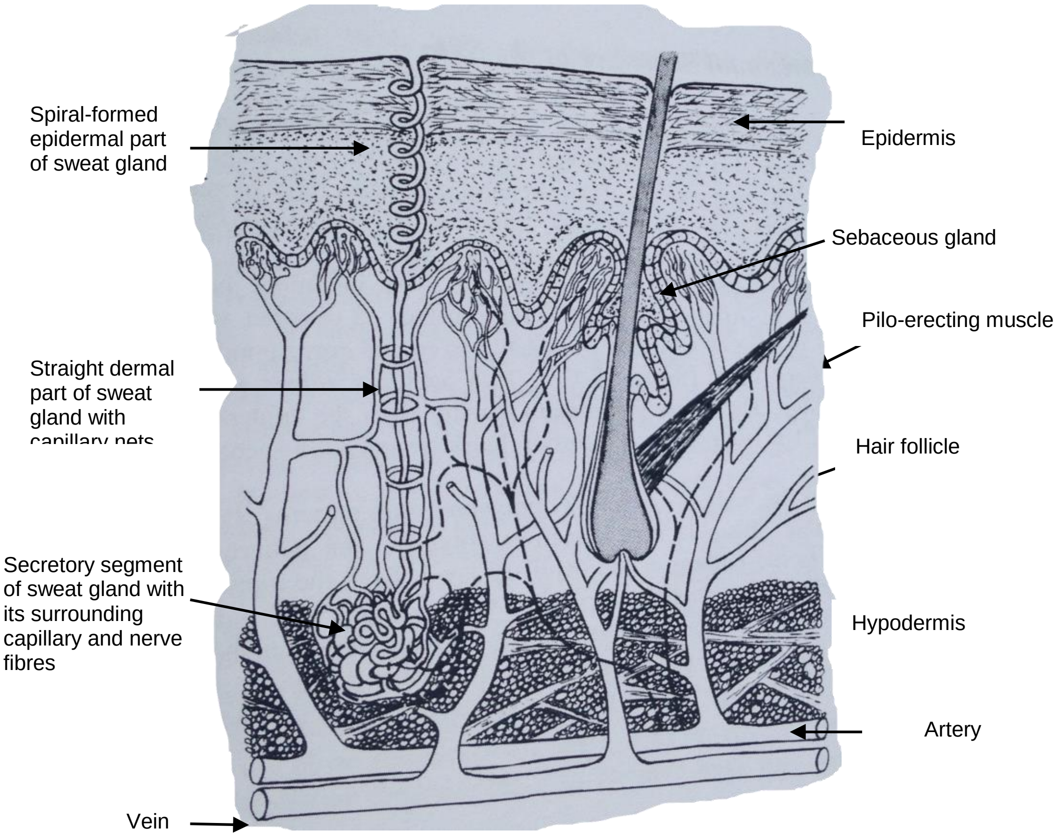


Figure 12. Schematic vertical section of the skin

The efferent sympathetic innervation is indicated by dashed lines (Adapted from Boucsein 2012)

The electrodermal phenomena refers to the changes in the electrical properties of the skin when sweat is secreted by the eccrine glands (Dawson, Schell and Filion 2007). Early research on electrodermal activity (EDA) was mainly from studies on cats, and although these results cannot be entirely synonymised with humans, understanding relating to the central origins of EDA was attributed to them (Boucsein 2012). As early as 1878, when researchers Herman and Luchsinger performed a stimulation study on a cat's sciatic nerve, documenting a sweat response to the stimulation, and at the same time an electrical current was detected (Neumann and Blanton 1970). Throughout the 1920s and 1930s research continued on the electrodermal phenomena, linking sweat gland activity with electric current flow. Herman also reported that when a human's body produced copious amounts of sweat, these sites had a greater electrical current than other locations on the body (Neumann and Blanton 1970).

With further explorative research throughout the years, EDA is currently the most frequently applied human bio signal in the psychophysiology research domain (Dawson, Schell and Filion 2007: 159, and Lykken and Venables 1971). Although, as with any complex body system there are still gaps in our understanding of all the factors, with reference to the central and peripheral mechanisms of EDA. What has been established through the years is a clear construct of the methodological procedures of EDA and it is hoped that with continual use of EDA in the research community, further elucidation on the central mechanisms will be gained (Boucsein 2012).

The main reasons for the popularity of this biosignal is the relative ease in which an electrodermal response can be obtained after a stimulus has been introduced, and the direct relationship between the intensity of this response and the intensity of the presented stimulus (Boucsein 2012). EDA is widely used in laboratories and field studies, due to the relative low cost and simplicity of the equipment required, making it a viable and popular choice for researchers working in the psychophysiology domain.

An advantage of measuring EDA from other biosignals, is that the activity solely reflects sympathetic branch of the ANS without any corruption from the parasympathetic branch (Braithwaite, Watson and Jones 2013). Although this is not unique, it is relatively unusual as many of the ANS's actions have an influencing factor from both divisions of the ANS. The innervation of the eccrine glands, that are responsible for EDA has been debated through the years, and centred around the cholinergic neurotransmitter, acetylcholine, which is mainly associated with being a parasympathetic neurotransmitter rather than the adrenergic nerve fibres of the sympathetic nervous system, which uses adrenaline as a neurotransmitter (Venables and Christie 1980). Studies have now provided convincing evidence that EDA is controlled by the SNS without the influence of any parasympathetic activity (Shields *et al.* 1987, Wallin 1981).

The term electrodermal activity was first introduced by researchers Johnson and Lubin in 1966, although, at that time and occasionally still seen today, the older notation of galvanic skin response or galvanic skin reflex was used. This term is no longer recommended as it implies that skin is a galvanic element which does not fully encompass the variety or complexity of the electrodermal phenomena. Other terms have also been attached to EDA, with an example being in the neurology domain, where researchers will often use the term peripheral autonomic surface potential (PASP) (Boucsein, 2012). An accepted list of terminology was produced in 1967 by the Society of Psychophysiological Research and is generally seen now in the majority of research papers (Brown 1967).

Mental or emotional sweating appears to occur mainly on the palms and soles of the body when the individual is under stress conditions or in an aroused state and is independent of the surrounding ambient temperature (Millington and Wilkinson 1983, Storm 2001). Millington and Wilkinson reported that emotional sweating also occurred in the axillary, forehead and genital areas. Contrary to this view point, Allen, Armstrong and Roddie (1973) reported that emotional sweating was seen at all other body sites and proposed that the amount of sweating was directly proportional to the number of sweat glands and not to distinct locations sites on the body. Bouscein (2012) commented that further investigation on the specific role of the palms and soles in emotional sweating was required.

A proponent for EDA research is Professor Rosalind Picard from MIT Media Lab, where she is the director of Affective Computing Research Group. Picard and her research team have created wearable sensors to assist in the communication and measurement of emotions. These wearable sensors have been used in a wide range of behavioural studies as well as providing information on sleep and seizure patterns (Poh *et al.* 2012, Sano, Picard and Stickgold 2014). Research studies have been carried out to validate ambulatory measurements of EDA from the wrist and ankles in comparison to the traditional EDA palmar sites (Val Dooren, De Vries and Janssen. 2012).

2.1.2.1 Applications of EDA

As EDA is a sensitive measure of sympathetic arousal, and is associated with cognition, attention and emotion, thousands of papers have been generated on the applications of EDA, covering a wide spectrum of domains (Boucsein 2012). Within the medical field EDA is studied to diagnose specific skin diseases. In the condition hyperhidrosis, where sweat is mainly dispersed from the palms, a repeated cycle is seen (Boucsein 2012). As the individual responds to an emotional event there will be an increase in sweat production, which is felt by the person, and as he or she is experiencing this uncomfortable response, a second emotional response is evoked, thus forming a positive feedback loop with further sweat being produced (Boucsein 2012). Recording EDA has also found favour in behavioural and personality studies, with focus often directed on schizophrenia research. In individuals that have schizophrenia, research has indicated that dysfunctions in their EDA may hold relevant information that can provide an insight into the development of symptoms (Dawson and Schell 2012).

Outside the medical profession, EDA measurements have been studied in human-computer interaction (HCI), where mental and emotional stress are key areas of focus for psychophysiologists to study (Schaefer and Boucsein 2000).

In aviation research, EDA has not had the general acceptance that cardiovascular indices have. Pilots appear to have a poor compliance with EDA monitoring, which may be due to the monitoring equipment being visible (Boucsein 2012). Aviation studies that have utilised EDA, do so, as part of a battery of measurements alongside the cardiovascular indices and EMG measurements. In these studies, pilots are commonly being monitored with respect to emotional stress and workload (Kahabka *et al.* 1986, Lindholm and Cheatham 1983). Although in the Lindholm and Cheatham study there was an indication that EDA demonstrated some superiority over heart rate (HR) in detecting persistent emotional strain, the researchers did not use EDA as a psychophysiological metric in any other studies that they conducted with pilots (Boucsein 2012).

In 2002, Wilson ran a small study on 10 male pilots where they had to perform flight manoeuvers in a small single engine aircraft. EDA was measured with ECG, EEG and EMG, with results from this study showing that EDA, measured as non-specific skin conductance response (NS.SCR) frequency, peaks over specific events, such as take-off and landing in parallel with peaks in the HR recordings and also in the subjective workload rating scale that the pilots completed.

An interesting laboratory study using EDA to determine a pilot's vigilance status was conducted by Haarmann, Boucsein and Schaefer (2009). The importance of this study was to assess the threat of hypovigilance (due to monotony) when situational demands change rapidly. Hypovigilance is of concern, especially in long haul flights where automation of pilot tasks is normal. Although the study was conducted in a laboratory, the use of EDA as a measure of any operator's vigilance may be seen as a promising development, and worthy of further investigation.

Although EDA application within the aviation industry can provide insightful data, as mentioned earlier there appears to be a pilot compliance problem. It will be of interest to see if the wearable sensors, that are emerging as reliable alternatives to the traditional EDA placement sites, will encourage a new generation of studies (Poh, Swenson and Picard 2010, Picard, Fedor and Ayzenberg 2015).

With relevance to this thesis, EDA has been used in automotive studies to assess the level of stress and workload when driving a vehicle. In this discussion stress is referred to as a negative occurrence, which is caused when a driver's workload has increased. In other research areas stress can be thought of as two separate components, namely 'distress' and 'eustress'. Distress would be the negative component, whereas eustress is termed a good stress, where a stressed situation would lead to a positive outcome (Kopin, Eisenhofer and Goldstien 1988).

Early studies have indicated that EDA was increased when drivers travelled on urban roads compared to driving on rural roads and decreased when roads were edged-lined compared to unmarked (Brown and Huffman 1972). In a more recent study, Healey and Picard (2005) conducted a physiological sensing study, to determine a driver's overall stress level. Participants were asked to drive a planned route that incorporated periods of rest, highway and city driving conditions. Physiological data included; EDA, ECG, respiration and electromyogram (EMG). Results indicated that for the majority of drivers, skin conductance and heart rate metrics were closely related to their driver stress level, which had been captured using video recordings.

MIT researchers from the AgeLab and New England University Transportation Centre have conducted numerous physiological sensing trials using EDA and heart rate to assess levels of cognitive demand under driving conditions (Mehler *et al.* 2009, Mehler, Reimer and Wang 2011, Reimer *et al.* 2010). To assess cognitive demand, a verbal response delayed digit recall task, termed n-back has been devised to represent similar cognitive resources that are used when interacting with an in-vehicle device (Mehler, Reimer and Dusek 2011). The task has three levels of difficulty which can be presented to the participants to assess the physiological metrics over levels of task difficulties and determine if the EDA and heart rate can differentiate between these different levels of difficulty.

A study in 2012 which assessed the sensitivity of EDA and heart rate for discriminating between the three levels of the auditory n-back cognitive task across different age groups found that both skin conductance and heart rate significantly increased with each incremental increase in the n-back digit task (Mehler, Reimer and Coughlin 2012). Authors discuss that with these findings a confidence in EDA and heart rate to measure degrees of mental workload should be developing and applied to HMI design projects. Perhaps the MIT researchers were keen to validate their earlier findings relating to EDA as a psychophysiological metric that could differentiate between cognitive demand levels. Contrary to the MIT findings results from the large European HASTE project, indicated that skin conductance was sensitive to visual but not auditory secondary tasks during a simulation trial and that there was no significant differentiation across the different levels of task difficulty (Engstrom, Johansson and Ostund 2005).

2.1.3 Cardiovascular Physiology

Cardiovascular physiology is the study of the function of the heart, blood vessels and blood (Fox 2011). Vessels transport blood containing vital substances such as oxygen and nutrients to every cell in the body, whilst metabolic waste products including carbon dioxide are transported in the blood to the external environment via the lungs, kidneys and liver. The importance of the transportation of blood cannot be underestimated, as it is important to remember that if transportation of blood ceases, a state of unconsciousness would result within approximately 30 seconds and irreversible brain damage may occur within a three to six-minute period (Mohrman and Heller 2014).

Cardiovascular measurements have been widely used in clinical, laboratory, and field studies and have a wide range of applications. Their use has been accepted due to their 'real-time' reliability, non-invasive nature and ease of recording.

The heart is a four chamber, muscular pump that propels blood separately but simultaneously into the systemic and pulmonary circuits by the alternating relaxation and contraction of the chambers (Mohrman and Heller 2014). In the normal heart beat two types of cardiac muscle cells are involved; the cells of the 'conducting system' that control and coordinate the heart beat and the 'contractile' cells that produce contractions to propel the blood through the heart (Fox 2011). Each heartbeat begins with an electrical action potential generated at the sinoatrial node that is found embedded in the right atria wall close to the superior vena cava.

These specialised conducting cells of the sinoatrial node are referred to as the pacemaker cells (Mohrman and Heller 2014). The pacemaker cells depolarise spontaneously towards the threshold for the formation of an action potential to occur (Martini 2006). Any of the cells of the conducting system can potentially generate an action potential but the cells of the sinoatrial node depolarise and reach threshold first, therefore it is the pacemaker cells that determine the heart rate in a normal beat. In the absence of neural or hormonal influence the sinoatrial node would produce an intrinsic heart rate of approximately 100 beats per minute (Widmaier, Raff and Strang, 2006).

Unlike skeletal muscle cells, cardiac muscle cells are electrically linked to each other and therefore when an action potential is initiated in one cell, the impulse spreads throughout the cell and has the ability to start action potentials in adjacent cells which are joined by structures named intercalated discs (Fox 2011). This process leads the heart to function as a single cell. This unique behaviour of cardiac muscle is often termed a functional syncytium. The physiology of this lies within the intercalated discs where there are minute cytoplasmic open channels, which provide a source of contact between the intracellular fluid of the neighbouring cells (Martini 2006).

As the action potential spreads from cell to cell throughout the right and left atria they both contract at the same time (Mohrman and Heller 2014). The impulse now depolarises the next conducting component, the atrioventricular (AV) node. The AV node is located at the base of the right atrium and forms the start of the conducting pathway to the ventricles (Mohrman and Heller 2014). After initiating the impulse in the AV node the impulse travels throughout the interventricular septum, a structure that divides the left and right ventricles (Fox 2011). This conducting pathway includes specialised conducting fibres that are termed the bundle of His (Widmaier, Raff and Strang, 2006).

The AV node and the bundle of His are the only electrical link between the atria and the ventricles due to a non-conducting layer of connective tissue that separates the cranial and caudal chambers (Mohrman and Heller 2014). An important characteristic of the AV node is that it has a slow conduction rate of impulse, which creates a short delay between atrial and ventricular contraction. This delay is important as it allows atrial contraction to be fully completed before ventricular contraction begins (Mohrman and Heller 2014). The impulse is now conducted rapidly and caudally through the bundle of His, which branches and eventually leaves the interventricular septum to enter the walls of the right and left ventricles (Fox 2011).

The left and right bundle are connected to the Purkinje fibres, which are large conducting cells that distribute the impulse to the rest of the ventricular myocardial cells (Martini 2006). This rapid distribution of the impulse causes the right and left ventricles to contract simultaneously, beginning at the apex of the ventricles and moving cranially (Martini 2006). This specialised conduction system of the heart enables each heart beat to follow a specific sequence. In a normal heart beat both atria contract together followed by a brief pause before both ventricles contract simultaneously. Finally, the heart relaxes and refills with blood before the next beat is initiated by the sinoatrial node pacemaker cells (Fox, 2011) Figure 13.

Figure 13. Electrical conduction pathway of the heart (BioBook 2014)

2.1.3.1 Autonomic Nervous System's Role on Heart Rate

The ANS is not the only system that has an influence on the heart but it is one of the most important (Ernst 2014). The SNS and PNS play an important and dual role in regulating and maintaining heart function. Sympathetic postganglionic neurons are to be found in both the sinoatrial node, atrioventricular node (AV), the conduction system, myocardial fibres and the coronary vessels. By releasing adrenaline, the beta-adrenergic receptors on the cell membranes of the sinoatrial node cells are activated, which gives rise to a positive chronotropic and inotropic effect, to increase heart rate and strength of contraction. In addition to specific adrenaline being released, circulating adrenaline from the adrenal medulla helps to increase this local effect of increased heart rate (Fox 2011). Adrenaline speeds up spontaneous depolarisation of the sinoatrial node cells and therefore raises the heart rate above the intrinsic level. Heart rate above the intrinsic level usually occurs during exercise or arousal (sensory / emotional).

The majority of parasympathetic neurons are located within the SA node and AV node sites with a fewer number of neurons being found within the atrial and ventricular myocardium (English and Jones 2012). The parasympathetic neurons release acetylcholine on the muscarinic cholinergic receptors located on the cell membranes of both SA and AV node cells which results in an increase in the resting potassium conductance into these cells and causes hyperpolarisation of the sinoatrial cells which slows down conductance through the AV node and results in decreasing the heart rate (English and Jones 2012). Heart rate is also decreased by Ach increasing the threshold for firing the SA node cells, which is the opposite action from the sympathetic system (Iversen, Iversen and Saper 2000).

The sympathetic system innervation is slower than the parasympathetic system, such that for the heart rate to rise there is an approximately a one to three second delay, reaching a steady state after 10 to 30 seconds (Ernst 2014, Malpas 2010). This slow heart response is considered to be due to the adrenaline diffusion rate within the slow-moving adrenaline signal transduction system (Malpas 2010). Parasympathetic changes occur rapidly with the vagal nerves exerting a beat-to-beat control of the heart function. Complex interactions between both autonomic branches allows the heart rate to be adjusted throughout the day. As mentioned earlier although the two main neurotransmitters are Ach and adrenaline there are several other mediators that have influences on the heart. An example of the complex interactions between the systems would be that the transmitter Neuropeptide Y (NPY) can be released by the neurons of the SNS and have an inhibitory influence on the neurons of the PNS (Ernst 2014).

Originally the interaction between the two branches of the ANS was described as a continuum with parasympathetic and sympathetic activity at each end of the pole. This classical model proposed by William Cannon inferred a mutual influence of the two branches where if parasympathetic activity decreased there would be an increase in sympathetic activity. Research studies in recent years have challenged this model and studies have indicated that influences from higher neural systems may produce reciprocal, independent or coactive changes in both branches of the ANS (Berntson and Cacioppo 2004). Studies in heart failure patients have been able to corroborate these findings (Porter et al 1990). Berntson and Cacioppo (2004) now believe that the PNS and SNS are tonically active during most normal and emergency conditions and operate in combination with each other as well as the somatic motor system. Although often the PNS and SNS produce opposing effects on their target tissues, it is the tonic balance between the PNS and SNS that helps to maintain a healthy internal environment despite the often changing external conditions (Iversen *et al.* 2000). In general, it is still reasonable to state that during rest the parasympathetic system dominates, whereas during activity the sympathetic prevails (Furness 2006).

An electrocardiogram is a visual representation of the electrical events occurring in the heart that can be captured from the surface of the body by a procedure termed electrocardiography (Cunningham and Klein 2007). The electrocardiogram is a frequently applied diagnostic tool when examining electrical abnormalities of the heart. Electrocardiography involves placing electrodes on the surface of the body at specific locations that capture the different voltages and then these voltages or electrical events are either displayed on a visual display unit (VDU) or printed out on a paper strip (Martini 2006). By manually inspecting the electrical events valuable information on the different components of the heart signal can be gained. A normal ECG has several different components which are often termed waves, each wave corresponds to the activity of a separate heart chamber (Mohram and Heller 2014), (Figure 14).

Figure 14. Representation of waves in a normal ECG (Student nurse diaries 2006)

The first small deflection in figure 14 is termed the P wave and represents atrial depolarisation (Mohram and Heller 2014). Approximately 25 milliseconds after the start of the P wave the atrial chambers will start contracting and blood will be forced into the ventricles (Mohram and Heller 2014). The large deflection in the ECG is termed the QRS complex and represents ventricular depolarisation. The ventricles will start to contract soon after the R wave peak. This large deflection is due to the muscular composition of the ventricles being far greater than the atria and therefore the electrical signal is stronger (Mohram and Heller 2014). The next small deflection is the T wave which represents ventricular repolarisation. There is no wave seen for atrial repolarisation as this occurs at the same time as ventricle depolarisation and is masked by the QRS complex (Mohram and Heller 2014).

As the waves in an ECG relate to specific electrical events in the heart, the time between these waves can also give additional information to the state of the heart. These time periods are termed intervals and segments, (Figure 15).

Figure 15. A normal ECG displaying intervals and segments (Scientific Research 2012)

The first important interval on an ECG is the PR interval which corresponds to the time between the start of atrial depolarisation and the start of ventricular depolarisation. The R-R interval, which is the time between the successive R waves corresponds to the time between ventricular contractions and therefore can be used to determine ventricular rate or more commonly known as heart rate (Fox 2011).

2.1.3.2 Applications of Heart Rate to assess workload and stress

Paper presented at the HCI International Conference 2013,
Autonomous Control in Military Logistics Vehicles: Trust and Safety Analysis,
Gempton, N., Skalistis, S., Furness, J., Shaikh, S. and Petrovic, D.

Measurement of heart rate is used daily within clinical settings and provides information to clinicians regarding their patients. Used alongside other measurements and tests it can guide clinicians to treatments or further investigations. Heart rate can also be used as an evaluation tool in a wide range of domains. Within aviation studies as early as 1917, it was used to monitor Italian pilots (Roscoe, 1992). The simplicity of recording and analysis of heart rate has given this measurement popularity throughout the decades (Kramer, 1990).

Heart rate recordings have been applied in a wide range of research topics that explore human information processing and mental workload. In addition to the aforementioned, heart rate has been used constantly as a marker in many sport disciplines and with mobile devices, individuals can now monitor their own heart rate during the day’s activities as well as monitoring it during sleep at night. Individuals who train regularly using cardiovascular exercises have lower resting heart rate than individuals that do not exercise regularly (Martini 2006). Heart rate has also been assessed when exploring human emotional states. Kreibig (2014) has demonstrated that several of the positive emotions have produced an increase in heart rate, these include emotions of joy and pride. An increased heart rate has also been linked with memory enhancement, when participants have viewed emotional arousal stimuli (Jennings and Hall 1980). Within the psychophysiology domain HR has been intrinsically linked with emotional studies to demonstrate the orienting response and how this is reflected in a deceleration in HR when participants are initially presented with an emotional arousal material (Abercrombie *et al.* 2008). Studies have also shown that tonic HR is raised higher for emotional arousal material than neutral stimuli HR (Cahill *et al.* 1994, O'Carroll *et al.* 1999). In the gaming industry HR has been used to evaluate user experience with entertainment technologies (Mandryk and Inkpen 2004). Heart rate has proved to be a reliable, unobtrusive measure that is easy to collect and by far the most frequently applied psychophysiological measurement in laboratory and field research studies (Wilson 2002).

In the aviation industry, early studies by Roscoe (1979), measured heart rate to examine pilot workload during Harrier ski-jump take-offs, compared to conventional short take-offs from a runway. From the data collected the study concluded that ski-jump take-offs were no more difficult to execute than conventional take-offs. In another study heart rate was used alongside pilot performance and subjective ratings to assess an automatic ground collision avoidance system in high performance military aircraft. The pilots flew on a collision course to the ground and then activated the avoidance system as close to the ground as they felt comfortable. With each test run, heart rate increased and also strongly correlated with the self-subjective rating for pilot anxiety that they completed after each flight (Hunn and Camacho 1999).

Whilst HR measurements have been used consistently in the aviation literature to assess the workload of pilots and air controllers for a considerable time (Wilson 2002), the literature in the automotive industry has been rather contradictory in terms of the sensitivity of indices to assess workload. This may be due to different methodology strategies being employed or due to a limited quantity of literature in the research domain (Son *et al.* 2010).

Research scientists from the MIT AgeLab in the United States use a working memory verbal response delayed task (n-back) to produce similar cognitive resources that are employed when drivers interact with voice commands from in-vehicle devices (Mehler, Reimer and Dusek 2011). It is not only in-vehicle devices that utilise cognitive resources, a simple conversation between the driver and passenger may use cognitive resources that can distract the driver from the primary task. This verbal response task has different degrees of difficulty which enables researchers to challenge the drivers' cognitive workload whilst participating in an on-road driving task. Studies have shown a stepwise increase in heart rate as the secondary task becomes increasingly more difficult (Mehler et al. 2008, Coughlin, Reimer and Mehler 2009, Mehler, Reimer and Coughlin 2012).

Brookhuis, De Vries and De Waard (1991) found that heart rate increased when participants were engaging in a telephone conversation at the same time as driving a vehicle, with another study showing an increase in HR when drivers entered a roundabout (Brookhuis and De Waard 2001). Richter *et al*. (1998) applied a range of physiological measures to identify unsafe curvature in established road networks before ergonomically designing new roads which would improve driver performance. This was achieved by first monitoring participants whilst they drove on roads with varying curvature levels. The authors concluded that heart rate data was the most sensitive of the physiological measures.

However, other studies using HR and other physiological indices have not found significant differences in HR between trial tasks. A large European project, HASTE, that ran studies assessing levels of difficulty on two different tasks types in simulator and 'on road' conditions found no significant differentiation across the different task difficulty levels (Engstrom, Johansson and Ostlund 2005). A road safety study in 2004 compared the responses of elderly participants and young participants in a simulator with different visual road configuration. The aim of the study was to assess how much visual road information was required to drive safely. A range of different parameters were measured, but with reference to HR there was very little difference found between the different road presentations (De Waard, Steyvers and Brookhuis 2004).

Heart rate will continue to be applied in many research areas and clinical applications, as it is a reliable, unobtrusive measure that is easy to collect and analyse. It has been the most frequently applied psychophysiological metric in the laboratory and in field research studies. However, with the increasing demands on pilots and vehicle drivers, cognitive load is an area that the research community will always investigate to gain a greater understanding of how individuals will react to new technology. When designing new technology, the major consideration is to lessen the distraction of the operator from their primary task, in this case driving a car or piloting an aircraft. Researchers will always seek new determinants that may offer additional and new information on their selected field.

An area that has gained significant interest is heart rate variability, as a marker of predicting cardiac mortality and evaluating the complex interplay between the sympathetic and parasympathetic divisions of the autonomic nervous system. Interest in these areas has accelerated in the research community, with physiologists and a collaborative input from engineers, physicians, physicists and mathematicians actively researching this area in a wide range of domains (Kamath *et al*. 2013). This multidisciplinary group alongside advancements in computer technology and signal processing algorithms have greatly increased the understanding of the electrical heart rhythms that are captured in an electrocardiogram (ECG) (Kamath, Watanabe and Upton 2013).

2.1.4 Heart Rate Variability

The easiest description of heart rate variability (HRV) is simply the variation between each heart beat in time. This variation can be measured in many different ways, using an ECG, pulse waves and heart tones. In practice, the most practical and precise method is to measure the distance between the R–R wave deflections that are seen on an ECG. The time measurement is usually recorded in milliseconds, see Figure 16.

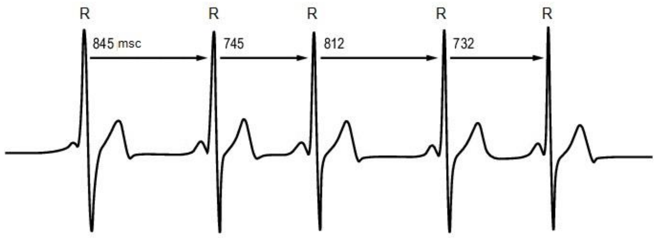


Figure 16. Illustration of variation between a R-R interval time series (adapted from Burnoutprotector 2016)

As the autonomic nervous system is considered the primary heart rate regulator, by studying the variability that occurs in the intervals between each heartbeat, it should be possible to determine an individual's functional state of their autonomic nervous activity (Khandoker *et al.* 2013). By applying mathematical models and digital processing applications, measurements can be gained on an individual's level of parasympathetic and sympathetic modulation of the heart rate (Khandoker et al. 2013). Inferences can then be generated, with regard to their autonomic nervous system and, over time predications may assist in clinical diagnosis and treatment, in a wide range of domains. Heart rate variability centres around the rhythm of the heart and within this rhythm there are oscillations either decreasing or increasing depending on various internal and external factors. This rhythm is essentially heart rate variability (HRV). In general terms, a healthy individual will display a large degree of variability, whereas a low variability within the R-R interval data, is often seen in an individual that has an underlying or current medical condition (Figure 17). In panel A of Figure 17, the illustration shows a considerable amount of variability is present, whereas panel B shows a relatively flat signal with far less variability.

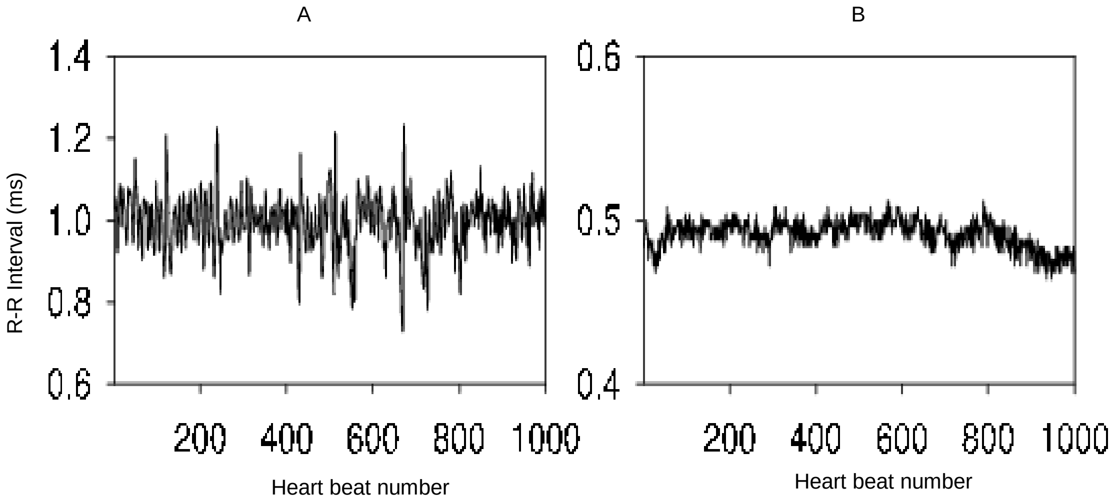


Figure 17. R-R interval series of a healthy individual
(A), and an individual with heart failure (B) (adapted from Goldberger *et al.* 2000)

This beat-to-beat variability of the heart rate is frequently assessed in clinical settings using an electrocardiograph which produces a recording of the heart signal. When examining an electrocardiogram for assessment of the electrical activity in the sinoatrial node it would be assumed that the P-P intervals should be the area of interest as they reflect the pacemaker cells, however, due to the small amplitude of the P wave, determination of this wave is often difficult and unreliable for accurate measurements (Kamath, Watanabe and Upton 2013). Therefore, the HRV measures are determined by the R-R waves which are prominent and easier to detect and read. Although the R-R waves reflect the fluctuations in the atrioventricular conduction they have been recognised to mirror the modulations of the sinoatrial node fairly accurately (Kuusela 2013).

Some of the earliest studies were carried out by Hon and Lee in the 1960's. Their studies focused on the foetal HRV during labour and the relation it had to foetal health (Hon and Lee 1963). They recorded in 1965 that foetal distress was preceded by changes in the interbeat intervals. These changes that occurred were recognised before any alterations in heart rate were detected (Hon and Lee 1965). In 1978, Wolf et al. highlighted that there was an association between a reduced HRV and a higher risk of post myocardial infarction (MI) mortality. The propensity for lethal arrhythmias and enhanced sympathetic or reduced parasympathetic activation focussed efforts to develop a quantitative marker of autonomic activity (Wahab 2012). By the 1980s HRV had been established as a marker of autonomic tone of the heart and as a strong independent predictor of mortality after an acute MI (Bigger *et* al. 1992, Kleiger *et al.* 1987, Malik *et al.* 1989).

Further advancement in HRV analysis came when Akselrod *et al.* (1981) demonstrated that by using power spectral analysis the heart rate signal could be decomposed into specific frequency bands. Further investigation using pharmacological interventions illustrated that the sympathetic and parasympathetic divisions of the autonomic nervous systems contributed to the heart rate power spectrum (Akselrod *et al*. 1981). The prospect of being able to quantitatively assess the autonomic nervous system has seen the number of publications regarding heart rate variability flourish over the last three decades. Between 1981 and 1990 there were 269 publications on heart rate variability whereas during the period from 2001 to 2010 there were been over five thousand (OVID Technologies 2012).

During this period of interest in heart rate variability, there has been progress in understanding of the complex components that contribute to the signal, however, there are still large areas within the research field still to be determined. These areas include the determinants of the signal and how they relate and impact the sympathetic and parasympathetic nervous systems. There is also a lack of standardisation on the different methodology approaches and procedural processes that are taken when planning and using HRV in studies. This is perhaps why physicians, even cardiologists, are reluctant to use HRV metrics as part of their routine clinical assessments (Kamath *et al.* 2013).

In 1996 there was recognition of standardisation problems which led to the European Society of Cardiology and the North American Society of Pacing and Electrophysiology setting up a specific Task Force to address these issues (Task Force of The European Society of Cardiology and The North American Society of Pacing and Electrophysiology 1996). Specific aims for the Task Force were to develop correct nomenclature and definitions, measurement methodologies, relationships between physiological and pathophysiological states, appropriate applications and future research. The Task Force publication provided researchers from all fields with a set of guidelines to inform their research methodology. In doing so, it was aspired that future research would follow similar protocols and there would be a level of homogeneity between research publications (Task Force 1996).

Since the publication of the 1996 Guidelines there has been a surge of research papers being published on HRV which have benefitted from the Task Force's recommendations. Although, since the Task Force's recommendations further technological advances have provided additional mathematical models for processing and analysing biosignals, there is now a requirement for a new review to take place that would encompass the advancements seen in HRV studies

(Nunan, Sandercock and Brodie 2010). Another area of weakness that the Task Force highlighted was the lack of population studies with longitudinal reassessment and to date, this is still being identified in many publications (Nunan, Sandercock and Brodie 2010, Sztajzel 2004).

As yet, there is still no 'gold' standard in all areas of HRV research, which perhaps makes it an exciting field to explore, as there are few definitive rules that have gained full consensus throughout the HRV research community. One example of this would be the duration of time that your signal should be recorded. The Task Force recommended in 1996 that it was important to establish a recording time. They suggested that the recording time should be divided into short-term and long-term epochs. Short-term recording should be over 5 minutes, whereas the long-term recordings would be over 24 hours. As recording equipment and processing technologies have significantly advanced since the Task Force's report, in principle it is now possible to measure HRV for minutes, hours and days and produce valid results (Ernst 2014). There have been several studies conducted to assess the reliability of measuring HRV over short time periods from seconds to minutes, with findings indicating that a short-segment approach can be appropriate to assess short periods of task demands (De Rivecourt *et al.* 2008, Stuiver *et al.* 2012).

Another area that continues to be assessed and debated is the different techniques available to analyse HRV, and which of these techniques produce the best representation of HRV. One paper reported that at present there are 18 different measurement methods to determine the HRV of an individual. This number of measurement parameters and the methods used to determine HRV will only increase further, as different mathematical models of the time signal are being actively researched by mathematical and engineering domains (Khandoker et al. 2013). To date, the most frequently applied methods of measuring HRV are the time domain and frequency methods.

Time domain methods are the simplest and usually the preferred choice when analysing long-term recordings and are considered markers of an overall heart rate variability. The selected measures described in Table 1 all use R-R interval differences. The term normal-to-normal (NN) interval is often used instead of R-R intervals, as it establishes that all the R-R intervals are between adjacent QRS complexes and have resulted from sinus node depolarisations (Task Force 1996). A selection of the time domain methods are described in Table 1.

Table 1, Selected HRV time-domain measures (adapted from Khandoker 2013)

HRV Measure	**Units**	**Description**
mRR	ms	Mean RR.
SDRR	ms	Standard deviation of RR intervals.
RMSSD	ms	Root mean square of successive differences of RR intervals.
NN50 count	pairs	Number of pairs of adjacent RR intervals differing by more than 50 ms.
pNN50	%	Ratio of NN50 count to the count of all RR intervals expressed as a percentage.

Calculation of the above measures is relatively straight forward with no time-consuming computation, compared to that required in other methods. The limitation with the time domain method is that it represents an overall picture of heart rate variability but is unable to provide differentiation between the sympathetic and parasympathetic branches of the autonomic nervous system. To gain a greater insight into autonomic modulation, frequency domain analysis is applied.

2.1.4.1 Frequency Domain Analysis

Complex analysis has to be performed to examine the distribution of the frequency components of HRV. Power spectral density (PSD) analysis is one such method that decomposes the heart signal into various frequency bands and quantifies their relative intensity, termed variance of power (Sztajzel 2004, Kamath, Watanabe and Upton 2013). Algorithms are applied to compute the PSD function, which transforms the time domain signal into frequency components. Fast Fourier transform (FFT) and Pan and Tompkins algorithms are two that are frequently applied in heart rate variability studies (Tarvainen *et al.* 2014).

In short-term recordings the spectral power is divided into three frequency bands: high frequency (HF) 0.15-0.40 Hz, low frequency (LF) 0.04-0.15 Hz and very low frequency (VLF) 0.003-0.04 Hz. If spectral analysis is being applied to a long-term recording then an ultra-low frequency component (ULV, 0-0.003) can also be calculated (Ernst 2014). Another frequency band that has been reported in some cognitive studies (Duschek *et al.* 2009), is termed the mid-frequency band (0.08-0.15 Hz) (Huang *et al.* 1997). A selection of studies have stated that the MF magnitude is inversely related to an individual's degree of effort during a cognitive task (Bouscein 2000, Van Roon *et al.* 2004). Berntson *et al.* (1997)

disagrees that the MF is a separate band, and considers it as a variation of the LF band, and suggests that it should not be referred to as a separate component in studies.

The high frequency band correlates with heart rate variations that are linked to the respiratory sinus arrhythmia and are dependent on the respiration pattern. The fluctuations within this band are due to the efferent parasympathetic nervous system activity (Ernst 2014). There has been a number of studies exploring the influence that controlled breathing or paced breathing has on the frequency components of HRV. Sinnreich *et al.* (1998) stated that using a metronome with a breathing rate of 15 respirations per minute, placed all the participants HF power securely within the HF band, and further commented that the use of paced breathing would not be necessary to obtain reliable results during short-term recordings.

Another study that examined the effect of respiration on short-term variability indicated that the time domain indices were not significantly dependent on respiration rate, but the frequency indices were affected (Schipke, Pelzer and Arnold 1999). The authors commented that the location of the frequency band ranges are important, describing that if an individual had a respiration rate of 30 breaths per minute, this could still be in the HF band if it had been extended to 0.50 Hz (Schipke, Pelzer and Arnold 1999). Another comment from Schipke, Pelzer and Arnold (1999) was that a controlled or paced respiration may modify results and increase the parasympathetic activity. In a static recording it is highly feasible to control and use devices to pace the breathing rate, but in an active and task orientated study, paced or controlled breathing would cause participant distraction and may contribute to frequency disturbances that were caused by the participants controlling their breathing rate, rather than the responses to the task that was presented to them.

Goldberger et al. (2001) stated that as there were large inter-individual variations in HRV due to the complex relationship between the parasympathetic activity or vagal activity. In some individuals this could be shown by an initial increase in their HRV which was also reflected by an increase in vagal activity until a ceiling or plateau effect was reached. Beyond this point, if there was further vagal activity, HRV decreased rather than increase. Goldberger *et al.* (2001) proposed that this was due to a HRV saturation from an acute autonomic stimulation.

Fluctuations in the low frequency band are more complex to determine and open to debate. Earlier researchers stated that the LF was a marker of sympathetic modulation (Kamath and Fallen 1993, Malliani *et al.* 1991, Montano *et al.* 1994) whereas others argued that the LF component included sympathetic and parasympathetic influences (Appel et *al.* 1989, Akselrod *et al.* 1981). In agreement with the latter view, the majority of studies and research now state that the LF component is modulated by both parasympathetic and sympathetic activity, with its origin in the baroreflex feedback loop (Lanfranchi and Somers 2002). Arterial baroreceptors are triggered by a change in blood pressure, resulting in an adjustment of the heart rate through the central nervous system via parasympathetic and sympathetic activity (Lanfranchi and Somers 2002, Frenneaux 2004). An increased LF measure has been associated with mental and physical stress shown in a large cohort of different research fields (Ernst 2014).

The exact origin of the VLF has been debated, with no consensus of agreement, with researchers stating that it should be treated as noise (Khandoker et al 2013), whereas others believe that the oscillations in this frequency band are attributed to thermoregulation or to hormonal influences (Kamath, Watanabe and Upton 2013). Another view on the VLF band has been that it reflects sympathetic activity and is a factor in physical activity (Frenneaux 2004). Bigger *et al.* (1993) introduced the ULF band and stated that it reflected circadian and neuroendocrine rhythms. The ULF band is only included in long-term recordings and knowledge relating to the physiological correlates of this band are not fully understood (Task Force 1996).

Another measurement that is often applied in heart rate variability studies is the LF/HF ratio, which has been attributed to reflect the sympathovagal balance and quantifies the relationship between the sympathetic and parasympathetic nerve activities (Ernst 2014). Although there has been criticism relating to this HRV metric as an indirect measure of sympathovagal balance, as it reflects autonomic fluctuations rather than an absolute measure of autonomic nerve activity, (Eckberg 1997, Billman 2013) it is still generally reported in the majority of studies to provide a global picture of the sympathovagal balance (Ernst 2014).

Phase-rectified signal averaging, fractal and chaos theory methods are alternative models that are being purposely researched to gain a greater understanding and clearer view of complex biosignals, such as a series of R-R intervals. Some studies suggest that fractal analysis may be able to recognise abnormal RR fluctuations more effectively than time domain or frequency methods (Khandoker 2013).

2.1.4.2 Heart Rate Variability Applications

The measurement and application of HRV appears to have very few boundaries and although originally in the domain of cardiologists, HRV has widened its appeal in other clinical areas as well as stimulating enthusiasm in a diverse group of research areas. HRV is now being applied in many clinical assessments with examples in critical care medicine (Prietsch, Knoepker and Obladen 1994), chronic renal failure (Forsstrom *et al.* 1986) and fibromyalgia (Raj *et al.* 2000). Other areas where HRV is being applied as a measurement tool are, obesity (Birch, Duncan and Franklin 2012) exercise (Hellard *et al.* 2011) stress (Gillie, Vasey and Thayer 2014) and affective states (Katsis *et al.* 2011).

Although HRV is often reported in studies as a psychophysiological measure, as mentioned earlier it has not yet established a ‘gold’ standardised methodology and procedural protocol, and therefore, there continues to be heterogeneity across studies. The reasoning for this lack of standardisation is perhaps the very nature of the complexity and variability within the signal, and as yet, not a full comprehensive understanding of the cross-play between the different components that are represented within the frequency bands and how they relate to an individual’s physical and or mental state has not been reached (Khandoker *et al.* 2013.

For this thesis, it is the application of HRV and its relationship with emotional responses that are of interest in the first two studies. As HRV is influenced by sympathetic and parasympathetic nerve activity, it makes it a promising measure to explore as it may have the potential to illuminate physiological responses that participants experience when exposed to an emotional stimulus. In the final study it is the user experience and user engagement that will be examined, and HRV will be assessed to determine if different responses can be detected when participants interact with the in-vehicle interface systems.

A study by Riener, Ferscha and Aly (2009) applied HRV to monitor a driver's affective state when travelling a routine journey to and from location of work. For this study the selected frequency measurement was the LF/HF ratio used as an indicator for sympathovagal balance. Although this study involved only one participant the trial was conducted over a two-week period and included over 20 logged journeys. Driving routes were divided into morning and evening journeys and the varying road characteristics were mapped and time stamped. Results indicated that on specific points of the journey where the volume of traffic was high, the LF/HF ratio was also raised, indicating an arousal state of the driver.

Researchers, Mehler, and Reimer from the Massachusetts Institute of Technology (MIT) AgeLab and New England University Transportation Centre, regularly apply physiological indices to assess driver's workload levels in 'real' environments as well as simulated studies. With other collaborators from the MIT Agelab they have consistently found that HR has been sensitive to a range of different workload tasks. However, as many papers suggest that HRV metrics can illuminate additional information, that HR data in isolation cannot provide (Aasman, Mulder and Mulder 1987, Mulder 1992), they were keen to assess the merits of HRV in comparison to HR in a driving task with an additional cognitive workload challenge (Mehler, Reimer and Wang 2011). Results indicated that time domain metrics (except SDNN) detected an increase in cognitive workload between the single driving task to the most challenging cognitive workload task, termed ‘high’. The cognitive workload task was an auditory digit recall test, termed ‘n-back’ and has been discussed earlier. Although the time domain metrics had detected the driver’s physiological changes, they were less robust than HR and skin conductance levels (SCL). Both HR and SCL were able to distinguish the change between the driving task and the ‘low’ additional task demand (0-back). An interesting finding, and in contradiction to the Reiner study, was that the HRV frequency metric LF/HF was unable to differentiate between the two conditions (0-back and 2-back). The HF and LF both decreased to the change in workload. In an earlier real-world driving task study, using physiological measures to detect drivers stress, the HRV frequency indices, including LF/HF, performed well and in a similar way to HR (Healey and Picard 2005).

Rigas *et al.* (2011) also used the LF/HF HRV metric to assess drivers’ stress and fatigue levels in a simulated environment, along with mean R-R, the author noted that they were the best physiological measures to detect stress although their discrimination power was not high between the different classified fatigue levels. As the author discussed this could have been attributed to their circadian rhythm, as some participants were conducting the study during the night.

HRV has also been used as a psychophysiological metric to determine how the autonomic system responds to different emotional stimuli and situations that provoke emotional responses from the individual. The concept behind examining HRV is to illuminate changes within the parasympathetic and sympathetic branches of the autonomic nervous system that other physiological measures cannot achieve. HRV has been performed when measuring discrete emotions

(Kreibig 2010), emotional regulation (Gross and Feldman-Barrett 2011), and is associated with emotion recognition studies (Quintana *et al.* 2012).

Quintana *et al.* (2012) recruited sixty-five participants for the study which involved Reading the Mind in the Eyes Test (RMET). The RMET has been used in a wide range of studies using healthy participants although it was first developed to assess social cognition in adults with autism spectrum disorders. The test involves participants identifying different emotions when they are presented with images of only the human eye region (Baron-Cohen *et al.* 2001). The aims of Quintana's study were to test the hypothesis that heart rate variability (HRV) was related to social cognition and support the Porges' polyvagal theory.

Porges' polyvagal theory highlights that the role of the mammalian autonomic nervous system has developed to enable survival, reproduction and social interactions of mammals. In humans this may be seen by the inhibitory role that the vagal nerve has on the heart's sinoatrial node. This inhibitory role is demonstrated by the example, of an individual being in a threatening situation, whereupon the activity of the vagus nerve is inhibited which facilitates a series of physiological responses which assist in the individual's survival. These responses would include an increase in heart rate thereby increasing blood flow to the limbs. By contrast, when the external environment is non-threatening and safe, vagus nerve activity increases at the sinoatrial node which slows the heart rate down, encouraging social behaviour and homeostatic processes. Porges polyvagal theory surmises that strong social interactions are promoted by a relaxed physiological state (Porges 2003).

Quintana's research and subsequent paper was the first to examine the relationship between emotion recognition and HRV in healthy adults. The HRV metric that the authors focussed on for this study was the high frequency band (HF) which strongly correlates with the activity of the parasympathetic nervous system. Results from this study appear to agree with Porges' polyvagal theory, as participants that had a higher HF HRV were able to identify the emotional state that was being portrayed by the images in the RMET, to a greater extent than individuals with a lower HRV. The results suggest that the HF component of HRV may provide a promising novel marker for emotion recognition (Quintana 2012). As a reduced HRV has been found in many mental health disorders (Kemp *et al.* 2010, Hasin *et al.* 2007), this study may contribute knowledge on social cognition deficits and the role that the ANS has in social cognition and behaviour (Quintana 2012). However, another conflicting finding was highlighted as higher levels of stress predicted an increase in HF, whereas a study by the same author found that high levels of anxiety caused a decrease in HF. Existing studies support this view that a decrease in HF is often seen in stress related disorders and other psychiatric illnesses (Kemp *et al.* 2010).

2.2 Rationale and Introduction to Study One

The previous chapters have highlighted the many complexities that are involved in the pursuit of trying to understand how individuals react to different presentations of stimuli. There is not a unified single model that can explain the nature of all complexities involved in this process. There is however, a number of ways in which we can begin to build up a more comprehensive understanding, that encompasses all the components that make that effect of how we interact with everyday objects / systems. As outlined in the previous chapters, we understand a great deal in terms of the underlying physiological processes that underpin our physical response to different experiences.

This research examines some of these processes and connections that occur when an individual reacts to a new or different stimulus. As individuals, we enjoy pleasurable experiences, with evidence suggesting that we perform better and achieve more if the interactions we have with products or devices are pleasurable and enjoyable (Norman 2013). When an experience is good, it will be repeated whereas if the experience is negative, there will be no further engagement with the product or device and certainly no future interaction with it.

In the automotive manufacturing industry, the need for user satisfaction and enjoyment is key. If the manufacturers sell cars, then the business is profitable successful and will grow. Although this is a straight forward prediction, if we factor in the unique, unpredictable and complex individuals, we now have a far more difficult job at predicting what car they will desire to purchase. As discussed in Chapter One, in-vehicle HMI systems are now an important and key differentiator when a customer purchases a new car.

The need for manufacturers to 'get it right' is vital in this competitive industry. For designers and engineers the evaluation process of HMI systems is crucial for the success of their system and ultimately the new car. Although an evaluation process may have been part of the design pathway, there appears to have been a reluctance to investigate the physiological parameters that could provide an enriched picture, regarding a range of responses that a user would have when interacting with a new system. In order to establish more robust data in relation to how potential customers perceive their displays, the use of physiological measurements offers a method by which this could be realised.

The research outlined in this thesis will explore the use of physiological metrics to address the assessment of usability in relation to in-car vehicle displays, whilst also aligning these finding with the cognitive elements that we already know. This mapping between physiology-psychology could potentially offer the automotive industry a means by which a method can be adopted to achieve a better understanding of which design elements can lead to a more pleasurable and useable experience.

3 Chapter Three – Study One

3.1 Overview

In order to investigate a formal process by which user interaction with a visual is assessed, it was imperative that the methodology adopted to conduct studies needed to be robust. Establishing such a methodology was fundamental to the overall research aims in order to test the study hypotheses within the research objectives. In order to construct such a methodology a 'bottom-up' approach was taken, which enabled the consideration of methodological aspects and study design in a layered concept, with each layer being a step closer to a simulated 'real experience' that a driver may encounter when interacting with an infotainment HMI system. By assessing the different individual elements within each study, there was an opportunity to evaluate its suitability before continuing with the next study.

In order to assess both physiological and psychological processes, this research has examined both subjective and objective approaches that may be used to assist in the evaluation of automotive HMI systems. The first aspect of this research was to investigate the nature of the physiological measurements that could be employed, in essence the physiological responses that could be detected from the participants when they were presented with different stimuli. The use of self-report scales were employed to validate the same stimuli reflected in the participants' subjective responses. The use of self-report scales was also an important element to the current research as many automotive HMI evaluation processes rely heavily on these scales for feedback and evaluation. By running the HRV measurements in combination with the self-report scales, it would be possible to see if there was a relationship between the two different types of data and whether physiological indices could provide a degree of validation for subjective measures. Therefore, the research question throughout this thesis would be:

To what extent can physiological metrics validate self-report scales in the automotive HMI evaluation process?

3.1.1 Study One: Aims and Hypothesis

As discussed, the aim of the first study was to conduct an initial study in order to develop and establish a methodology and define a procedure that could be successfully implemented in the subsequent studies. At this early stage in the thesis it was important to assess whether heart rate variability metrics were sufficiently sensitive to detect, and be influenced by, emotional stimuli when they were presented to participants. By using a set of recognised validated emotional words, that have associated levels of emotional associated with them, the study's aim was to record physiological measures to determine if any physiological differences could be found when the ANEW words were placed into High, Neutral and Low categories. It was also important in the first study to compare the participants HRV measurement with their own personal rating of the ANEW words, using a validated self-report scale. This would be a valuable consideration throughout the whole thesis as it was a practical link for the automotive HMI evaluation process.

Another important consideration for this study was to test the physiological equipment and determine its feasibility for future studies in terms of both repeatability of measure and realistic application within an applied context (such as an automotive HMI configuration).

Hypothesis:

H_1: The use of a HRV metric would detect participant's responses to presented emotional stimuli

H_0: There would be no significant changes in the HRV metric detected, in participant's response, between presentation of different emotional stimuli.

H_2: There would be an association between the participant's HRV metric and their subjective rating to the presented different emotional stimuli.

H_0: The HRV metric would not reflect the participant's own rating of the different emotional stimuli.

Paper accepted at the Doctoral Colloquium Session at the Automotive UI 2013 Conference, Eindhoven. Exploring Heart Rate Variability for Automotive Human Interface Evaluation (Furness 2013).

3.1.2 Methodology

Study One was laboratory based. In accordance with Coventry University regulations an online ethics application was completed and submitted to Coventry University Ethics Committee where the application was reviewed and permission was granted to run this study (Ethics No: P7949). The study also adhered to the British Psychology Society (BPS) (2016) code of conduct for experimentation with human participants. Each participant was required to read the informed consent forms and participant information sheet before the study commenced (Appendix 1 and 2). Recorded HRV LF /HF data was assessed by a Shapiro-Wilk's test of normality, it was determined that the data did not conform to a normal distribution, and therefore a non-parametric test was applied. A Kruskal-Wallis test was conducted to determine if there were differences in the HRV LF/HF ratio between the three different groups (high, neutral low) sourced from the ANEW stimuli. Subsequently, a post hoc pairwise comparison were performed using Dunn's (1964) procedure with a Bonferroni correction for multiple comparisons to determine where the differences were in the data. Statistical significance was accepted at the $p < 0.05$ level.

3.1.3 Design

The study design was constructed using a within subject's design, comparing participant's physiological responses between presentation of different emotional stimuli.

3.1.4 Participants

The first study coincided with Stage 3 Undergraduate Biomolecular Science final year project students' studies, who were allocated three weeks for data collection. This enabled a 'pool' of willing volunteers to participate in the first study. Recruitment of students was facilitated by the researcher's supervisor. Twenty Coventry University students volunteered for this study. After the data was recorded from the participants, only ten data sets were able to be analysed due to the integrity of the signal. From the ten participants 4 females and 6 males, with a mean age of 21 years. A recruitment flow chart is illustrated in Figure 18 and outlines the original number of participants tested and gender balance.

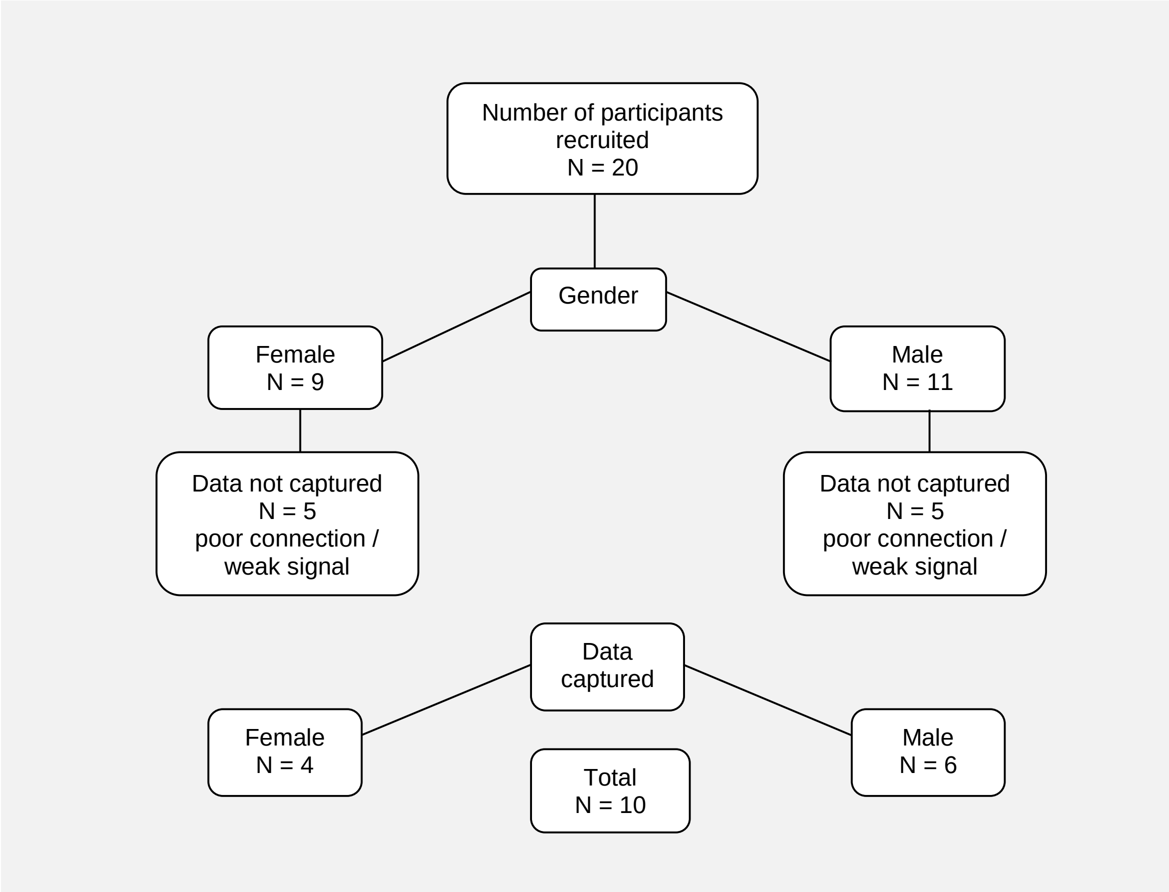


Figure 18. Study one, participant flow chart

3.1.5 Materials and Equipment

The emotional stimuli presented to the participants was developed by Bradley and Lang (1999) from the Centre for Emotion and Attention at the University of Florida. Words, used in the English language were the stimuli and came from a large catalogue entitled The Affective Norms for English Words (ANEW). There are over 1000 words which have been previously rated in terms of their affective dimensions of valence, arousal and dominance. To assess the affective dimensions, the Self-Assessment Manikin (SAM) system devised by Lang (1980) was used.

The SAM system comprises of bipolar icons that represent different values along each emotional dimension, (9-point rating scale), that help with self-expression over the three affective dimensions (Bradley and Lang 1999). Bradley and Lang (1990) determined that the SAM correlates well with the dimensions of valence, arousal and dominance. The SAM has a 9-point rating scale for each dimension (Figure 19).

Figure 19. Self-Assessment Manikin (SAM) for arousal, valence and dominance dimensions (Bradley and Lang 1994)

Figure 19 illustrates the icons that are scored by participants grading them. The first strip of icons (A) demonstrates the gradations for valence, from a smiling happy figure on the left to a frowning, unhappy figure on the right. Whereas icon strip (B) demonstrates the gradations of arousal from an excited figure on the left to a calm icon on the right. Icon strip (C) demonstrates the gradations of dominance, from a small figure on the left (dominated) to a large figure on the right (in control), (Bradley and Lang 1999).

The initial task for this study was to order the catalogue of words into three emotional groups based on their validated SAM mean rating score, supplied from the Centre for Emotion and Attention (1999), against the three affective characteristics of valence, arousal and dominance. The emotional groups were termed High, Neutral and Low according to their SAM score. The higher the score, indicated that each word was high in arousal content, high in valence content and high in dominance (Figure 19 narrative). Once the groups were ordered, according to their SAM score, the highest mean-rated twenty words from the High group were streamed, the middle twenty words from the Neutral group and the lowest twenty words from the Low group (Table 2).

A one-way ANOVA and Tukey Post Hoc statistical tests was applied to determine that there was a significant difference between the mean SAM scores of each of the grouped words. By ensuring that there was a significant difference in the SAM scores between the selected High, Neutral and Low emotional words, provided confidence that the selected words were clearly defined into the assigned emotional groupings and therefore provoke three different responses from the participants when they were presented with the words. Valence, $F(2,57)=875.248, p<.05$; Arousal, $F(2,57)=225.562, p<.05$ and Dominance $F(2,57)=392.181, p<.05$. The selected words were then formatted in PsychoPy (Figure 20) and presented in PowerPoint to the participants on a 40 x 31 cm screen, randomly to counterbalance any order of effects (Myers and Hansen 2011).

Table 2. ANEW words prior to random presentation

High Group	Neutral Group	Low Group
win	tool	loneliness
orgasm	black	failure
cash	lesbian	unhappy
thrill	razor	sad
kiss	news	sick
fun	market	gloom
graduate	ship	paralysis
victory	avenue	depressed
sexy	ketchup	depression
ecstasy	name	lonely
leader	hat	coward
desire	nursery	Inferior
promotion	doll	fatigued
admired	hit	poverty
excitement	office	dreary
fame	rough	loser
engaged	stove	Infection
erotic	concentrate	obesity
surprised	cannon	Illness
adventure	medicine	deformed

PsychoPy, is an open-source application for running studies in Python (Peirce 2009). PsychoPy provides the user with a large variety of stimuli that can be integrated into a study design. PsychoPy has two interfaces that can be utilised, one is termed the *'coder'* interface and is intended for individuals that prefer to program and a *'builder'* interface for individuals that do not wish to program. For all studies the *'builder'* interface was used, (Figure 20).

Figure 20. PsychoPy builder interface

the presentation of the recognition task and recording response and error rates (Peirce 2009)

The physiology equipment for this study included a *Suunto tc6* heart rate monitor, illustrated in Figure 21.

Figure 21. *Suunto tc6* heart rate monitoring equipment (Suunto 2011)

This piece of equipment has two components, the first being an elastic fabric band, with a transmitter positioned in the centre of the band, worn around the chest. The transmitter sends the heart signal to the receiver, which is a device, similar to a wrist watch. Data is then stored in the watch until the end of the study, when it is connected to a computer and downloaded for post-test analysis.

Before the transducer band was fitted to the participants, the electrodes on the band were dampened with water to improve sensor conductivity. If required, assistance was given to correctly place the chest band (Figure 22).

Figure 22. Correct placement of a *Suunto t6c* heart rate monitor (*Suunto* 2011)

To record participants respiration rate a thoracic respiration band (stethographic transducer) was secured around their chest, with the transducer fitment centred on their sternum. The respiration band did not have to be in direct contact with the participant’s skin and was placed on top of a light shirt or top as illustrated in Figure 23.

This material has been removed from this thesis due to Third Party Copyright. The unabridged version of the thesis can be viewed at the Lanchester Library, Coventry University.

Figure 23. Correct placement for thoracic respiration band (Biopac 2012)

For EDA recording two cupped reusable Ag/AgCl electrodes were positioned on the medial phalanges of the index and middle fingers of the participant's non-dominant hand (Figure 24) (Venables and Christie 1980), as the medial phalanges are less prone to movement and scarring effects. The EDA electrodes have an electrode cavity, where electrode gel is placed before attachment to the participant's fingers.

This material has been removed from this thesis due to Third Party Copyright. The unabridged version of the thesis can be viewed at the Lanchester Library, Coventry University.

Figure 24. Illustration to demonstrate position of EDA electrode (Furness 2016)

The recorded physiological data was stored on the *Suunto Training Manager* (2.3.0) then exported to *Kubios HRV* (version 2.1). *Kubios HRV* is an advanced software tool for studying heart rate variability and has been developed by the Biosignal Analysis and Medical Imaging Group (BSAMIG) at the University of Eastern Finland (Tarvainen *et al.* 2014). Kubios calculates a wide range of time and frequency variables using R-R interval data. For frequency analysis a power spectrum density (PSD) estimate is calculated for the R-R data. The R-R data is converted to an equidistantly sampled series by a cubic spline interpolation method before PSD estimation is performed by a Fast Fourier Transform (FFT) based on the Welch's periodogram method. By using this method, the data is divided into overlapping segments and the spectrum is obtained by averaging the spectra of these segments. This method helps to decrease the variance of the FFT spectrum (Tarvainen 2014). Kubios default settings were applied for the standard frequency bands as discussed in chapter two. A bandpass filter of 5 - 30 Hz was applied to reduce power line noise, baseline drift and other interference that may distort the measurements (Pahlm and Sornmo 1984).

At the start of each word being presented to the participants, a 30 second epoch was time stamped, this period of time was then manually inputted into the Kubios interface, where a range of HRV indices were derived. Although a range of HRV indices were calculated, for this study the main focus was on the LF/HF measurement, which is a ratio of the low frequency and high frequency component from the heart signal. This measure assesses the sympathovagal balance, which is reflected by the fractional distribution of power across the frequencies (Keenan and Grossam 2006, Malliani *et al*. 1994). LF/HF data was transferred to SPSS where inferential statistics followed.

3.1.6 Procedural Approach

On entering the laboratory participants were seated at a desk with a screen (40 x 31 cm) in front of them (50 cm) which was used to present the stimuli. Each participant had been requested to refrain from caffeine drinks for a minimum period of four hours before attending the session. Participants were asked to read and complete the informed consent form and health questionnaire. On completion of the forms the participants left their seat to be fitted with a heart rate monitor, electrodermal monitoring electrodes and a respiration transducer band (using the above protocol). The participants returned to their seat when the physiological equipment was fitted and were asked to close their eyes and relax for a period of five minutes, in order to record a resting baseline measurement (as per Hsia *et al.* 2009). Obtaining a stable baseline recording for the physiological measures was important as physiological results of the participants can then be compared against these measurements.

After the baseline recording, the researcher sat next to the participant to explain the task and give instructions on how to complete the SAM rating scale. The participants had an opportunity to ask any questions relating to the study before it commenced. An instructional slide was then presented to the participant with two words shown (not from study set) for a practice run at completing the SAM sheet. The researcher then positioned herself behind the participant for the start of the study.

Each participant was presented with 21 randomly selected words. Each word appeared on the screen for six seconds and during that period the participants had been instructed to concentrate on the word for the full length of time. The word was then removed from the screen and participants completed SAM, on their first reaction to the presented word (Figure 19). The participants were instructed to move quickly through the rating scale and avoid spending too much time thinking about each individual word. A period of 15 seconds was given to participants to rate each word over the three dimensions before an inter-stimulus interval of 15 seconds was given before the next word was presented (Lang and Bradley 1999). The participants completed the SAM to rate their personal reactions to the presented words. For each word the participant was instructed to move from left to right over the SAM rating scale and place a circle around the icon that best represented their immediate feelings of the presented word (Figure 19). An average score was then obtained for each word and image across the three dimensions.

Physiological measurements were recorded throughout the study and were later analysed. After all the words had been presented, participants were asked to close their eyes, before a second five-minute resting baseline recording was taken. Although no data was analysed from this period of time, it is considered as 'good practice' to run a second resting baseline, to ensure that the participants do not leave the laboratory environment in a stimulated state from the presented stimuli (Krahe *et al*. 2011). Study trial order is shown in Figure 25.

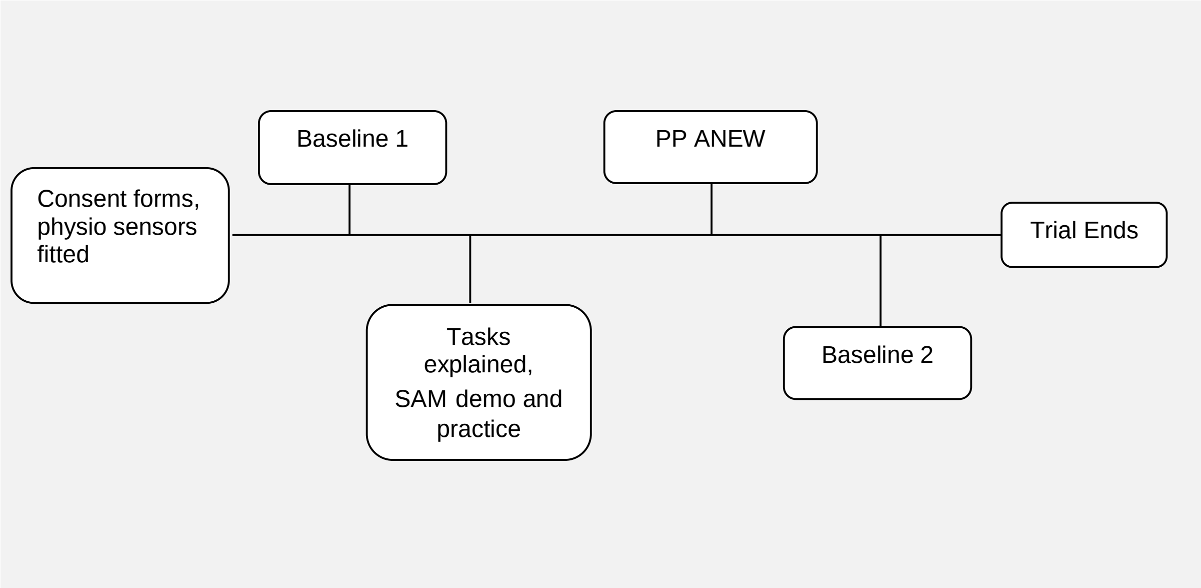


Figure 25. Trial order for study one

3.1.7 Results

Baseline readings were assessed for five minutes prior to test application with results summarised in Table 3.

Table 3. Summary of Physiological Baseline Measurements for 10 Coventry University Students

Heart Rate Mean (SD)	R-R Interval Mean (SD)	RMSSD Mean (SD)	LF/HF Ratio Mean (SD)	LF msecs^2 Mean (SD)	HF msecs^2 Mean (SD)
80.400 (9.318)	768.220 (85.089)	62.760 (57.355)	1.129 (0.974)	1872.117 (357.790)	2448.491 (539.781)

Distributions of HRV LF/HF ratio were similar for all groups, as assessed by visual inspection of a boxplot. HRV LF/HF ratios were statistically different between the different ANEW groups X^2 (2) = 19.51, p < .001. Post hoc analysis revealed statistically significant differences in HRV LF/HF ratio between neutral (mean rank = 46.27) and low (mean rank = 71.12) ANEW groups (p = 0.008), and neutral and high (mean rank = 86.61) ANEW groups (p <0.001). There was no difference between low and high ANEW groups. Figure 26 illustrates the graphical representation of the post hoc pairwise comparison with the mean rank score next to the grouped stimuli of High, Neutral and Low emotional words.

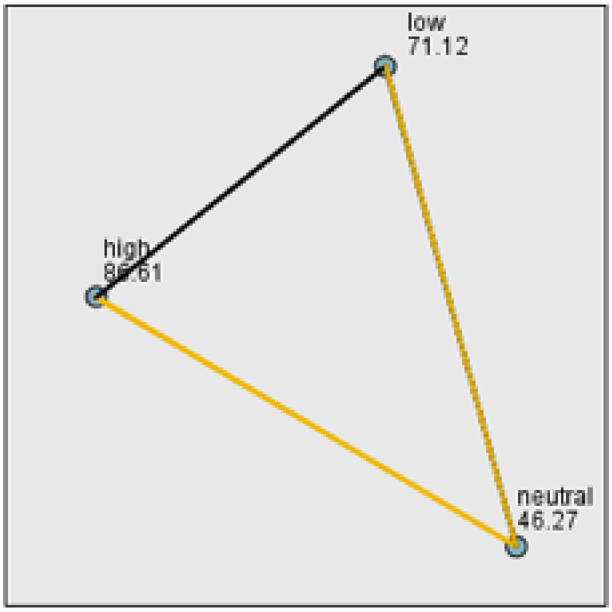


Figure 26. Post hoc, Pairwise comparison of HRV LF/HF data across ANEW groups (High, Neutral and Low). Significant differences found between the HRV LF/HF data and the different grouped stimuli are illustrated by yellow lines.

In addition to the LF/HF ratio, selected time and frequency domain measurements were analysed, although there were no other significant differences found between the different emotional grouped words, large differences compared to baseline readings were found in the LF power and HF power as summarised in Table 4.

Table 4. Summary of Physiological Measurements during presentation of ANEW stimuli

Condition	Heart Rate Mean (SD)	R-R Interval Mean (SD)	RMSSD Mean (SD)	LF/HF Ratio Mean (SD)	LF msecs^2 Mean (SD)	HF msecs^2 Mean (SD)
Baseline, pre stimulus	80.400 (9.318)	768.220 (85.089)	62.760 (57.355)	1.129 (0.974)	1872.117 (357.790)	2448.491 (539.781)
HIGH	84.320 (10.502)	734.520 (76.974)	47.579 (30.873)	5.231 (1.760)	735.977 (516.272)	688.241 (481.713)
NEUTRAL	82.163 (8.863)	741.000 (71.241)	56.794 (35.165)	2.013 (0.564)	1426.760 (571.631)	1019.023 (451.150)
LOW	83.462 (7.710)	738.064 (68.632)	54.013 (51.114)	3.664 (0.962)	2021.610 (556.880)	947.663 (450.633)

Data from the SAM assessment sheets were collected and processed. At this stage it was important to assess the validity of the SAM self-report scales that the participants completed against the validated ANEW ratings. Independent T-Tests were carried out and there were no significant differences across the three dimensions, demonstrating that the participants from this study were scoring the ANEW words in line with the validated ANEW scores.

A comparison of the mean HRV LF/HF ratio with the participants SAM self-report rating scale is shown in Figure 27. Results from observing the mean scores indicated that there was an association between the participants' physiological metric and their own rating of the ANEW.

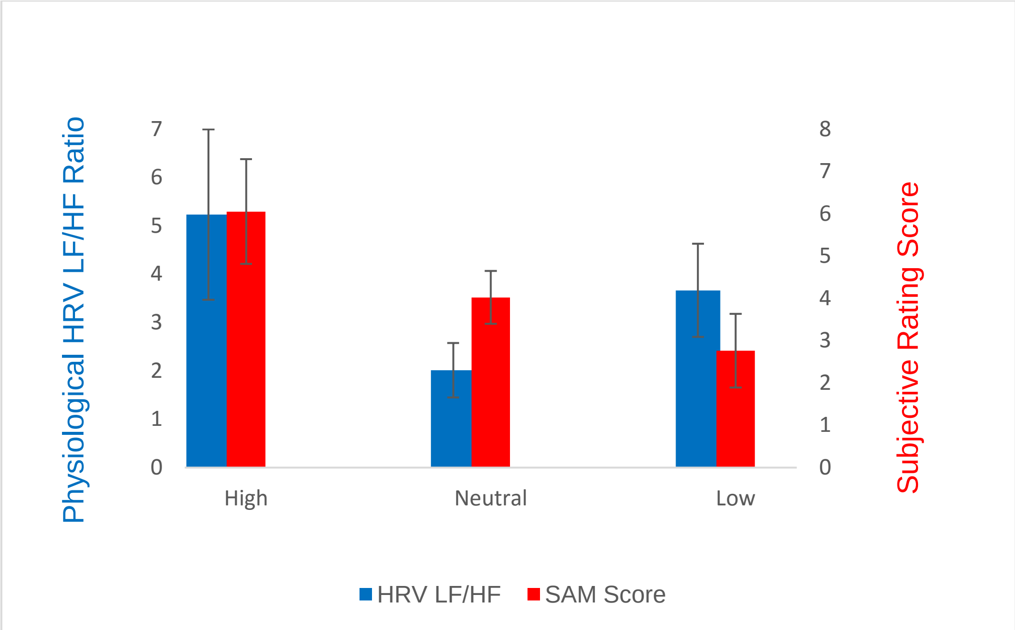


Figure 27. Comparison of HRV LF/HF physiological response to participant's self-assessment of the ANEW grouped emotional stimuli (High, Neutral and Low). LF/HF value is shown on the left axis with the self-assessment rating score (SAM) on the right axis. SAM score has a range between 1-9.

3.1.8 Discussion

By examining the differences between the three ANEW groups the results revealed that the total variance of a continuous series of heart beats, as determined by its frequency components (LF/HF), was able to detect changes during the presentation of the stimuli to the participants. The study also revealed that the participants SAM scores for the different grouped words were linked to their HRV metric. When the participants viewed words from the High emotional group, their HRV ratio increased, and at the same time the participants also rated these words as high against the three emotional dimensions of valence, arousal and dominance, in the SAM rating system. When neutral words were presented to the participants, their HRV ratio HRV metric was rating for the different a connection could be seen between the participants own subjective rating of the emotional words. Importantly, this was also reflected in their physiological responses.

The HRV metric used to assess the sympathovagal modulation was the LF/HF ratio. This measure was selected as it has been associated with arousal changes in different studies (McCraty *et al.* 1995, Riener, Frescha and Aly 2009). In McCraty's *et al*. (1995) emotional research the LF/HF ratio was linked with increases during both positive and negative emotions. For that research, the freeze-frame method was used to enact a positive emotion where the participants were instructed to disengage from unpleasant thoughts and focus their attention on their heart. This process promotes positive thoughts which can then be directed to thoughts of appreciation towards another individual or object (McCraty *et al.* 1995). The emotion of anger was produced by the participants recalling a situation that still provoked feelings of being frustrated or angry. Both emotional states were reflected by an increase in the LF/HF ratio, however, the greatest change was during the recall of being frustrated or angry.

In Study One, the focus was not on a specific discrete emotion, but rather in an arousal index. Participants LF/HF ratio increased when they viewed the emotional High grouped words compared to when they viewed the Neutral words. The words that were classified in the Low grouping also evoked a greater LF/HF ratio compared with the Neutral grouped words. This finding is consistent with other studies that indicate that cardiovascular effects will be seen with either pleasant or unpleasant stimuli compared to neutral stimuli (Bradley and Lang

2007). In this study the participants heart rate increased from their baseline and an observed difference between the different grouped words. Heart rate is known to increase when arousal levels are high, this is often seen when basic emotions such as anger and fear are evoked (Levenson 2011). However, when participants are shown images or film clips that are intensely unpleasant there is an initial deceleration of the heart rate rather than an increase. It was clear from the literature that emotions can be provoked by a range of stimuli that can be delivered in a laboratory setting. In Study One there was also a physiological response to the emotional words that the heart rate and HRV LF/HF ratio detected.

Although this study only used words in a laboratory situation, there has been research using this physiological metric to assess arousal levels in a ‘real’ driving environments, with this in mind it seemed an appropriate measure to assess in this first study. For example, the study from Riener, Frescha and Aly (2009) as discussed in the HRV section. In this study the time window for measuring LF/HF ratios were 60 second segments over the duration of the 20-30 minute journeys, whereas in study one a shorter selected time window of 30 seconds was used to capture the variation of the LF/HF ratios when the participants were presented with the emotional words. Although the parasympathetic activity occurs almost instantaneously, the sympathetic neural activity has a delayed onset of approximately three seconds before reaching a peak level, which can take up to 30 seconds (Franchini and Cowley 2012). By selecting 30 second segments, both parasympathetic and sympathetic neural activity would be reflected in the LF/HF ratio.

Notwithstanding the finding that HRV was indeed sensitive to detecting arousal fluctuation, there were several problems that arose whilst undertaking the study. The low number of participant's data that was able to be processed and the robustness of the physiological equipment were significant issues. Although stated as two problems, they were very much interlinked. The poor stability of participants signals was due a number of factors, but primarily stemmed from having a shared laboratory space, which was not conducive in gaining a clear, clean signal. The shared laboratory environment increased the noise artefacts and interference as well as disrupting the focus of the participants. The other major drawback from Study One was that the additional physiological measures of respiration rate and EDA were not able to be analysed due to their signal instability. In Study One the physiological monitoring equipment hardware were captured on separate devices, which produced a large crossover disturbance and poor signal pick-up. Although the running of this study highlighted certain issues, it allowed the methodology to be tested and reviewed for future studies.

Study One assessed whether the emotional stimuli (in this study, emotional words, ANEW) could provoke a physiological response, that the selected measures, heart rate and HRV LF/HF ratio could detect. By using the physiological indices, the participant responses could be quantified. It also demonstrated that the subjective rating scale showed a high degree of association between their rating of the emotional words and their physiological responses to the same stimuli.

Based on the promising use of HRV (LF/HF ratio) the focus for next study used images as well as words to evoke participant responses and compare the two data sets. Study One revealed significant differences in relation to the emotional saliency of the ANEW words, but, in an applied context (such as an automobile) the individual would be confronted more with visual information as opposed to written words. Indeed, the very nature of cognitive processing is very different between these two activities, with different encoding processes (see 1.2.6). The next study was designed to investigate whether the visual presentation of emotional imagery could afford a similar effect in terms of physiological response, and in addition to the emotional words and images being presented to the participants an emotional recognition task will be presented to the participants at the end of their word and image presentation.

Emotional studies have indicated that emotionally arousing images can be detected quicker than neutral images and that the intensity of the emotion can also have an effect on information processing, such as memory (Kensinger and Corkin 2003). Research has also examined negative emotions and the effect they may have on subsequent cognition tasks. Studies have revealed that following a brief exposure to negative stimulus tasks, impairments to subsequent tasks follow, these include the n-back task (Kensinger and Corkin 2003), short-term memory retrieval (Dolcos *et al.* 2006) and perceptual identification (Anderson 2005, Arnell, Killman and Fijavz 2007, Ihssen and Keil 2009). These findings may have an important relevance in the driving environment, if the driver if becoming frustrated with the interactions they are having with the HMI system, it may lead to further frustration with the system as they will begin to make additional errors, leading to disengagement with the system.

Although the sample size was small in Study One, there was some evidence to suggest that by presenting different emotional graded words to participants, the HRV metric was able to detect the physiological changes in the autonomic nervous system's activity when the participants viewed the different grouped emotional words.

4 Chapter Four – Study Two

4.1 Rationale and Introduction

As was seen in Study One the HRV metric was of sufficient sensitivity to detect emotional stimuli that was presented to the participants. The stimuli used in Study One was written words, whereas driving is a visual task and the nature of the interaction between the driver and the in-car systems is primarily visual. With this in mind, the second study involved an image task, where emotional images were presented to the participants to assess their response, in terms of a physiological response to the image task and secondly to assess their subjective response to the image task. Presentation of emotional words will also be included to compare against the image task as well as assessing repeatability from Study One.

Study Two would also only focus on the R-R data from the participants, with heart rate and HRV ratio being determined. The rationale for this was to ensure that the biosignal was as stable, with as few disturbances as possible. Although EDA and respiration measurements were planned as part of the methodology in Study One, due to various factors, as described earlier, no data from these signals could be analysed. Although this was of concern, empirical evidence is clear on the response that EDA demonstrates when presented with emotional stimuli (Boucsein 2012), and therefore at this stage it was not central to Study Two. However, what was important for Study Two was to further assess the validity and repeatability of HRV with the presented emotional stimuli. As discussed earlier (2.1.4.2) there was no 'gold' standard for HRV assessment (Khandoker *et* *al.* 2013), and therefore it was important to consolidate the findings from Study One, especially since participant numbers were small.

In Study Two participants were assessed on their responses when they were presented with words and images. The words being used for this study were from the validated ANEW directory that was used in Study One. In addition to the stimuli of words participants were also presented with images from the validated International Affective Picture System (IAPS) directory. As this project followed a 'bottom-up approach', it was also another step taken to the final study that would involve the participants interacting with a visual interface. In order to assess other aspects of human interaction, participant's recognition and response times to the different emotional stimuli were also recorded. This would allow for the evaluation of the cognitive processing response to the different presentations of emotional stimuli.

4.1.1 Study Two Hypothesis

H_1: The use of a HRV metric would detect participant's responses to presented visual emotional stimuli

H_0: There would be no significant changes in the HRV metric detected, in participant's response when presented with visual emotional stimuli.

4.1.2 Methodology

In contrast to Study One, this study was more controlled and was based in a dedicated laboratory, thus allowing the use of the HRV metric (LF/HF) in isolation. By adopting this approach, it was believed that this would limit the number of extraneous variables and afford a more likely prospect of obtaining robust physiological signals. To this end it was decided that by running only one biosignal, the signal would be cleaner, have less artefacts and therefore an anticipated larger data set could be collected.

Ethical approval was granted from Coventry University Ethics Committee for Study Two (Ethics P13879)

4.1.3 Design

The study was a within-subjects design with each participants viewing emotional stimuli comprised of both words and images. Participant responses to the different stimuli were examined along with the subjective self-report rating scale.

4.1.4 Participants

To recruit participants for the second study an email was sent by the researcher's Director of Study to all Health and Life Science staff and post-graduate students. On receiving an email from an interested individual the researcher sent, by email, a Participant Information Sheet (PIS) and the respondent was encouraged to email any questions or concerns that they may have regarding the study. The PIS informed the prospective participant about certain health conditions that would exclude them from the study as well as advising the participants about the requirements for attending the study, this included information regarding refraining from caffeine drinks for a four-hour period before the study. If an individual wished to volunteer for the study an appointment was arranged that was convenient to them. The recruitment data capture flow chart is illustrated in Figure 28.

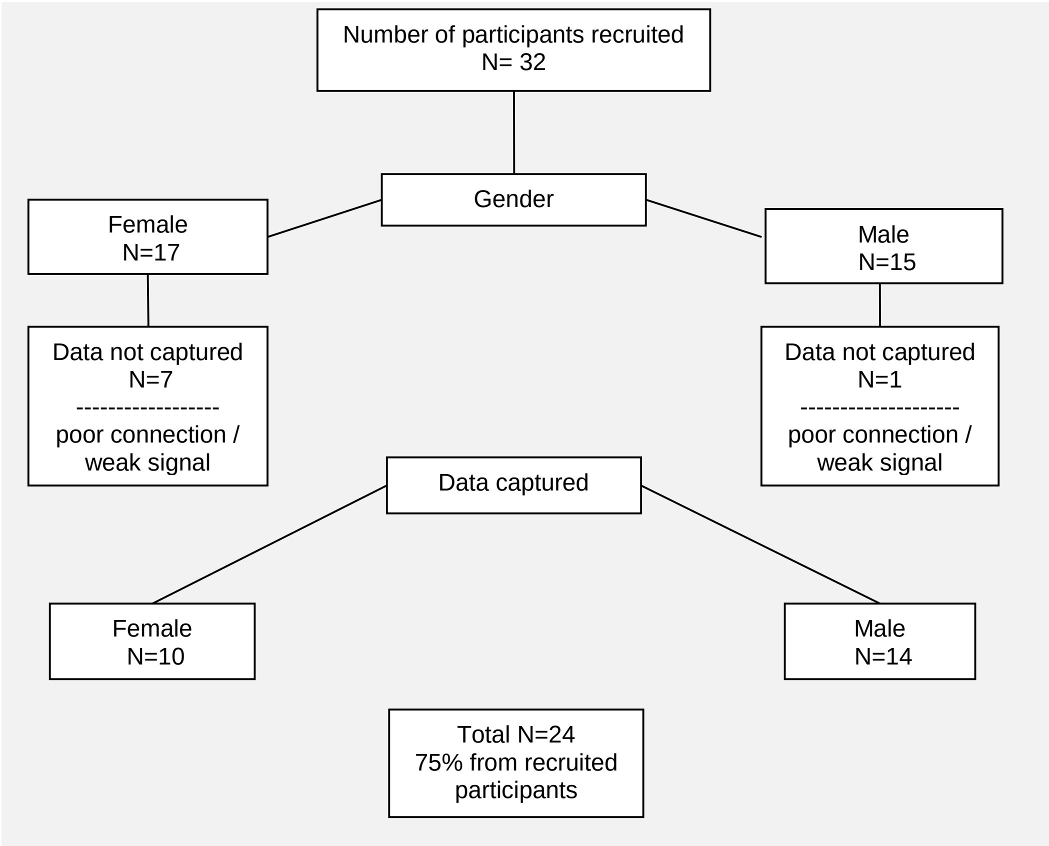


Figure 28. Study two recruitment flow chart

Data was collected from 24 Coventry University staff and students. The following table represents the number of participants in the study with their gender and age details.

Table 5. Number of Participants, Gender and Age details

Participants	Gender	Age (SD)
24	10 Female 14 Male	38 (13.410)

4.1.5 Materials and Equipment

The images used were from the International Affective Picture System (IAPS) from the University of Florida (Lang, Bradley and Cuthbert 2008). To obtain the catalogue of images a request was made from the researcher's Director of Studies directly to Lang and Bradley. The catalogue of images was then emailed to the Director of Studies and forwarded to the researcher. There are over 1000 images in the IAPS catalogue and similar to the ANEW system they have been rated over the three affective dimensions of valence, arousal and dominance using the SAM rating scale. As per Study Two, the Affective Norms English Words (ANEW) stimuli were used, which allowed the selection of stimuli that had also been rated over the same three emotional dimensions (Bradley and Lang 1997; 2008). The words and images were then grouped into three categories depending on their rating. The three categories were termed High, Neutral and Low. A representation of three images in their categories is displayed in Figure 29.

This material has been removed from this thesis due to Third Party Copyright. The unabridged version of the thesis can be viewed at the Lanchester Library, Coventry University.

Figure 29. Hand drawings representing images from the IAPS catalogue and rated against the emotional dimensions of valence, arousal and dominance. Image (a) is an illustration of IAPS 5263 from the High emotional group,(windsurfers), image (b) represents IAPS 7041 from the Neutral group,(baskets) and image (c) represents IAPS 9560 from the Low grouping (oiled seabird), (Furness 2016).

The IAPS catalogue number for all images used in this study are located in Appendix 8. An example of the ANEW words assigned to the classified emotional groups are illustrated in Table 6.

Table 6. Example of ANEW words assigned to classified groups

High Group	Neutral Group	Low Group
494 WIN (mean 7.83)	866 MARKET (mean 5.02)	260 LONELINESS (mean 2.89)
ANEW 494 (win), ANEW 866 (market) and ANEW 260 (Loneliness): validated mean ratings across emotional dimensions of arousal, valence and dominance (Lang and Bradley 1997)		

The Self-Assessment Manikin (SAM) used in Study One was also adopted for this study. A full description and illustration of the self-report rating scale was discussed in Chapter 3 (3.1.5).

The physiological monitoring equipment required for this study was a heart rate monitoring device, identical to the device used in study one (see Chapter 3, 3.1.5). As this study focussed on the HRV metric, no other physiological equipment was used. In the previous study, there had been an issue with the transmitter picking up the signal from the heart, with many participants' data unable to be processed. Therefore, for this study an electrode gel was applied to the sensors on the chest band, with the aim of improving the quality of the signal.

A projector, with an accompanying screen was utilised for this study to display the words and images. At each side of the screen was an A4 poster, to remind participants about the correct key they should be using in the recognition task. A keyboard was placed on the desk, in front of the participants, during the recognition task. The same process for analysing the data was performed in Study One (Chapter 3, 3.1.5). Presentation of the recognition task was through PsychoPy, which was used in Study One and full description discussed in Chapter 3 (3.1.5). is an open-source application for running studies in Python (Peirce 2009).

All recorded data was either inputted onto Excel sheets or directly into SPSS (version 22) for analysis. Data was first assessed by a Shapiro-Wilk's test to determine if the recorded data was normally distributed. For normally distributed data one-way ANOVAs and Post Hoc tests were applied, where the data was not normally distributed a Kruskal-Wallis test was applied.

4.1.6 Procedural Approach

A quiet laboratory space was acquired for this study where participants were invited to sit behind a desk for the duration of the study. The emotional stimuli were projected in front of them on a 90 x 90 cm screen, using a PowerPoint presentation. Participants completed an informed consent form and a short medical questionnaire, relating to any medications they may have been receiving, or devices that they had internally which could impact on the biosignal.

Participants were then given a demonstration of the correct placement of the heart rate monitor, and if required assistance was given. Participants then positioned themselves behind the desk, where the researcher also sat for a short period of time to give participants instructions regarding the tasks ahead. Once the participants were happy with the format of the study, the researcher sat a short distance away from the participant behind a desk, which had two personal computers for delivery of the tasks. At the end of the instruction period the researcher checked that the wrist mounted heart rate transceiver was continuing to capture the biosignal.

The lights in the room were switched off, and a five minute ‘plain vanilla’ baseline measurement was recorded prior to any study stimulus being presented. A ‘plain vanilla’ baseline was selected for this study, as it has been suggested that by engaging the participants minimally it may produce a more stable baseline than a traditional ‘resting’ baseline (Jennings *et al.* 1992). An aquatic image clip was selected for this purpose (as per Piferi *et al.* 2000). A 'resting' baseline can be problematic, due to differences in mood and emotional states that the participants have on entering the laboratory, which can be reflected in their baseline recording. In a traditional 'resting' baseline, participants are asked to close their eyes and relax, which may be an unrealistic request, as participants are often in an unfamiliar environment, with physiological equipment attached, and in addition to these factors they may not have met the researcher before.

An additional two baseline measurements were recorded during the study. The second baseline was recorded after the first stimuli set was completed but before the presentation of the second stimuli set started. The justification for this was to ensure that the participants started the second presentation in a similar emotional level as at the start of the first presentation. A final baseline was presented to the participants at the end of the study, this baseline was presented to the participants as it is 'good practice' that the participants leave the laboratory in a similar emotional state as when they entered, rather than a stimulated state facilitated by the experiment (Krahe *et al*. 2011). A representation of the laboratory layout is illustrated in Figure 30.

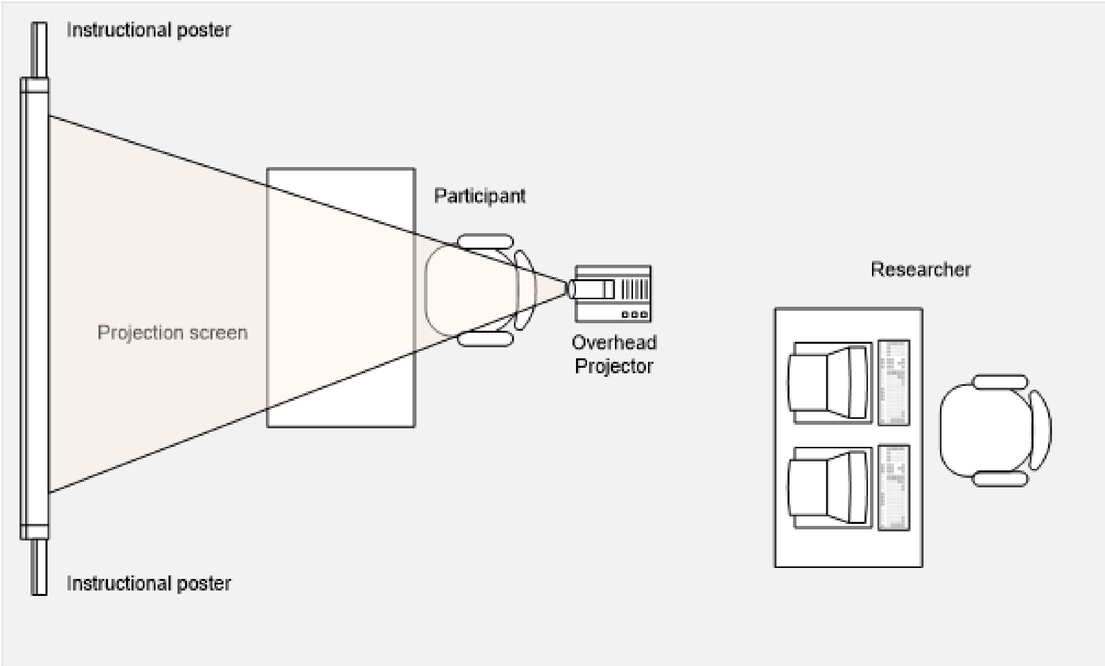


Figure 30. Laboratory layout for study two

Before the start of the trial, each participant was permitted a practice ‘run’, where they were presented with an image or word (not part of the stimuli set to be tested) and had to complete the SAM rating sheet. If the participants had any difficulties with this task the researcher was able to ‘walk’ them through the procedure. Instructions were also displayed on the screen for each participant to read. Each participant viewed 21 words and 21 images which were randomly delivered to counterbalance any order of effects. Each word and image was displayed for a period of six seconds. During this time the participants had been instructed to concentrate on the word or image and remain focussed on the presentation for the full length of time. Only when the word or image was removed from the screen could the participants complete the SAM rating sheet. Participants had 15 seconds to complete the self-assessment scale before a normalising object (in this instance a single dot) appeared on the screen. This inter-stimulus interval lasted for 20 seconds, which allowed the previous response to dissipate before the next word or image was presented (Lang and Bradley 1999).

Concurrent physiological measurements were used to assess any modulations when the participants were presented with different emotive words and images. After the final word or image was presented, the second task commenced, which was a recognition task. In addition to the 21 words or images that the participants had already seen, a further 39 unseen images or words were presented. The additional words or images were evenly acquired from the three emotional categories; High, Neutral and Low. The total number of pictures presented in the image task was sixty, which has been used in a study assessing emotional picture processing, in this study the images also came from the IAPS catalogue (Bublatzky *et al.* 2010). In Study Two the recognition task assessed the participant's reaction time to the different images and words as well as their error rate.

The participants were instructed to use a keyboard and select the 'right arrow' to indicate that they had already viewed the word or image in the previous task. If they had not viewed the word or image in the previous task they selected the 'left arrow'. They were instructed to complete this task as quickly as they could. Signposting on the screen was also provided relating to these instructions. The 60 words and images were randomly presented to the participants, through PsychoPY (Pierce 2009). Figure 31 illustrates the order in which the tasks were completed during the study. The duration of the study from the participant entering the laboratory to leaving it was approximately one hour and 15 minutes.

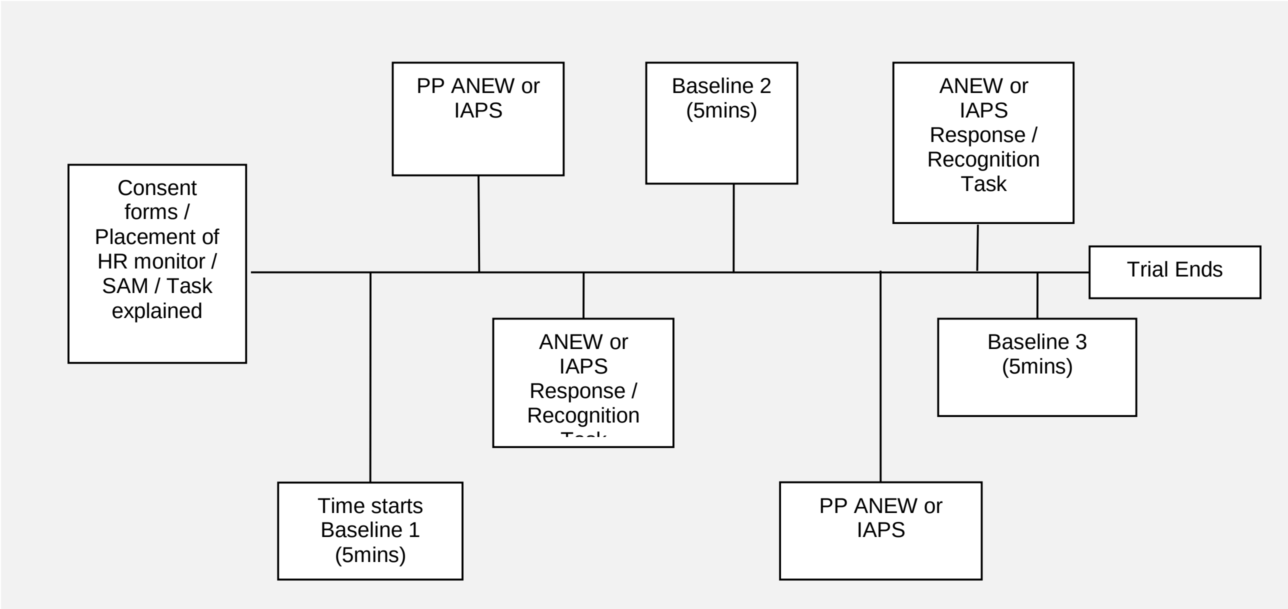


Figure 31. Task order for study two; tasks from left to right:
Participant completed documentation and had HR monitoring device fitted, aquatic baseline for 5 minutes, then presentation of emotional words or images (dependent on study matrix), presentation of 60 words or images for recognition task, baseline 2 (as before) then sequence repeated using different emotional stimuli, completed study with baseline 3 (as 1 and 2).

4.1.7 Results

The average baseline measurements for the participants of study two are detailed in Table 7.

Table 7. Participants average baseline measurements for HR, HRV LF/HF ratio and R-R intervals

Participants	Baseline HR (bpm) (SD)	Baseline LF/HF Ratio (SD)	Mean RR (ms) (SD)
24	68.031 (11.150)	1.220 (0.770)	901.21 (152.11)

A Kruskal-Wallis test was applied to the heart rate and HRV LF/HF ratio data to determine if there were differences in the heart rate and LF/HF ratio between the emotional grouping of words (ANEW) and images (IAPS).

No significant differences were found in the heart rate data between the different groupings of words and images ($p > 0.05$) but significant differences were found in the HRV LF/HF ratio. A Post hoc Pairwise comparison was applied using Dunn's (1964) procedure with a Bonferroni correction for multiple comparisons. LF/HF ratio was statistically significantly different between the emotional groupings of words and images. IAPS (Images), $X^2(2) = 32.160$, $p < 0.001$. Post-hoc analysis revealed statistically significant differences in LF/HF ratio between the Neutral (mean rank = 18.35) and High (mean rank = 49.42) ($p < 0.001$) and Neutral and Low (mean rank = 41.73) ($p < 0.001$). No significant differences were found between high and low IAPS groups, Figure 32.

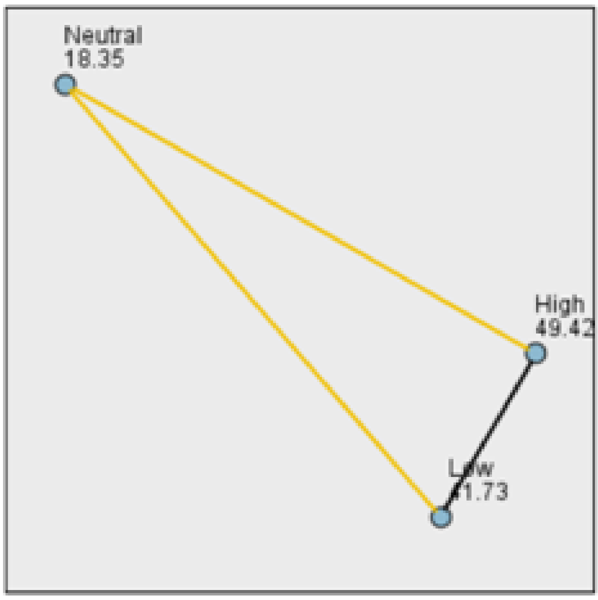


Figure 32. Post hoc, Pairwise comparison of HRV LF/HF data across image (IAPS) groups (High, Neutral and Low). Significant differences found between the HRV LF/HF data and the different grouped images as illustrated by yellow lines, mean rank values are stated next to the emotional groupings.

Findings also revealed that there were differences within the ANEW grouped words. As above a Kuskal-Wallis test was performed; ANEW (Words), $X^2(2) =$

26.685, $p < 0.001$. Post-hoc analysis revealed statistically significant differences in LF/HF ratio between the Neutral (mean rank = 16.79) and High (mean rank = 47.81) ($p < 0.001$) and Neutral and Low (mean rank = 44.90) ($p < 0.001$) but not between the High and Low ANEW groups, Figure 33.

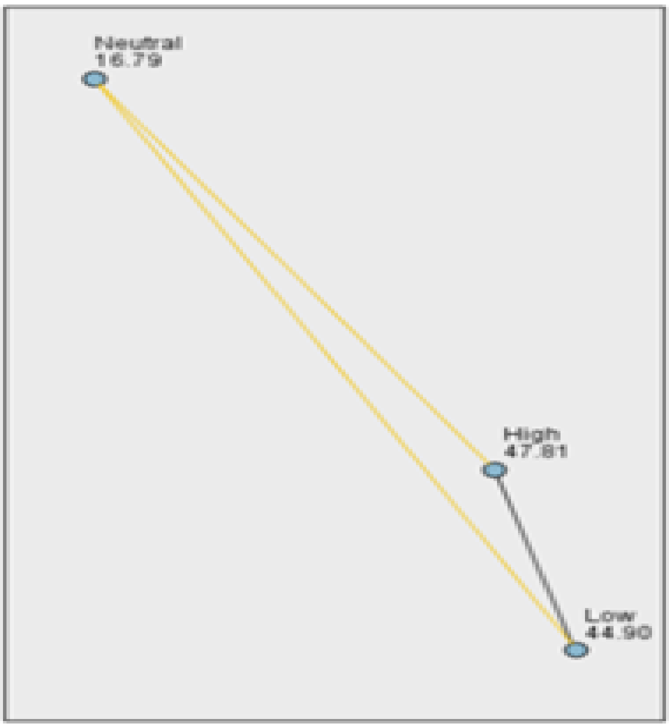


Figure 33. Post hoc, Pairwise comparison of HRV LF/HF data across word (ANEW) groups (High, Neutral and Low). Significant differences were found between the HRV LF/HF data and the different grouped words as illustrated by yellow lines, mean rank values are stated next to the emotional grouped words.

At the end of each presentation of the ANEW and IAPS stimuli, a recognition task was conducted with 'seen' and 'unseen' images and words. Participants responded as quickly as possible and inputted whether they had seen the image or word on the previous task. Response times were recorded and a one-way ANOVA was conducted, results revealed that there were significant differences in the response times to the differently grouped stimuli, see Figure 34. $F(5,138) = 3.720$, $p < .005$. A post hoc pairwise comparison test was conducted to determine which groups of stimuli produced significantly different reaction times when the participants responded to them. Figure 33 displays the mean reaction time of the participants and differences in groups were between ANEW Neutral and ANEW Low; IAPS Low and ANEW Neutral and IAPS High and ANEW Neutral.

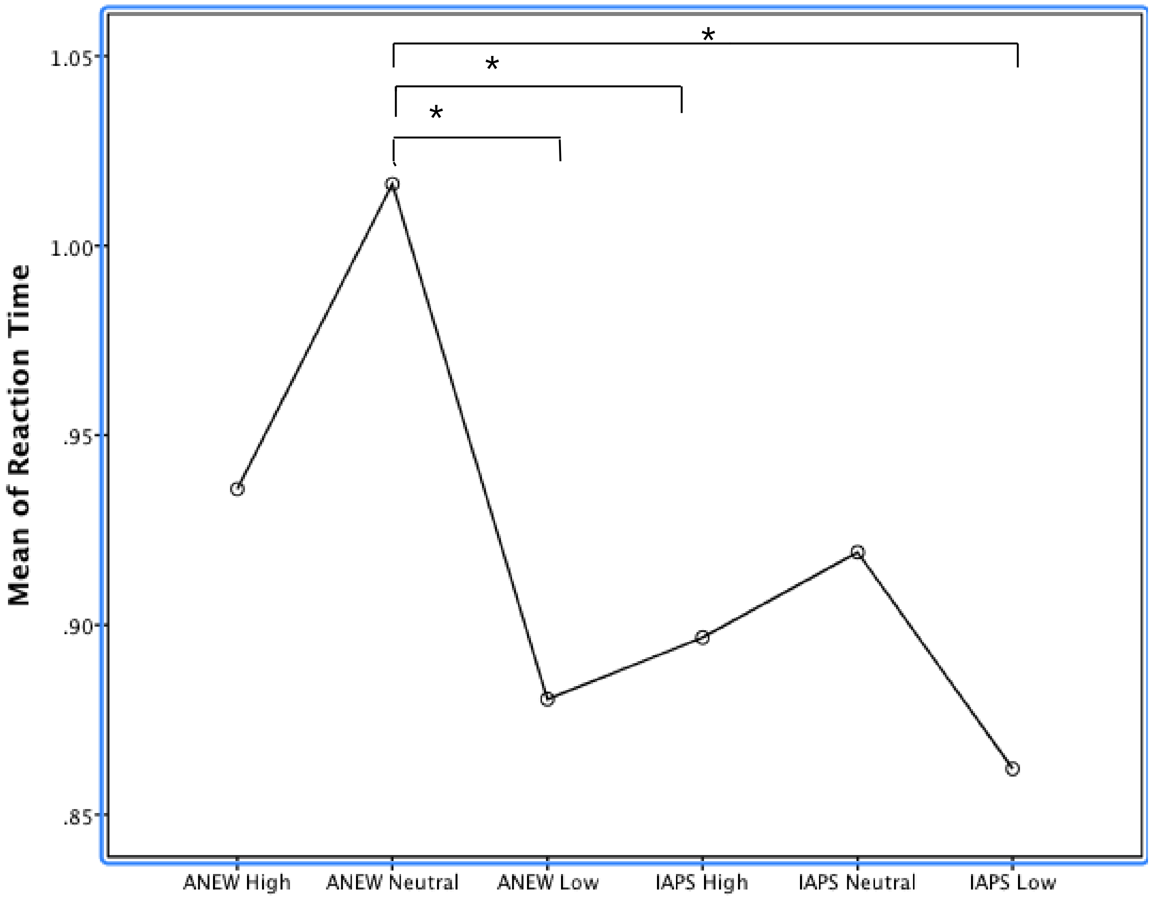


Figure 34. Participant recognition response times to different emotional stimuli. Participants were presented with 60 images and words, 21 were 'seen' and the remaining were 'unseen'. Participants responded as quickly as possible and selected the correct identification key to each word and image. *p<0.05

Error rates were also recorded, participants responding to the pictures correctly identified them with a 98.85% correct score. Participants responding to the words (ANEW) had a 92.67% correct response rate. A comparison of the HRV metric and the participants own subjective rating of the ANEW words and IAPS images, using the SAM were also assessed to determine if the participant's physiological measurement was related to their attached emotional feelings when presented with the stimuli, Figure 35. Correlation tests were also conducted on the SAM self-report scales and the HRV metric. There was a moderate positive correlation between the self-report SAM scale of the ANEW High grouped words and the HRV measure, r = .409; there was a moderate negative correlation between the SAM self-report scales of the ANEW Neutral grouped words and the HRV measure, r = .380. A negative moderate correlation was identified between the SAM scales of the IAPS High pictures and the HRV measure, r = -.305.

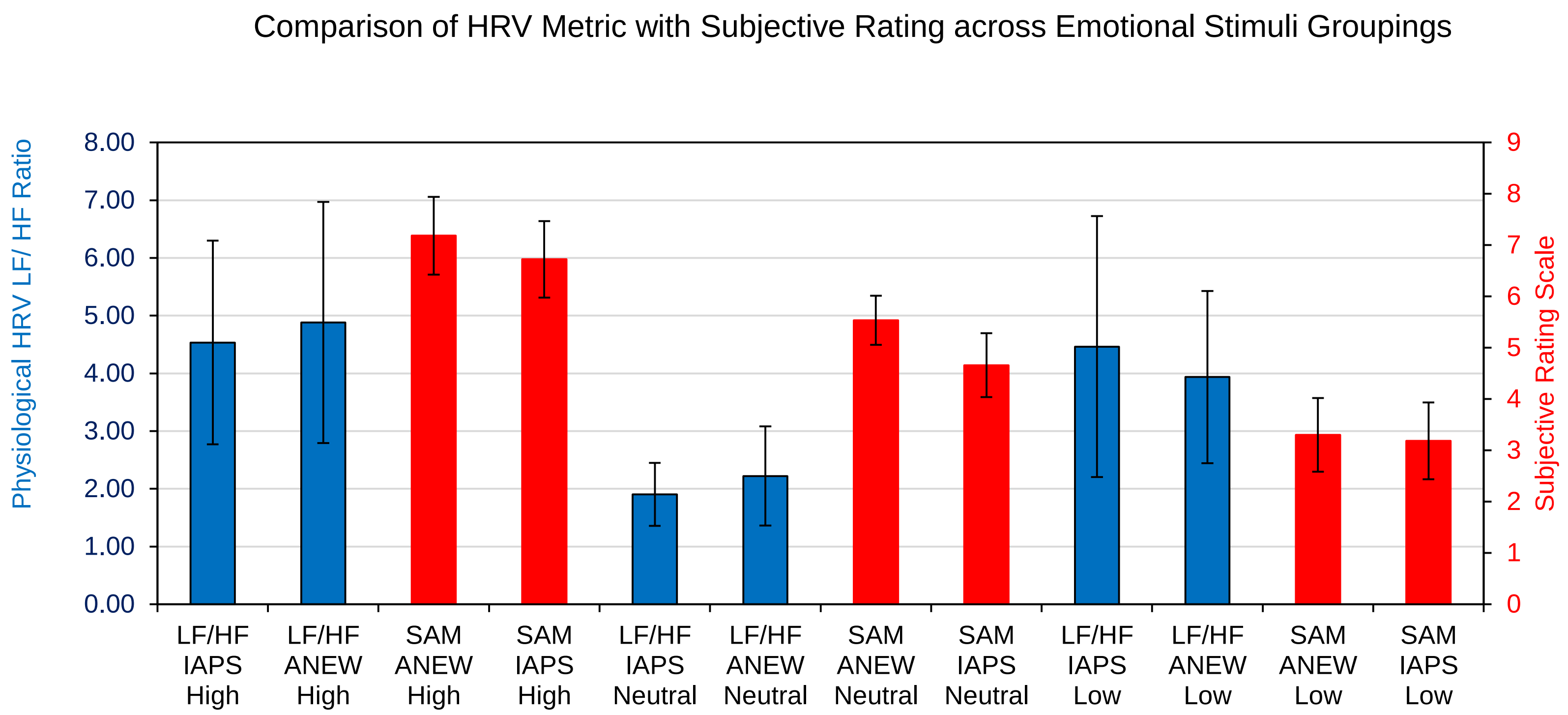


Figure 35. Comparison of HRV LF/HF ratio to different grouped emotional stimuli and mapped with participants' self-assessment value of words (ANEW) and images (IAPS). Each emotional stimulus (word or image) was grouped into High, Neutral and Low categories, according to their rating score over the emotional dimensions of valence, arousal and dominance. The HRV value is shown on the left axis (blue), whereas the self-assessment (SAM) value is shown on the right axis (red). SAM score has a range between 1-9.

4.1.8 Discussion

The data analysed from this study has indicated that the HRV frequency domain measurement (LF/HF ratio), can be used as an indicator for autonomic balance when participants were presented with emotional stimuli. In this study the stimuli applied were words (ANEW and images (IAPS) which possessed validated ratings over the three emotional dimensions of valence, arousal and dominance (Bradley and Lang 1997). The stimuli were then grouped into three categories depending on their mean subjective rating; namely High, Neutral and Low.

As discussed previously the LF/HF ratio is a HRV measurement used to assess sympathovagal autonomic balance. High values are thought to indicate the dominance of sympathetic activity with low values indicating the dominance of parasympathetic activity (Khandoker *et al.* 2013). The results indicate that when either a word or image from the High and Low groupings was presented to the participants their autonomic nervous system response was reflected by a significant change in their LF/HF ratio values. This reaction resulted in a rise in the ratio compared to words and images from the Neutral grouping and the baseline measurements taken prior to the tasks.

When examining which stimulus evoked a greater physiological response, although the difference was not significant, in both High and Neutral categories the ANEW system had produced a higher LF/HF ratio value than the IAPS system. This was also observed by the participants own self-assessment of the words and images. In the Low category the IAPS system produced a greater physiological response, with the LF/HF value being higher than the value when the ANEW system was presented. Due to the large inter-individual variation in the HRV measurement no significant differences could be identified between the different presentation of stimuli and physiological measurement.

The Self-Assessment Manikin (SAM) across the three emotional dimensions were higher for the ANEW categories of High and Neutral compared to the IAPS categories of High and Neutral. The images, however in the Low category were scored lower indicating that the nature of the pictures had a slightly more negative effect on the participants, no significant differences were found.

One of the key aspects of the study was to assess the results of the HRV measure parallel with the SAM results to find out whether this physiological metric could give validation to the subjective responses from the participants. The findings illustrate that the scores the participants gave to each word and image over the three emotional dimensions corresponded to their HRV metric. This would suggest that HRV does indeed reflect not only physiological response to emotionally-charged stimuli, but also the perceived emotional experience as well. A recent study by Williams *et al.* 2015 investigated how resting HRV may be associated with perceived everyday emotional difficulties. Participants completed different questionnaires relating to emotion regulation, rumination and anxiety, and results revealed that a lower resting HRV was associated with perceived greater difficulties in emotion regulation, anxiety levels and perceived ruminative tendencies. This study supports other work by Thayer and Lane (2000) that recognises the association between HRV and emotional regulation. Studies by Quintana *et al.* 2012 also found that higher levels of anxiety predicted decreases in HF HRV. In 2000 Thayer and Lane introduced the Neurovisceral

Integration Model, and proposed that the autonomic, attentional and affective systems were integrated by neural structures that would assist in understanding emotion regulation and dysregulation. Evidence supports that HRV not only provides information relating to the heart, but importantly has the potential to relate information regarding different structures of the brain that are closely linked to emotional and cognitive processes (Thayer and Lane 2000).

Previous findings in this area would have suggested that images may have provoked a greater physiological response than words (Holmes *et al.* 2008). However, in this thesis this was not been witnessed except in the Low category. Reasons for this may include unequal distribution of emotionality across the dimensions; where a High categorised image rated high in the valence dimension but not as high in the arousal dimension. This may also have relevance in the LF/HF ratio value of the Low categorised images. These images evoked a greater response than the words resulting in a higher LF/HF ratio value. It is worth noting that some of the images in the category may be classed as being highly emotionally graphic and scored low on the valence and dominance scale, participants may have found that they had a level of arousal when viewing these images compared to similar categorised words. This effect may have a strong physiological bases, which has been suggested by Kensinger and Schacter (2006), in which their study investigated the different processing areas of the brain for effects of valence and arousal from pictures and words. One of their findings was that an area of the brain, the prefrontal cortex (PFC) responded more to negative pictures than words.

Another consideration, this time relating to cultural differences, was highlighted during the trial debrief. A participant commented that two of the Neutral words viewed had very specific meanings to this individual and for both words the participant had rated them very high across all three emotional dimensions. This perhaps highlighted an issue of personal interpretation and also the question arises if this may be more profound in one stimulus than another. When examining the SAM scores the word 'Market' frequently scored higher over the three dimensions than the validated score. As ANEW was developed in the United States of America (USA) it may be worth considering if cultural differences have affected the rating of this word. A study by Levenson, Soto and Pole (2007) substantiates these findings, as they reported that subjective self-report scales were very susceptible to cultural influence when an emotional exposure had been presented.

After debriefing the participant who had commented on the anomalies for the Neutral words in the SAM rating scale, it was interesting to ascertain whether their LF/HF ratio had reflected this. On both occasions the participant's LF/HF ratio was considerably higher for 'Market' than for the rest of the Neutral words viewed by the participant. This physiological response validates that the participant was processing this word in a different way to other participants who had rated this word as a neutral word, and their physiological response reflecting the non-arousing aspect of the word. Interestingly this may reflect a physiological basis in that the response is a consequence of learning through exposure to specific cultural socialisation. Other findings have suggested a difference in both cognitive and visual processing based on cultural background (Choi, Choi and

Norenzavan 2004, Boduroglu, Shah and Nisbett 2010, Kuwabara and Smith 2016). We can begin to theorise that the individuals mental model is not only formed based on familiarity and experience of the stimuli they come into contact with, but also this may have a learned effect that is reflected in physiological mechanisms related to cognitive processing.

The other task that was conducted in Study Two was to assess the participants' recognition response times and error rates when unseen emotional stimuli were presented along with seen emotional stimuli. As Figure 34 shows the negative (Low) stimuli was recognised faster than the neutral stimuli, regardless of stimuli set. This finding has been consistent in emotion research, where the attention-based theory suggests that emotional stimuli are identified quicker as they are selected before the non-emotional content (Frischen, Eastwood and Smilek 2008) and therefore an emotion will be remembered better if the intensity of the emotion is increased (Johnston and Olson 2015, Kensinger and Corkin 2003). Research has also examined negative emotions and the effect they may have on subsequent cognition tasks. Studies have revealed that following a brief exposure to negative stimulus tasks, impairments to subsequent tasks follow, these include the n-back task (Kensinger and Corkin 2003), short-term memory retrieval (Dolcos *et al.* 2006) and perceptual identification (Anderson 2005, Arnell, Killman and Fijavz 2007, Ihssen and Keil 2009). These findings may have an important relevance in the driving environment, if the driver if becoming frustrated with the interactions they are having with the HMI system it may lead to further frustration with the system as they will begin to make additional errors, leading to disengagement with the system.

The present study also revealed that the unseen images received a marginally quicker response time to the seen images. Error rates were low in the study, with the participants responding correctly to the seen and unseen images, participants score was 98.85%, whereas the score for the ANEW words was lower at 92.67%.

As stated in the recruitment flow chart (see Figure 28) eight participants' data was unable to be utilised due to connection and signal problems. Although, in this study electrode gel had been applied to the HR monitoring chest band for every participant, with the aim of improving the signal, there was still a sufficient number of 'lost' data sets that gave concern. From interacting with the participants, it was clear that some of the individuals had difficulty fitting the HR chest band and although the researcher offered assistance, due to the participant's privacy it was not always possible to check the position of the chest band.

The researcher was also aware that the female participants struggled more with the positioning of the chest band and this was reflected in the higher number of poor data sets from the female participants. Although this was not mirrored in Study One, (Chapter 3) it may have been accounted for as the participants in the previous study were undergraduate students and therefore younger than the female participants in the current study, who were mainly Coventry University staff. The younger students in Study One appeared to be as comfortable as the male participants in fitting the chest band, whereas some of the female participants in the second study had no experience of using a HR chest monitor and were unfamiliar with the equipment.

There are some clear advantages of using physiological equipment that can record data visually in ‘real time’, one being that you are always able to see the signal from the start of the experiment. If there are any problems encountered, you can choose to stop the study if required and make adjustments if needed. Another advantage is that you can instantly see how the participant is reacting to the stimulus and record notes alongside the data as events occur, to assist in the later analysis. For these reasons, and the experience gained from Study One and Study Two, an electrocardiograph was obtained for the final study. By using this equipment, the strength of the heart signal should be improved which will lead to a ‘cleaner’ signal and improve the processing and analysis. Another important advantage of using a different recording system was that several biosignals could be recorded simultaneously and fed into the different channels of the equipment. By using the same equipment for different signals would eliminate the crossover disturbance that occurred in Study One.

In conclusion, Study Two has validated Study One results, with the findings that the HRV metric was sensitive enough to detect modulations of the sympathovagal balance when participants were exposed to two different sets of emotional stimuli. In addition, the participants also rated their feelings when exposed to the stimuli and their subjective ratings had a measure of association with the LF/HF ratio.

In the final study additional physiological measures will be used alongside HRV metrics, which will provide an opportunity to compare HRV with accepted and traditional physiological stress assessment techniques such as electrodermal activity and respiration. As described in Study One and Study Two, HRV has shown a high level of sensitivity to participants’ responses when presented with different emotional stimuli, and it will be of interest to compare this level of sensitivity with electrodermal activity during the same study.

Another layer of the study was to run a recognition task, which brought in elements of cognitive and attentional processing, which are two key elements that are required in a driving environment.

5 Chapter Five

5.1 Rationale and Introduction

The process of In-vehicle interface evaluation often relies on data collected from subjective responses obtained in user trials (see 1.2.7). By their very nature this information is often regarded with scepticism and thought unreliable (Jung and Willumeit 2000). Therefore, the possibility of applying an objective physiological measure to run parallel with the subjective measure would increase confidence in this data and allow for analysis to extract meaningful data that could be part of the evaluation process.

Heart rate variability, as a psychophysiological measure has the potential to assess and quantify the dynamic interactions between physiological requirements, mental, emotional and behavioural processes. As emotional states affect the activity of the ANS it would be reasonable to select physiological measures that are in some part, either controlled or influenced by the ANS. Although only HRV and heart rate were extracted and analysed from the R-R interval data in studies one and two, it would be advisable, due to the complex nature of the final study to have a range of physiological measurements that were linked to the ANS and therefore be compared to each other. The physiological indices for Study Three were HRV, HR, EDA and Respiration (see Chapter 2 for detailed information).

Before relating HRV analysis of affective responses to in-vehicle interface systems it was important to assess the sensitivity of HRV measures to emotional responses. By using a 'bottom-up' approach, the capability of heart rate variability to detect emotional responses was first assessed in Study One and Study Two before adding an in-vehicle interface system into the study design.

This was the third study in a series exploring the use of physiological metrics to validate self-report scales when they are used to capture subjective responses in the evaluation process of HMI systems. The first study focussed on methodology issues whereas the second study isolated one of the physiological measures, heart rate variability to determine if HRV could be sensitive to emotional images and words when presented to participants. From the previous encouraging results, the third study was designed to assess participants' responses when they interact with two different automotive touch screen graphical user interface systems (GUIs). Although study three was designed as a bench-top experiment, it was important to factor-in additional features in the experiment that would replicate some of the processes used when driving. A primary visual tracking task and a tertiary task were added. By adding these additional tasks, the participant's interactions and user experience with the HMIs would be influenced by them.

The aim of Study Three was to employ both objective and subjective metrics to assess user engagement during interactions with two different presentations of an automotive infotainment HMI. By examining the physiological and self-report scales, a greater understanding of users' responses would be determined and such information may provide important guidance within the early stages of in-vehicle design evaluation in terms of usability and user satisfaction.

In addition to comparing the participant responses to two different HMI presentations, and additional smaller sample of participants from an automotive industry research division would also be conducted. This would be a different dimension to the study and worthy of exploring the responses of a group of individuals who regularly participate in user trials and have direct industrial experience with HMI research, design and development. A comparison could then be interpreted between the different grouped participants.

By using two different interfaces which had the same task set-up it was hoped that the participants' level of engagement and usability experience would be different between the two interfaces. If this was indeed the case, the significant aesthetic and usability difference between the two HMIs (sharing identical tasks) would provoke significant physiological differences when interacting with the HMIs. This will form part of the mapping between physiological and psychological state in terms of usability for different HMIs.

While Study One and Study Two focussed on the different types of emotional stimuli, which were categorised over the dimensions of valence and arousal, the third study will focus on the discrete emotions that an individual may have when interacting with a HMI system, relating to user experience and user engagement.

Study Three, will therefore connect all the different components that have been discussed earlier and utilised in the first two studies.

With a clear understanding that cognition and emotion are closely interconnected and that emotions play an important role in a user experience. It will be of interest to assess these different elements in Study Three and investigate how they will impact on the perception that an individual forms regarding different representative automotive HMI systems.

Paper accepted at the Human Factors Conference 2015:
Exploring Heart Variability for Automotive Human Interface Evaluation
(Furness 2015)

5.1.1 Study Three Hypothesis

H_1 The user experience of participants will be affected with the interactions of a 'poor' HMI compared with a 'good' HMI and detected by physiological and subjective scales

H_0 Interactions with a 'poor' interface will not affect the participants' user experience

5.1.2 Methodology

The third study adopted a similar research philosophy to the previous studies one and two. Ethical approval was granted for this study by Coventry University Ethics Committee (Ethics P23519).

5.1.3 Design

The study design was a mixed design approach. Participants were grouped depending on the HMI allocation and task allocation. A between groups approach was then taken to compare the differences between the groups to the different

HMI and task allocation. A within subjects approach was also taken to assess how the different tasks affected the participants during the study. A study matrix was created to counter balance any order effects.

5.1.4 Participants

Participants were sourced from Coventry University and from an automotive company. All interested staff and students were advised to email the researcher for further information on the study. In addition to further follow-up emails to staff and research students, permission was granted by the Psychology department to put a notification regarding recruitment on their Faculty website.

On receiving an email from an interested individual the researcher sent, by email, a PIS and encouraged the respondent to email any questions or concerns that they may have regarding the trial. Information was gathered concerning if the potential participants were right or left-handed. If participants were left-handed they were advised that they would be participating in the study towards the end of the data collection period. This was due to the design set-up being slightly different. As only four left-handed individuals volunteered it was decided not to include them in this study. Figure 36, displays the participant recruitment flow chart for the participants from Coventry University.

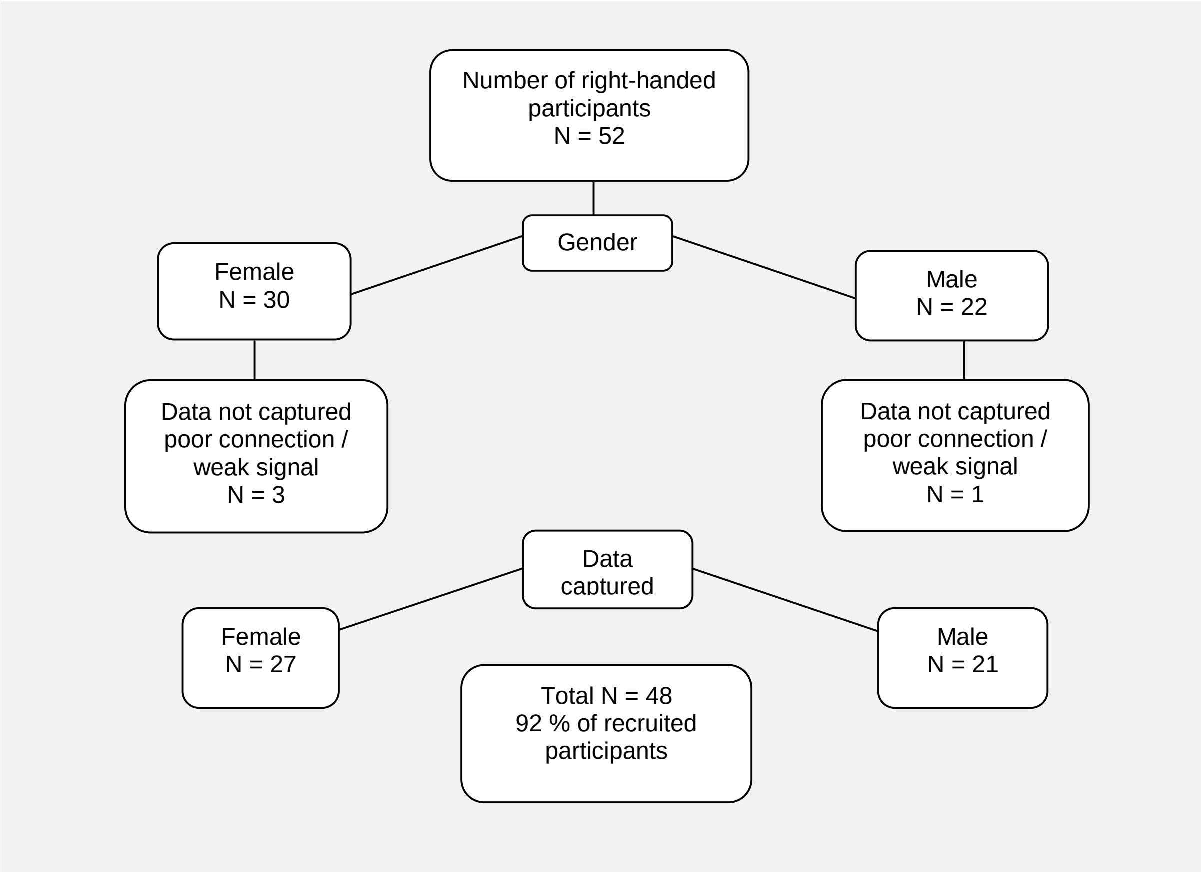


Figure 36. Study Three; Recruitment flow chart for participants from the academic community

The additional ‘pool’ of participants from the automotive company were employees and worked in their research division. The researcher’s industrial supervisor emailed the research team and set-up an appointment system for interested employees to select a time. A participant information sheet was emailed to each volunteer. This trial took place in the HMI laboratory at the International Digital Laboratory site on the Warwick University campus. Equipment and materials were simply transported to the HMI laboratory prior to the start of the study. Figure 37 displays the participant recruitment flow chart for the automotive company employees.

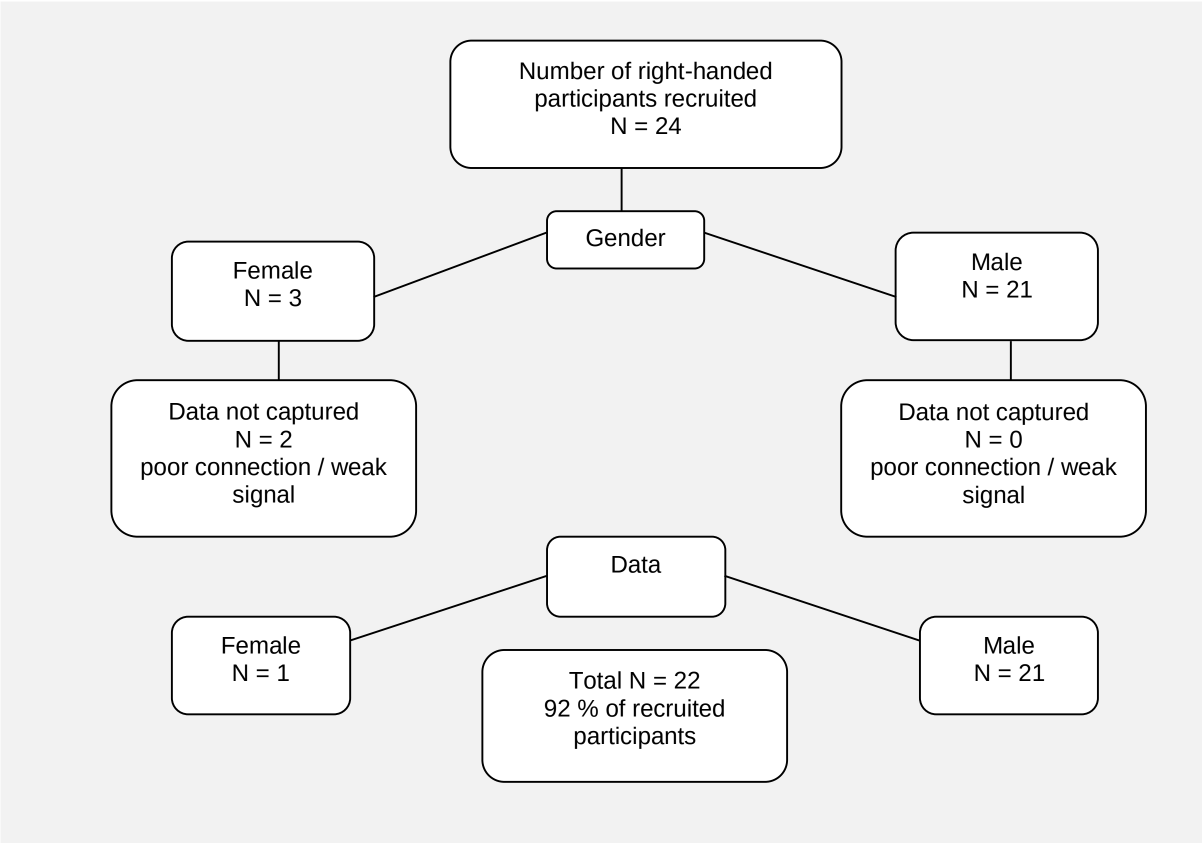


Figure 37. Study Three; Recruitment flow chart for participants from the automotive company

From the participants that were recruited from Coventry University, three groups were determined, and from the recruited participants from the automotive company, two groups were formed.

Participants were asked to abstain from caffeine drinks for at least four hours before taking part in the study. A participant information questionnaire was completed by each volunteer. The questionnaire included demographic questions and included space for any medications details to be disclosed to the researcher. If participants had been on medication that they thought may have an effect on the experiment, they would have been withdrawn.

On arrival at the laboratory, informed consent was obtained, before participants were asked to remove their shoes for a height and weight measurement to be taken and recorded. Participants were then invited to sit for the preparation and placement of the physiological equipment. The demographic information from the participant groups is shown in Table 8.

Table 8. Participants' grouped demographic information

Group	Participants from	(N)	Gender	(N)	Mean Age	(SD)	Mean BMI	(SD)
A	Academic group	20	M	(8)	36	(13.22)	24.54	(2.95)
			F	(12)				
B	Academic group	10	M	(6)	32	(14.51)	24.32	(3.27)
			F	(4)				
C	Academic group	18	M	(6)	28	(7.14)	26.20	(9.60)
			F	(12)				
D	Automotive group	12	M	(11)	30	(6.05)	25.87	(4.87)
			F	(1)				
E	Automotive group	10	M	(10)	30	(6.01)	23.50	(3.30)
			F	(0)				

5.1.5 Materials and Equipment

The HMI was designed to simulate a range of touch screen tasks that represented typical interactions that a user would experience with an in-vehicle infotainment system, and after consultation with the collaborating automotive company two graphical touch screen interfaces were provided.

The first interface used capacitive technology, was colourful and operated very smoothly. Capacitive displays can be operated with very light touches of the fingers and generally cannot be controlled with a mechanical stylus or gloved hand (Thaler and Wenzel 2012) see Figure 38, this interface was termed the 'good' interface. The dimension of the GUI was H 20 cm x W 30 cm. The second interface consisted of the same task set-up as the good interface but was presented on an android touch screen. The android touch screen also used capacitive technology but it was very 'sticky' to operate and control, with poor touch consistency. This GUI was termed the 'poor' interface. Although the range of tasks were identical to the good interface, the poor interface contained no colourful graphical icons as seen on the 'good' interface. The 'poor' interface had limited colour with the instruction headings were very small and positioned on the far left of the touch screen (Figure 39).

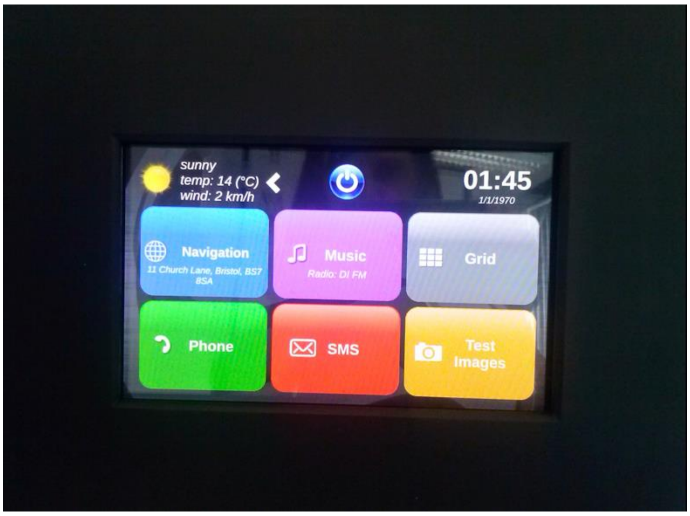


Figure 38. ‘Good’ HMI used in study three

Figure 38 shows the ‘Good’ interface displaying the colourful icons on the Home page, that help direct users. There were no icons on the ‘poor’ interface to guide users easily around the interface (Figure 39).

Figure 39. ‘Poor’ HMI used in study three

Participants interacted with three screens during the study, which are shown in Figure 40. The touchscreen HMI is positioned on the left and is a secondary task, the central screen presented the primary visual tracking task and the screen on the right was used to present the tertiary task.

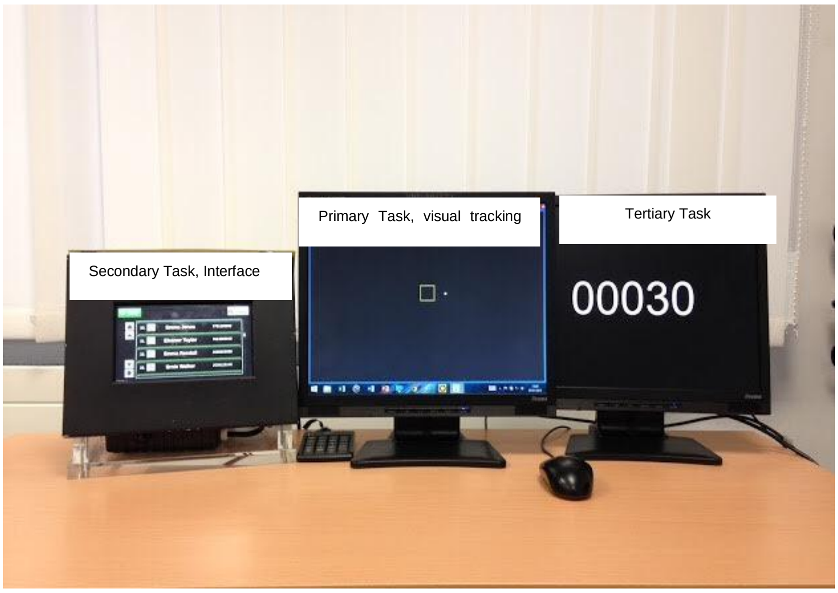


Figure 40. Participant screens for study three

Two identical visual display units (W 34 x H 27 cm), were used to present the tasks to the participants. The middle screen directly in front of the participant presented the aquatic video clip for baseline recordings and also to deliver the primary visual tracking task. The primary task was obtained from the open source, Hanover College, Cognitive Laboratory Studies. Correspondence with the designer of the visual tracking task enabled the researcher to request whether the duration of this task could be lengthened. This request was granted and the duration of the task was increased to 20 minutes (Krantz, 2014).

The task comprised of a small moving ball which travelled in different directions across the screen. The aim of the task was to keep the small ball inside the target box. The participant had control of the target box by moving a mouse with their right hand, see Figure 41. The settings for the primary tracking task could be adjusted for angle variation, speed and size of dot and target box size. The settings for the study are displayed in Table 9.

Table 9. Primary task settings

Visual Tracking Task Settings	
Angle range	360; range of direction
Speed of dot	5 (pixels per update)
Size of dot	10 (in pixels)
Target box size	60 (in pixels)

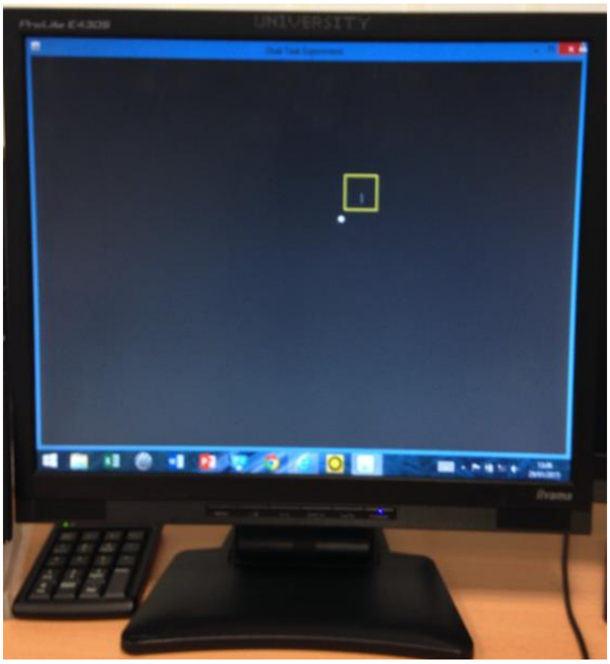


Figure 41. Primary visual tracking task

The second visual display screen was used to present the tertiary task, which was created using PsychoPY (Peirce 2009), (See section 4.1.4 for a full description). For this study a visual digit task was produced, and comprised of a series of five digits. Four of the digits were zero digits and one was a non-zero number. The participants were required to input the correct location of the non-zero number, working from left to right. An example is illustrated in Table 10. Over the duration of the study, the digit task was presented to the participants on three occasions, and on each occasion the participant had to attend to a string of three digits.

Table 10. Example of digit task (one string)

Digit Task	
Series presented to participant	00003
Participant's correct response	5

The touchscreen graphical interface, was positioned to the left of the participant and was operated by the participants left hand. The tertiary task was presented on the screen to the right of the participant. The participant interacted with this screen by using their left hand and inputting information into a keypad, which was brought forward at the start of the study and positioned at the left-side of the central screen.

Two subjective self-report scales were used in the study to gain user feedback on the graphical interface. One was a Visual Aesthetic Scale (Moshagen and

Thielsch 2012) and the second self-report scale was a System Usability Scale (SUS) (Brook 1986), see Appendix 4 and 5.

To calculate the scale value for the SUS the odd number items (1,3,5,7 and 9), are taken as the scale position, minus 1, whereas for even items in the questionnaire their score contribution is the scale position, minus 5. The sum of the score contributions is then added and multiplied by 2.5, this result is the SUS score which ranges from 0 – 100 (Brooks 1986). The score for the Visual Aesthetic Scale is simply by adding each of the values that have been assigned to the four different items. For each item there is a maximum score of 7.

At a desk behind the participant the researcher had three personal computers; one to display and record the participants' physiological data in *'real time'*, the second displayed the baseline recording and monitored the primary task. The third computer was used to present the tertiary task.

For the third study, a three-lead Electrocardiograph (ECG) was used to record the electrical heart signals. The ECG electrodes were positioned using the Einthoven's method (Martini 2006). The participant preparation was slightly more involved than the chest transducer band (Study One and Two) as disposable electrodes were placed directly onto the participant's upper body. Leads were then fitted onto the electrodes. For each participant, access to their clavicle area and lower left thorax area was necessary for correct placement. Skin had to be cleaned and dried before placing the pre-gelled disposable electrodes onto the appropriate sites of the upper body. The first electrode (negative) was positioned immediately below the right clavicle and the second electrode (positive) was positioned on top of the lower ribs on the left side of the body, as illustrated in Figure 42.

This material has been removed from this thesis due to Third Party Copyright. The unabridged version of the thesis can be viewed at the Lanchester Library, Coventry University.

In addition to the heart signal recording, the respiration rate and activity from the electrodermal system was also captured in the third study. A thoracic respiration band was secured around the participant's chest, with the transducer fitment centred on their sternum (as described in Study One and illustrated in Figure 23). When placing the transducer chest band for respiration monitoring, it is important that the band does not interfere with the ECG electrode placement. To record the participant's electrodermal activity (EDA), the electrodes were positioned on the medial plantar surface of the foot, below the ankle (Figure 43). This is not the preferred site for EDA, but due to the nature of the trial, the participants moved their hands and fingers when interacting with the trial equipment. It was decided that if the preferred site, which is the fingers, were used there would be a strong possibility that the signal would be disrupted.

Figure 43. Illustration and image of electrode sites for EDA recording, (A and B) discs (Bouscein 2012, Furness 2016)

The hardware equipment for recording the physiological data was a Biopac MP36R system (with the accompanying AcqKnowledge 4.4 software). The MP36R system is a four channel computer-based data acquisition system that records a wide range of data from different biosignals (see figure 44 and 45).

Figure 44. Biopac MP36R, physiological monitoring equipment (Furness 2016)

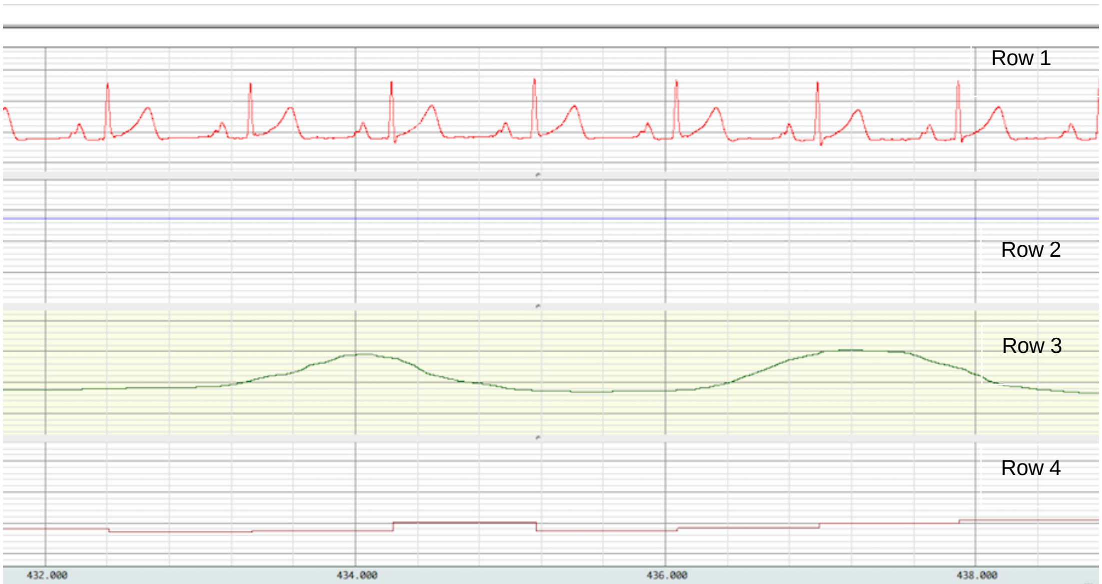


Figure 45. Screenshot of biosignals recorded during Study Three, row 1 ECG, row 2 EDA, row 3 respiration and row 4 heart rate (Furness 2016)

The AcqKnowledge 4.4 software was employed to view, edit, transform and analyse the incoming data. After attending training sessions provided by the suppliers of the Biopac 36R equipment and personal communication relating to the acquisition rate and filters, acquisition rate for all incoming data was set to 2000 samples per second. Rationale for this was that samples could always be removed, but samples can not be put back into the signal if they were not there originally (Braithwaite *et al*. 2013). Filters were applied to ECG data to reduce noise and arifacts that could potentially inhibit HRV analysis. A low pass filter was enabled (LPN) at 35Hz and a high pass filter (HP) at 0.5Hz. For the respiration channel a LP filter at 1Hz and a HP filter at 0.05Hz was applied. The EDA channel was set to 0.35Hz for tonic recording with a high pass filter of 0.05Hz. Before EDA recording can commence it is important that EDA callibration occurs. EDA callibration occurs automatcially using AcqKnowledge software but it is important to remember that for this procedure there must be no connection between the electrodes. The leads are detached from the disposable electrodes until the process is complete whereby reconnection with the participant is completed and recording can commence.

Disposable, pre-gelled with isotonic gel, snap electrodes with an accompanying lead set were used to record EDA activity (Biopac EL507 and SS57L). Electrical heart activity was captured using disposable electrodes and a fully-shielded cable with ‘pinch’ connectors (Biopac EL503 and SS2LB). Although both disposable electrodes have pre-gelled cavities, it was also recommended that for EDA recordings the use of a specially formulated gel was applied (Braithwaite et al., 2013). For this study Biopac Gel 101 was used. This gel is designed to mimic the salt concentration of sweat and contains 0.5% saline in a neutral base. Similarly, with ECG recordings an electrode gel was applied in addition to the already pre-gelled disposable electrodes. To record the respiratory effort of the participants a respiration transducer band was used (Biopac SS5LB). The transducer band measured the change in the thoracic circumference and was worn around the body at the level of maximum respiratory expansion. To achieve this, each participant was instructed to take a large inhalation followed by a slow exhalation. On completion of exhaling the respiratory band was tightened around the participant's chest. By observing the participant and following product instructions, the maximum respiratory expansion was approximately five centimetres from the participant's underarms (Biopac 2012). The respiratory transducer band was attached to an adjustable strap, allowing it to fit a wide range of individuals.

Additional equipment for study three included a stadiometer, weight scales and a video recorder and stand. Participants weight and height were recorded, which enabled their Body Mass Index (BMI) to be calculated. This was achieved by obtaining their weight in Kg, which was divided by their height in metres squared. Recording the participants was performed to act as a back-up for identification purposes only.

HRV data was processed by AcqKnowledge software and analysis programme, before it was either placed in Excel sheets of SPSS (version 22). ANOVA and Post Hoc testing was conducted.

5.1.6 Procedural Approach

On completion of the consent form, the EDA and ECG electrodes were attached first to the participants, as it was important to allow 10 minutes for the electrode gel to be absorbed into the skin. An important consideration for EDA recording is the ambient temperature, which should be approximately 22-24 degrees Celsius (Braithwaite *et al.* 2013). The researcher set the thermostat in the laboratory at 22 degrees Celsius for the duration of the study as well as ensuring that all windows were closed.

EDA sensors were placed directly onto the skin. Cleansing of the foot or other preparation is not advised, as it is important to preserve the electrical properties of the skin (Dawson, Schell and Fililon 2007). Participants now had their ECG electrodes sites prepared before securing the disposable electrodes in situ, (as described in Section 6.1.4.) The participant preparation was slightly more involved than in Study One and Study Two, as the electrodes were placed directly onto the participant's upper body. Leads were then subsequently fitted onto the two individual electrodes. For each participant, access to their clavicle area and lower left thorax area was necessary for correct placement. Skin was cleaned with a non-alcohol disposable wipe and dried before placing the pre-gelled disposable electrodes onto the appropriate sites of the upper body. The participants were now instructed to sit at the desk where they would remain until the end of the study (as seen in Figure 46).

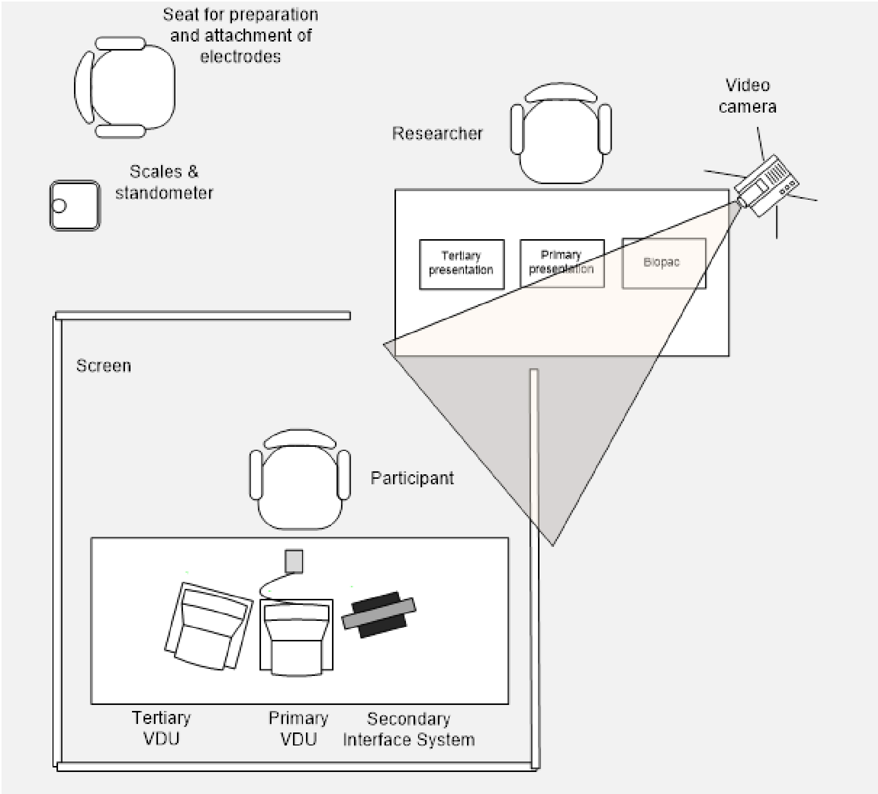


Figure 46. Laboratory layout for Study Three

The participants were handed a set of headphones and instructed to relax and watch the aquatic clip with accompanying music; to gain a baseline recording (as per the previous study). Further to this, a second identical baseline was presented to the participants at the end of the study. The duration of the baseline recording was five minutes, with recording of the physiological measures commencing at the beginning of the aquatic clip. By recording two baselines the researcher can omit the first and last sections of the recording and take the middle 2.5 minute sections from each baseline. This allows the participant to settle at the beginning of the period and then towards the end, when they are beginning to think about the tasks ahead.

After the baseline recording had ended, the researcher sat next to the participant for the familiarisation period. This stage involved the participants being informed about the trial and the different tasks that they would complete. Each screen was explained and the participants had the opportunity to ask any questions about the study. The familiarisation period also gave the participants the opportunity to see the interface for the first time, as it had been covered-up until the start of the familiarisation period. The participants were handed an interface guide that gave step-by-step details concerning each interface task. Five tasks were designated in order to allow the participants to interact with the infotainment system. For each of the five tasks a task flow chart was produced to deconstruct the different cognitive processes that were involved in completing each task (as in Appendices 9 - 13). The five interface tasks are described in Table 11.

Table 11. Interface tasks

Interface Tasks	Participant's Goal
Satellite Navigation	Locate a given address
USB Music	Find specific music track
SMS Message	Write an SMS message and send
Telephone	Locate a phone contact and call
Radio	Select a specific radio station

The participants were instructed to follow the guide and complete each of the tasks. The participants completed the tasks in their own time, but could only move on to the next task when the previous task had been correctly completed. The researcher stayed beside the participant during this period, in case they had any questions pertaining to the interface tasks. This familiarisation period also gave the researcher the opportunity to examine the physiological data being recorded, and if there was any signal instability adjustments could be made to the electrodes and leads during this time. Signal instability may have occurred due to kinking of the leads, or the position of the sensors may have moved. On completion of the familiarisation period the researcher re-positioned herself to sit at a desk, behind the participant. The participant's study area was screened to prevent distractions from the environment (see Figure 46).

An experimental matrix was created for the allocation of participants to different groups, depending on location, interface type and whether they would be completing the digit location tertiary task. The group dynamics are described in Table 12.

Table 12. Participant group allocation for study three

Recruited Participants	Label	Interface used	Tertiary task
Academic group	Group A	Good interface	Tertiary task presented
Academic group	Group B	Good Interface	No tertiary task presented
Academic group	Group C	Poor Interface	Tertiary task presented
Automotive group	Group D	Good Interface	Tertiary task presented
Automotive group	Group E	Good Interface	No tertiary task presented

On completion of the familiarisation period the researcher handed the participants a visual aesthetics questionnaire to complete (Moshagen and Thielsch 2013), (Appendix 4). This questionnaire related to their initial impressions of the interface they had just seen and interacted with. The questions related to the diversity of the interface, the attractiveness and the simplicity of the design.

The participant took control of the mouse with their right hand to begin the visual tracking task, which ran for 20 minutes. The aim of the visual tracking task was to keep the small ball inside the target box. The visual tracking task was included in the trial to keep the participants' eyes focused on a central area, while at the same time, attend to or monitor another task. Although this was not a driving task, the principle of a dual task is often seen in a driving environment, an example would be when drivers have to monitor their road position while monitoring their speed dials. When the researcher was explaining the tasks to the participants, they were instructed that for the duration of the study their full attention and focus had to remain on the primary task, and at no time during the study could they remove their right hand from the mouse, which controlled the primary task. An average tracking error was recorded during the trial.

When the participants initiated the tracking task to commence, they had a two-minute familiarisation period where they only had the tracking task to operate. After the familiarisation period the researcher verbally instructed the participant to complete their first interface task. The interface was set to display the 'home page' when the verbal instruction was given to the participant. Once the interface task was completed, the timing stopped when the interface screen returned to the 'home page'. This procedure was the same irrespective of the interface type. Between each interface task there was a two-minute section where the participants only operated the visual tracking task. All five interface tasks were completed by the participant (following experimental matrix). The participants used their left hand to interact with the interface, and they also used their left hand for the tertiary task.

In addition to the interface tasks, one of the conditions required participants to complete a tertiary task to complete while they were interacting with the interface. This involved a digit task that would appear on the right-hand screen, (see Figure 39). Participants did not know whether they would be attending to the tertiary task until it appeared on the right-hand screen in the participants' peripheral view. All participants were given the same instructions regardless of their participation in the tertiary task and were instructed to attend to the tertiary task as soon as they were aware of it. Participant response times were recorded in relation to the digit task. During the familiarisation period the tertiary task was explained to the participants and each participant demonstrated their understanding of the task by completing several examples, which were included in the study instructions.

After completing the five interface tasks and the primary task, each participant was asked to complete a System Usability Scale (SUS) (Brook 1986). This is an industry recognised scale that is easy to administer and provides valid results even when the sample is small (Sauro 2011). The SUS rating is scored out of 100, although it is not a percentage, it is considered as a percentile ranking. A SUS score above 68 is deemed above average (and thus an acceptable usability score achieved) and below 68 is considered below average (Brooks 1986). A final baseline recording was taken, before the researcher removed the monitoring equipment from the participants.

On completion of the study the researcher held a short post study debrief to obtain any further feedback from the participants regarding the study. Comments that had been recorded during the study were also discussed with the participants to ensure the accuracy of them in context to the interface. This qualitative data was for assessed for frequency of words and phrases then inputted into a ‘word cloud’ generator (Wordle 2014) for a visual display of feedback.

The duration of the study was approximately one hour between the time the participant entered and left the laboratory. The study order is illustrated in Figure 47.

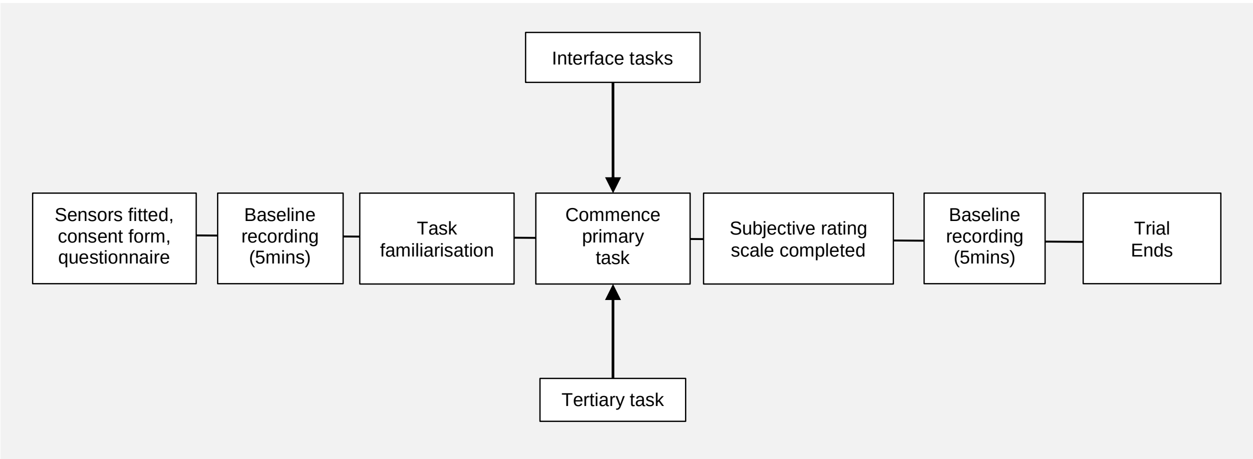


Figure 47. Task order for Study Three

5.1.7 Results

Demographic details revealed the following information (see Table 7). A Body Mass Index (BMI) was calculated for each participant by obtaining their height and weight measurement, (see 5.1.5). For this study an underweight BMI is < 18.5, a normal BMI range is between 18.5 – 24.9 and an overweight BMI score is between 25 – 29.9. An obese BMI score is considered 30 or more (NHS Choices Information 2014). Two participant groups had an overweight BMI score, one from the academic community (Group C = 26.20 SD 9.60) and the other from the automotive industry (Group D = 25.87 SD 4.87). The other three groups recorded normal BMI scores. A BMI score was calculated as there has been research conducted into the effects of obesity on HRV, and studies have indicated that HRV can be reduced in a person that is obese. By taking a BMI measurement, assessment of an individual that had an obese BMI score could be observed to see if their data had any anomalies that would affect the mean results for that group.

An average tracking error score was recorded over the duration of the trial. Each screen was updated 20 times per second and the distance from the box to the target (in pixels) was averaged over the trial (Krantz, 2014). Figure 48 illustrates the different participant groups average tracking error over the duration of the experiment.

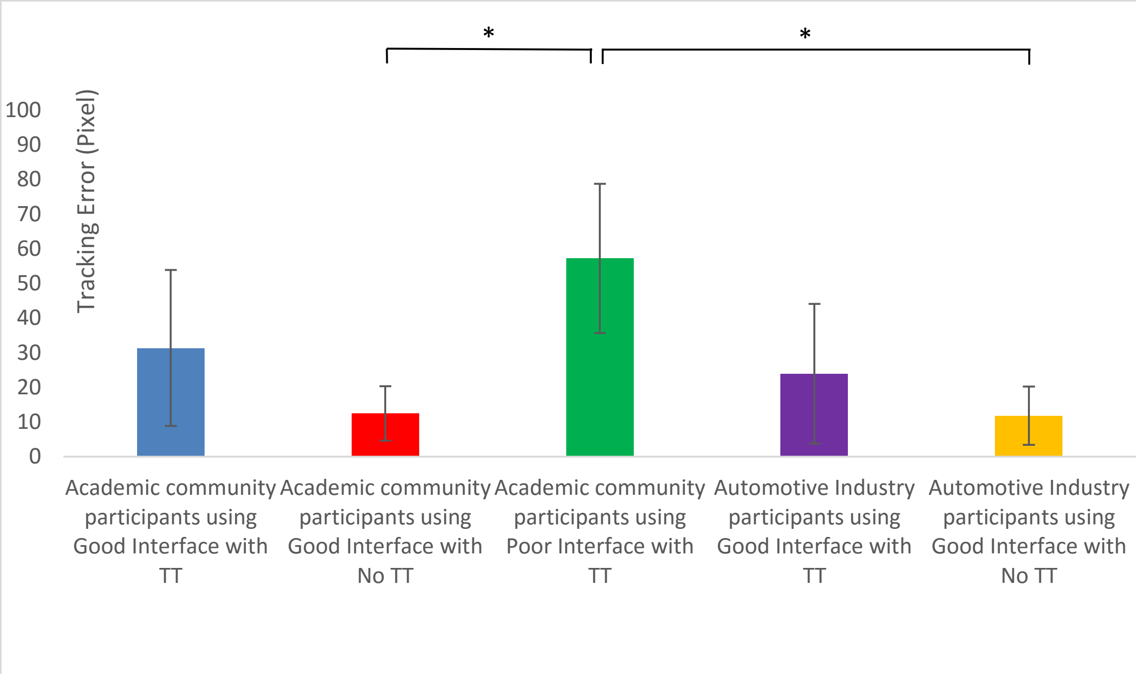


Figure 48. Participants tracking error score
One-way ANOVA with post-hoc test revealed a significant difference between the group using the poor interface and the two groups using the good interface with no tertiary task $p < 0.05$ (* represents significant result).

Figure 48 demonstrates that the academic group using the poor interface and attending to the tertiary task made significantly more errors than the groups that were interacting with the good interface. Their average tracking error was 57.22 (SD). A one-way ANOVA and Post Hoc test revealed that there was a significant difference between this group of participants and the academic group using the good interface F(4,70=P<0.05) and also the automotive group using the good interface F(4,70=P<0.05). The two groups that interacted with the good interface and had no additional tertiary task had very similar average tracking error scores. The academic groups tracking error score was 12.46 and the automotive groups tracking error score was 11.84.

Figure 49 displays the information relating to the average response times that were recorded when the participants reacted to the digit tertiary task.

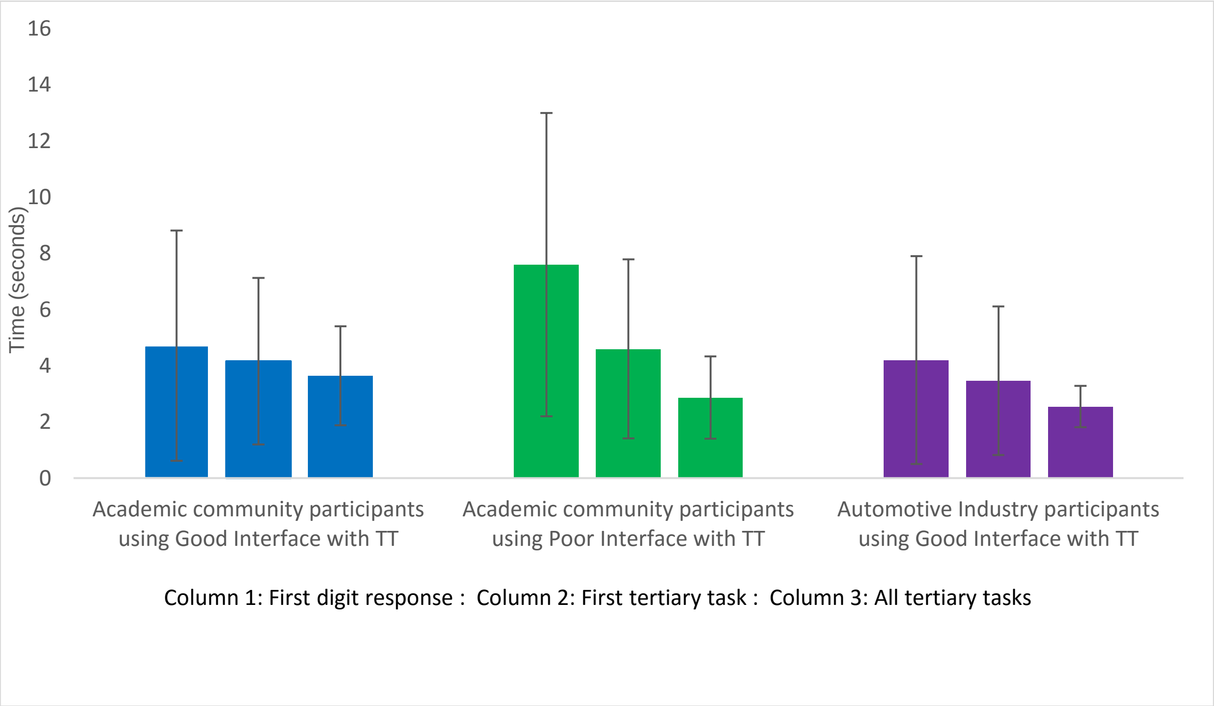


Figure 49. Reaction times to digit tertiary task;
participants attended to three tertiary tasks during the study, each task comprised of three digit strings. The digit strings were four zero digits with one non-zero digit. The participants had to select the correct location of the non-zero digit.

The academic group using the poor interface had the slowest response time to the first presentation of the digit tertiary task (7.6 seconds) and the slowest time averaged over the three digit strings of the first tertiary task (4.6 seconds). There were no significant differences between the groups.

Participant interface task completion times were recorded over the five different interface tasks. The satellite navigation task took the longest time to complete over all participant groups. In all participant groups, except group C (academic group using poor interface), participants completed the tasks in a similar time order, with the music track selection task next followed by the telephone call task. The radio selection task was the quickest interface task to complete for all participant groups except group C (academic group using poor interface). In group C however, the completion times for the task were in a reverse order apart from the satellite navigation task, which took the longest time to complete. No significant differences were revealed between the different groups and their interface completion times, this was reflected by observing the mean completion times across the groups, see Figure 50.

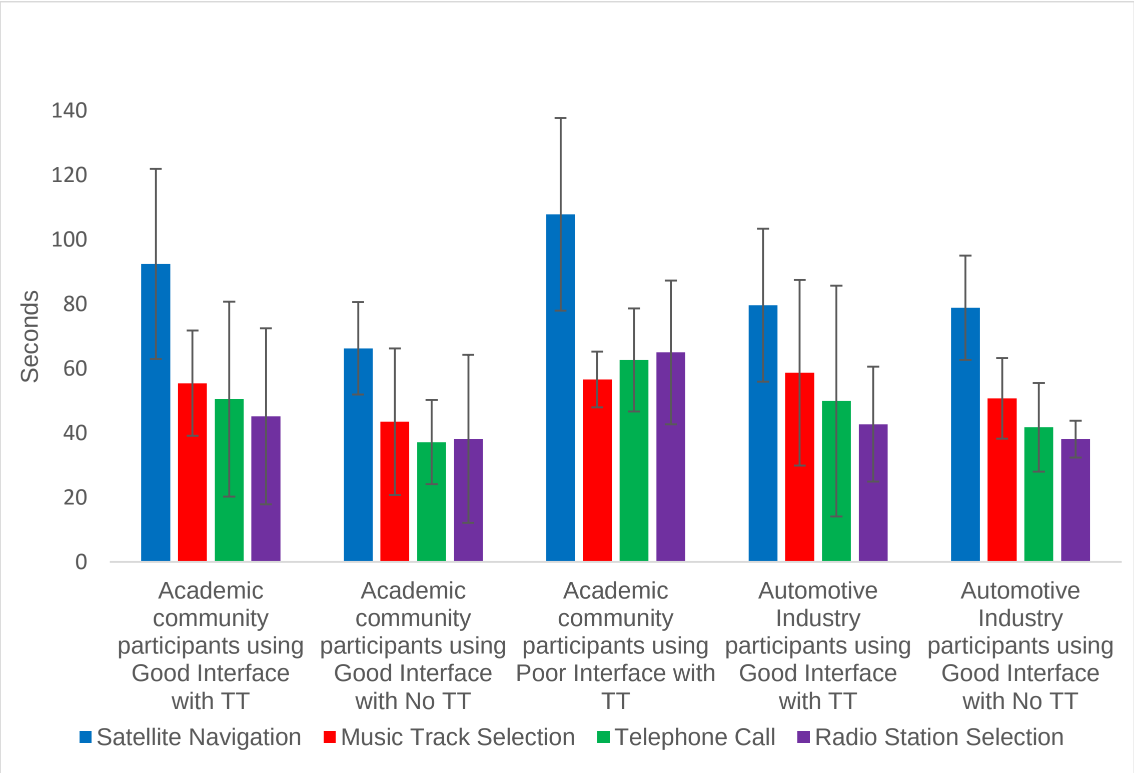


Figure 50. Interface task completion times;
Interface completion times were recorded over four interface tasks for each group.

5.1.7.1 Physiological Results

Respiration rates (breaths per minute) were recorded throughout the study, there were no significant findings between the different groups or within the groups as the participants progressed through the trial tasks (Figure 51). Heart rate data (beats per minute) was also recorded and similar to the respiration rates there were no significant differences between the different groups or within the groups as they attended to the different tasks. Figure 52 represents the heart rate change of the participant groups during the study.

The group using the ‘poor’ interface had a greater magnitude of change in their heart rate when interacting with the interface. Average tonic skin conductance levels were recorded throughout the duration of the study. Figure 53 displays the skin conductance data. Repeated measures ANOVA revealed significant findings from baseline to primary task in academic groups using the ‘good’ interface and the automotive group with the tertiary task ($p<.05$). Significant findings from baseline to interface tasks in the two academic groups using the ‘good’ interface and the automotive groups with the tertiary task ($p<.05$). Significant results were also indicated from baseline to interface tasks in the two academic groups using the good interface and the automotive groups using the good interface with tertiary task ($p<.05$).

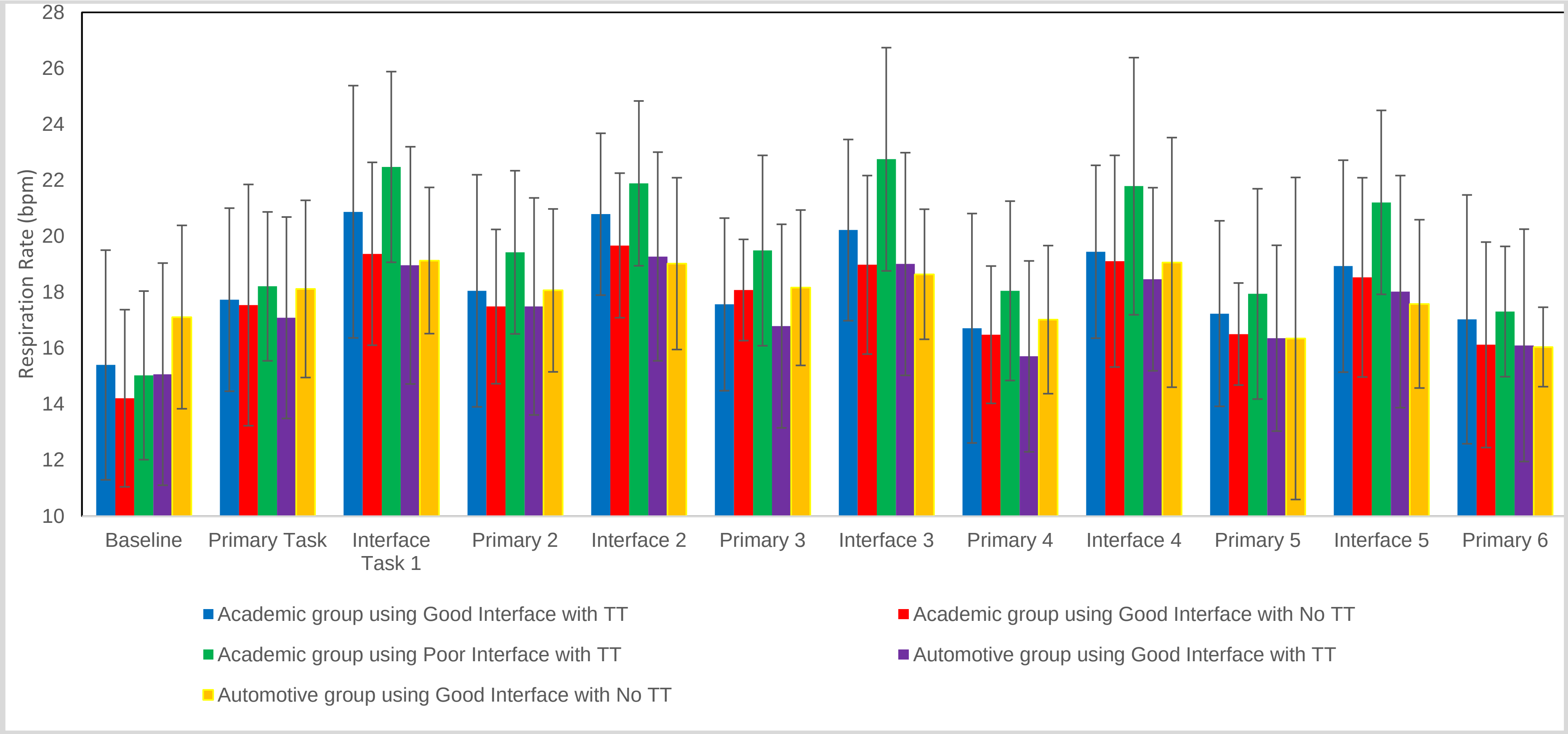


Figure 51, Participant groups average respiration rate across the trial tasks;
The study was divided into primary task sections and interface task sections, primary tasks lasted for 2 minutes whereas the interface tasks depended on the participants' completion time. The study progressed from the baseline to primary task 6. During the primary task sections, the participants only attended to the primary visual tracking task, whereas during the interface tasks, participants also completed an interface task at the same time as they were attending to the primary task.

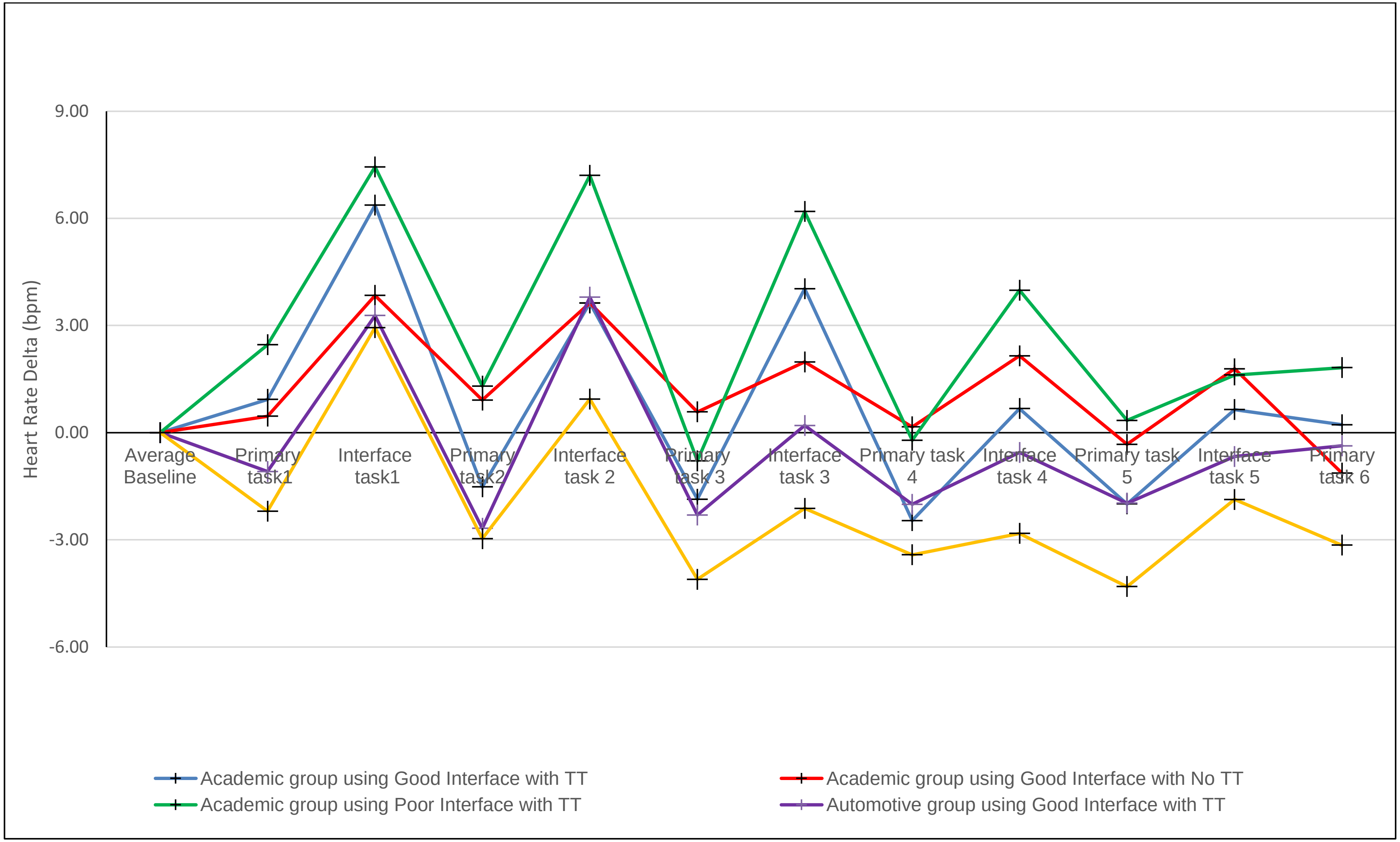


Figure 52. Participant groups heart rate change from baseline, progressing through study tasks

Average tonic skin conductance levels were recorded throughout the duration of the study. Figure 53 displays this information.

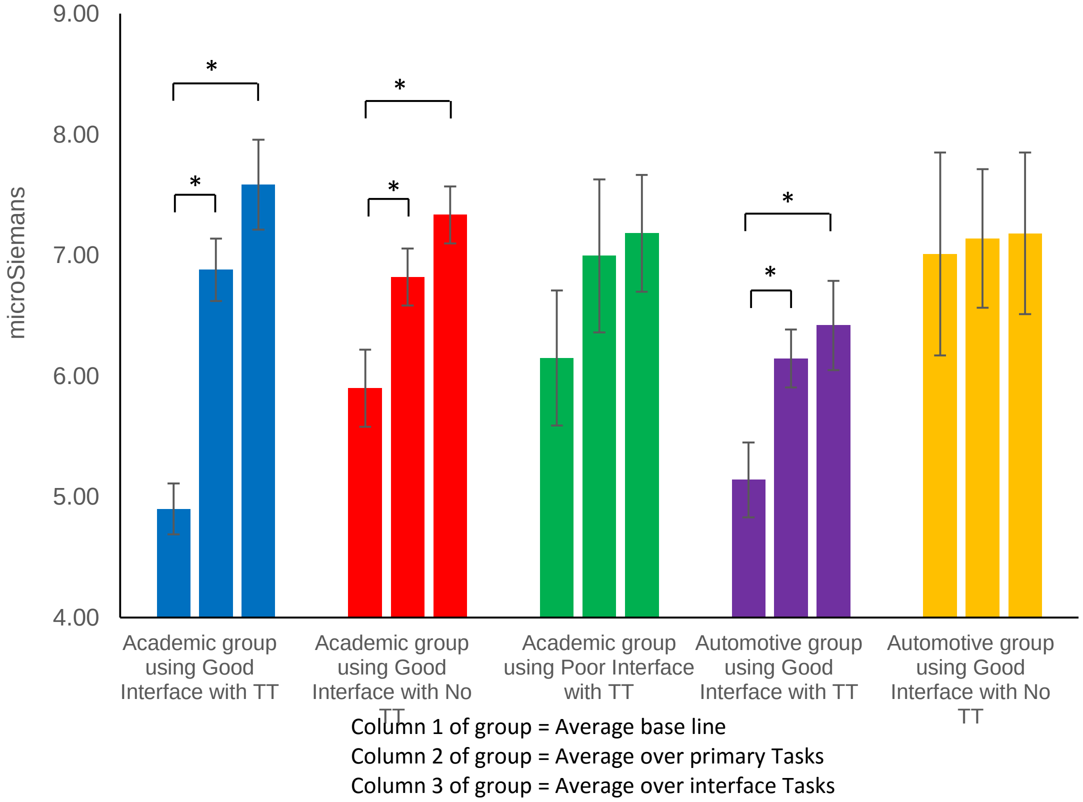


Figure 53. Average skin conductance level

One-way repeated measures ANOVAs were conducted to determine whether there were statistical differences in mean skin conductance levels during the

For each participant group, physiological measurements recorded during the study are summarised in the tables below. The first three rows refer to average heart rate (bpm), average tonic skin conductance levels (mS) and average respiration rates (breaths per minute). The remaining rows are HRV indices. The first two HRV rows refer to time domain metrics and the remaining rows are frequency domain metrics. Percentage changes from baseline to tasks are stated for HRV LF and HF power.

Table 13. Academic group using good interface with tertiary task
(Group A): physiological measurements, mean and (standard deviation)

Variable	Baseline	Primary Task 1	Interface Tasks	Interface with TT	Change from Baseline to Primary Task	Change from Baseline to Interface Task
HR	73.754 (6.495)	74.68 (9.801)	80.12 (10.51)	77.78 (10.60)	0.93bpm	6.37bpm
SCL	4.901 2.301	6.953 2.126	7.586 2.203	7.230 2.331	2.05 mS	2.68 mS
Resp. Rate	15.403 *4.106*	17.733 3,270	21.762 2.893	19.414 3.088	2.33 bpm	6.36 bpm
RMSSD	34.882 19.940	37.297 20.060	33.853 18.838	22.227 17.238	2.41	-1.03 ms
pNN50%	16.002 17.649	16.225 17.242	11.184 13.798	13.804 15.445	0.22	4.82%
LF Power	811	856 5.9%	340 -58.08%	436 -46.2%	44 ms^2	471ms^2
HF Power	646	772 19.5%	369 -42.88%	474 -26.6%	126 ms^2	276 ms^2
LF / HF Ratio	2.85	2.34	0.46	2.62	0.51	2.39

Table 14. Academic group using poor interface

with tertiary task (Group C): physiological measurements, mean and (standard deviation)

Variable	Baseline	Primary Task 1	Interface Tasks	Interface with TT	Change from Baseline to Primary Task	Change from Baseline to Interface Task
HR	74.691 (7.482)	77.150 (9.024)	82.132 8.876	83.775 7.598	2.46 bpm	7.44 bpm
SCL	6.152 3.600	7.713 4.222	7.882 4.074	8.215 4.149	1.56 mS	2 mS
Resp. Rate	17.113 4.182	21.004 2.662	21.420 3.402	21.523 3.989	3.89 bpm	4.31 bpm
RMSSD	38.845 15.486	42.853 18.106	32.724 14.013	39.653 15.512	4.01ms	6.12 ms
pNN50%	14.85 10.444	13.08 13.396	11.186 13.798	9.41 19.945	1.77 %	3.67 %
LF Power	1282	826 -35.57%	128 -90.01%	375 -70.7%	456 ms^2	1153 ms^2
HF Power	783	698 -10.85%	202 -74.20%	246 -68.6%	85 ms^2	581ms^2
LF / HF Ratio	2.26	1.35	0.35	2.14	0.91	1.91

Table 15. Automotive group using the good interface

with tertiary task (Group D): physiological measurements, mean and (standard deviation)

Variable	Baseline	Primary Task 1	Interface Tasks	Interface with TT	Change from Baseline to Primary Task	Change from Baseline to Interface Task
HR	71.842 8.886	70.673 10.68	75.048 7.63	71.960 9.873	1.17 bpm	3.2 bpm
SCL	5.143 2.951	6.590 2.956	6.427 2.814	6.702 2.927	1.45 mS	1.28 mS
Resp. Rate	15.072 3.972	17.084 3.601	18.202 3.275	19.23 3.980	2.01 bpm	3.15 bpm
RMSSD	38.542 15.662	44.337 13.074	33.180 15.126	43.871 17.001	5.79 ms	5.36 ms
pNN50%	17.320 11.223	23.590 10.435	16.304 12.021	14.118 12.795	6.27 %	1.02 %
LF Power	1002	853 -14.8%	441 -56%	758 -24.35%	149 ms^2	761 ms^2
HF Power	788	963 22.2%	566 -28.17%	632 -19.8%	175 ms^2	222 ms^2
LF / HF Ratio	2.94	1.21	0.35	2.79	1.73	2.59

Table 16. Automotive Group using good interface with No tertiary task (Group E): physiological measurements, mean and (standard deviation

Variable	Baseline	Primary Task 1	Interface Tasks	Change from Baseline to Primary Task	Change from Baseline to Interface Task
HR	75.34 7.805	73.143 5.740	78.282 6.789	2.2 bpm	2.94bpm
SCL	7.013 3.083	7.867 3.455	7.186 3.449	0.85 mS	0.17 mS
Resp. Rate	17.11 2.321	18.11 3.166	20.018 3.073	1 bpm	2.9 bpm
RMSSD	26.685 14.772	30.731 15.799	22.510 15.964	4.05 ms	4.17 ms
pNN50%	14.010 11.324	15.321 13.978	11.456 12.024	1.31 %	2.56 %
LF Power	834	931 11.3%	418 -49.88%	97 ms^2	416 ms^2
HF Power	548	479 -12.59%	372 -32.11%	69 ms^2	176 ms^2
LF / HF Ratio	2.85	2.34	0.46	0.51	2.39

The academic group using the poor interface has large percentage changes in their LF and HF power from their baseline measurement to the interface tasks as shown in Table 14. All participant groups showed a decrease in their LF and HF power from baseline to interface tasks

5.1.7.2 Results from self-report scales and participant feedback

Participants completed a visual aesthetics questionnaire after the familiarisation period of the study. The results were mapped accordingly to participant groups. Average group values for the aesthetic scale are shown in Figure 54. Questions related to the diversity, attractiveness, simplicity and design of interface (Appendix 4). The value range of the questionnaire was between 0-35.

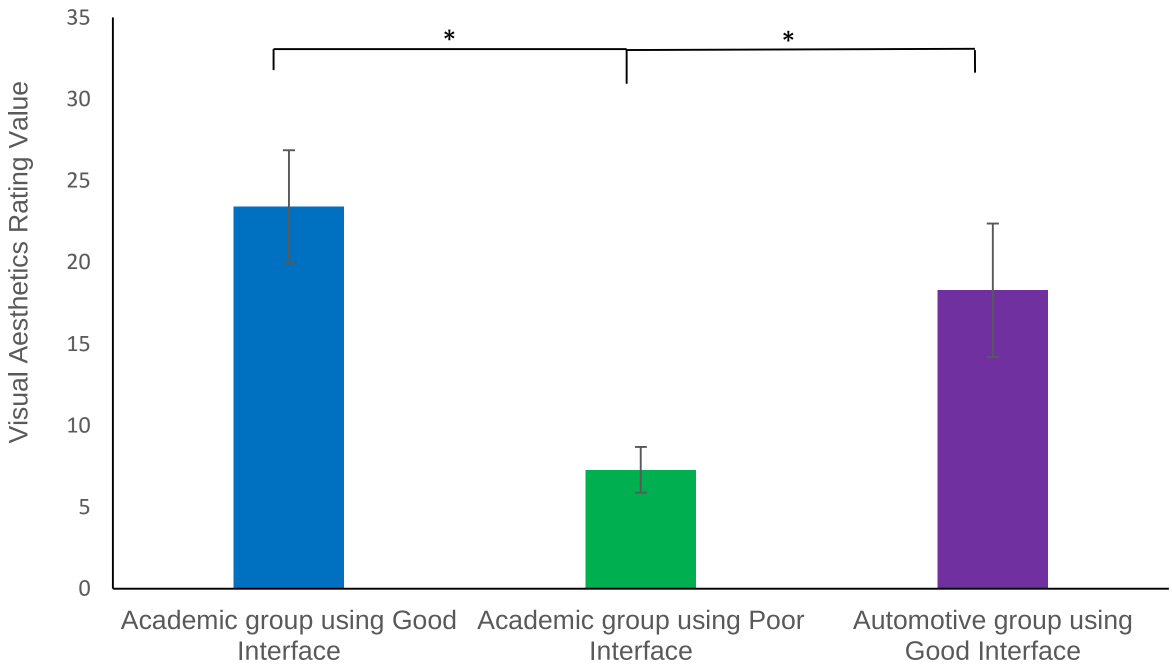


Figure 54. Participant Visual Aesthetics values

A one-way ANOVA test with a Tukey Post Hoc test revealed statistically significant differences between participant groups that were using the good interface and the group of participants that were using the poor interface,

$F(2,67)=136.514, p < 0.05$.

In addition to the visual aesthetic scale, participants also completed the System Usability Scale (SUS) to determine the participants' opinion on the specific interface they were interacting with. A SUS scale of above 68 is regarded as above average with a score below 68 is regarded as below average. The results are displayed in Figure 55.

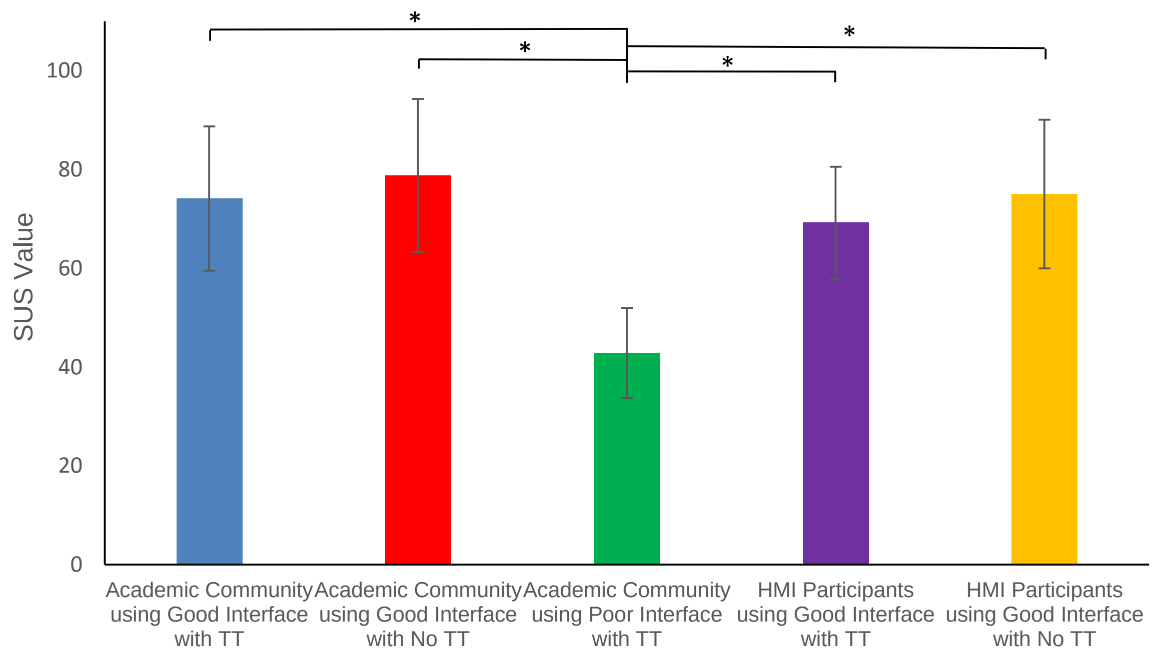


Figure 55. System Usability Scale scores

A one-way ANOVA and Tukey Post Hoc tests were applied and revealed that there was a significant difference in the participants that were using the ‘poor’ interface and the participants using the good interface, independent of the sectors they came from, or whether they had tertiary tasks to complete during the study,. $F(4,67)=19.841$, $p<.05$

Qualitative responses that were heard by the researcher as the participants interacted with the interface were recorded and additional comments were recorded during the post study debrief session. Comments were then assessed for recurring words and phrases from the participants, which are shown in Figure 56 and Figure 57. The ‘word clouds’ shown below are a visualisation technique which highlights the frequency of the words from the participants with a larger text size (Wordle 2014). The group using the poor interface made several negative comments about the interface with four comments frequently used: Difficult; Not responding; Not working; and Can’t see.

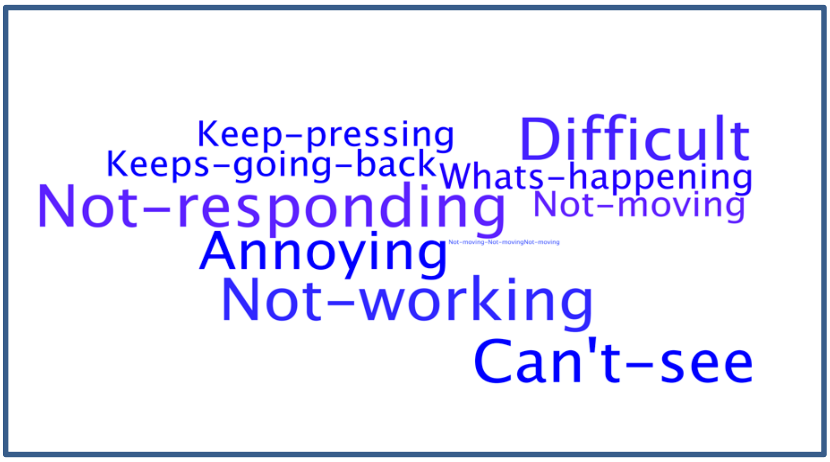


Figure 56. Feedback from participants using the poor interface

Figure 57. Feedback from participants using the good interface

Feedback from the groups that had interacted with the good interface produced several positive comments, with ‘Easy’ and ‘Clear’ being the two most frequently used.

5.1.8 Discussion

5.1.8.1 Trial Tasks

Study three was considerably more complex to run than the two previous studies. There were many different components to integrate into the study, which subsequently increased the number of study variables that needed to be controlled to ensure a robust and controlled design.

The results from Study Three clearly indicated that the group of participants that were interacting with the poor interface, were in agreement that the interface was neither aesthetically pleasing nor easy to use (as per Figure 54 and 55). This was also reflected in the human performance tasks (task completion times and average tracking error), thus suggesting that a more challenging interface influences human interaction not only with that interface but with also other primary (vigilance) and tertiary (digit problem-solving) tasks.

A low score was attributed to the aesthetics of the interface by the participants that were using the poor interface (7.28 SD1.41). The aesthetics questionnaire covered interface facets such as; attractiveness, simplicity, colour and design of the interface. This was the first rating scale to be completed by the participants and was given to the participants before they had begun performing the tasks within the study. This result was perhaps not surprising, as previous studies suggest that users quickly form an impression of the HMI (Lindgaard *et al.* 2006, Tractinsky *et al.* 2006, Van Schaik and Ling 2009). This was also found by Tractinsky (1997) using HMI displays, but in a different cultural setting. Experiments was conducted in Japan and results indicated that high correlations existed between the aesthetics of the ATM and the perceived ease of operating the screen, reinforcing the earlier studies that aesthetics and usability are very much entwined. Similar findings that link aesthetics and perceived usability have been found in evaluating displays such as web pages (Schmidt, Liu and Sridharan 2009), mobile phones (Sonderegger and Sauer 2010) and previously in automotive displays (Lavie, Oron-Gilad and Meyer 2011).

Prior to participants entering the laboratory, all interfaces had been covered-up until the start of the familiarisation period, thus we can assume that the participants' opinions, (similar to the previous research just stated) had developed very rapidly. This initial impression formed by the individual can be compared to the formation of a mental model; whereby the individual's perceived aesthetic values and prior experiences lay the foundation for any interaction experience that follows. Indeed, it has been found that user's perception of aesthetic qualities of an interface has an effect on actual usability (see Moshagen et al 2009, Sonderegger and Sauer 2010, Ben-Bassat, Meyer and Tractinsky 2006).

Although there was no significant difference between the other groups, the automotive participants had rated the aesthetics of the interface lower than the other academic groups. This may be indicative of the nature of participants belonging to the research division within the automotive company experiencing and being exposed to a range of new generation interfaces that would influence their impressions of the interface being shown to them. Thus we could argue that this group of individuals would possess a very different sort of mental model attributed to HMI design.

With the connection between aesthetics and usability, it was important to assess the usability of the system after the participants had completed the tasks. There was a significant difference between the participants using the poor interface from the rest of the groups. All participant groups, except the academic group using the poor interface, rated the usability of the interface with an above average score. The group using the poor interface rated it with a score of 49, which was below average. A score below 68 is regarded as below average and a score above 68 is regarded above average (see Sauro 2011). Participants from the academic group and automotive group that were using the good interface but had no additional tertiary task rated the usability of the interface (74 and 75), which is higher than the automotive group that had the tertiary task (69). However, this group did not rate the interface as high as the academic group using the good interface when having to attend to a cognitive tertiary task (78). The lower usability score may be attributed to the increase in cognitive workload that was introduced, suggesting that a cognitive tertiary task had an influence on participant's perceived usability.

The connection between aesthetics and usability was demonstrated in this experiment as the group using the good interface rated the aesthetics with the highest value, then proceeded to score their perceived usability of the interface with the highest score after the interaction with the interface. This result was also found by Tractinsky *et al.* (2000). Physiology findings also reflected the participants perceived usability and aesthetic appeal. The academic group using the good interface with no tertiary task and the two automotive groups had lower heart rates with smaller magnitudes of change throughout the trial, indicating that they were having a more positive experience as they interacted with the interface. The two automotive groups also had a smaller decrease in their HF power as they interacted with the interface compared to the academic group using the poor interface and the academic group using the good interface with the additional tertiary task. Increased HF power had been linked to more positive moods and experiences as well as a calm and cheerful demeanour (Geisler et al. 2010, Kemp and Quintana 2010).

Cognitive performance was also monitored during this study; namely reaction time and error during the HMI tasks. The primary task within the study, and explained to participants, consisted of a visual tracking task. This was chosen as it mimicked the high visual cognitive load, similar to what we would expect a driver to adopt in terms of cognitive processing. Significant differences were found between the tracking error of the group using the poor interface and all the other groups. From this result it was evident that the group using the poor interface had found it more difficult to keep the moving ball in the box. It is also worth remembering that the group using the poor interface also had the additional tertiary task to attend to during the study and this was also causing a distraction.

Although no other significant differences were found between the other groups that also had to complete tertiary task, there was still a large difference in their tracking error. The two other groups that did not receive a tertiary task were able to record far lower tracking errors in the primary task with less inter-individual variation, which was reflected in smaller standard deviations. By giving the participants an additional task to complete, extra mental workload was placed on the participants, resulting in a higher likelihood of errors in the primary vigilance task.

Average response times for the tertiary task were also recorded and the academic group using the poor interface had a much slower response time. Figure 49 demonstrates that as the participants progressed through the study their average response times to the tertiary task were faster than compared with their first presentation of it. Although the groups using the poor interface task was slower to respond at the first instance as they progressed through the experiment they got faster over time.

The academic group using the poor interface was slower to process and respond to the first presentation of the tertiary task, which was due to this group having to focus and work harder with this interface to complete their given task. Evidence would suggest that an individual's cognitive processing capacity is full they are unable to attend to another task especially if the task from the same modality, as described by the Multiple Resource Theory (see Chapter 1 (1.2.2), (Paas 1992, Mousavi, Low and Sweller 1995). The participants using the poor interface were already attending to a visual tracking task, in addition to this they were now having to attend to another visual task presented to them – albeit on a different screen. This would present a competition for cognitive resource and some studies have suggested that a tunnelling-of-vision effect may occur, in which participant's attention is restricted to the centre of the visual field (Foyle, Dowell and Hooey 2001). It is also worth noting that there may be a strong likelihood that after the first presentation of the digit task, the participants were primed for it to occur again.

From the four interface tasks that were compared, the satellite navigation task took the longest time. On reflection, when viewing the video playback, it was possible to see that the participants struggled with the interface when they were selecting the radio station. The other two tasks involved a scrolling action whereas the radio selection task required a pressing action to move to the desired station. Comments from the participants during this task were noted and placed in the word cloud in Figure 56 ('keep pressing'). Although this should have been the quickest task, the inconsistencies that the poor interface demonstrated, made this task considerably longer than the radio selection task completed by the other groups. The group using the poor interface took over 19 seconds longer than the next longest time (academic group) to complete this task. Observing the participants during the study, it was obvious that those interacting with the poor interface struggled with the HMI in order to complete the tasks. Comments ranged from 'I can't get the screen to work', or 'I keep pressing the right icon but nothing is happening', (as indicated in terms of frequency in Figure 56 and 57).

The participants who experienced the good interface did not perceive these problems with their HMI. The results also indicate that by having the additional tertiary task to complete, affected the length of time that the participants took to complete the interface tasks. The two groups that were using good interfaces and had no tertiary tasks to complete were quicker in completing the interface tasks.

From the two subjective self-report scales and task specific results, the findings indicated that the academic participants who were exposed to the poor interface were having a noticeably difficult interaction experience with the HMI – reflected in both subjective and human performance measures. The application of physiological measures allowed the assessment of any associations between both subjective perceptions and performance behaviour with a direct mapping to emotional experience indicated by their physiological response. In a study that investigated everyday perceived difficulties, Williams *et al.* (2015) mapped the Emotion Regulation Scale (ERS) with HF HRV. Results revealed that everyday perceived emotion difficulties were associated with lower HF (vagal) HRV. Thayer and Lane (2000) discusses that in the Model of Neurovisceral Integration an increase in HRV can regulate negative emotions.

5.1.8.2 Physiology

Three different biosignals; ECG, EDA and respiration were captured for this study to enable a broader view of the different interactions that the participants were experiencing and how the different biosignals would reflect these different interactions. In the first two studies the participants were in a very much more controlled environment, where there was little movement throughout the whole study. This was a much more dynamic study where the participants were involved in either two or three activities at the same time.

The average heart rate of the academic groups increased from baseline as they began the trial tasks (see Figure 52). However, the automotive groups heart rate decreased from baseline to the first primary task. Due to large individual physiological variation in heart rate there was no significant differences found between participant groups and between tasks. Although there were no significant differences between the groups the observed mean from Figure 52 shows that the participants heart rate increased from the primary task to the interface task. Heart rate can vary greatly from individual to individual depending on many different co-factors, these include age, sex, health condition and level of fitness as well as each individual's own intrinsic level of heart rate.

Several interesting points can be raised from examining the heart rate data. As the participants progressed through the study and interacted with the interface their heart rate began to plateau after the third interface task and after that point the magnitude of the heart rate change was less for the duration of the trial. Although this occurred with the group using the poor interface, it was not until participants completed interface task 4 had been completed that their heart rate change decreased (Figure 52). As mentioned before there were no significant differences indicated but as seen from Figure 40 the individual variation in the participant's heart rate contributed to the very large standard deviations. A consideration relating to this would be that the participants were being asked to perform a range of simultaneously which they all would respond to very differently.

This highlights that as the tasks were introduced to the participants they were under a considerable amount of stress to complete them which is evident in their heart rate and respiration rate (see Figure 51 and 52). But, by the time they had reached and completed interface task 3 (groups using good interface) and 4 (group using poor interface) they were habituating to the tasks. Habituation occurs after the participants have completed the same task several times and these effects are indicative of the learning process (Boucsein 2012). However, as can be seen in Figure 48, the participant group that was using the poor interface, habituated later in the task and not to the same degree. This groups heart rate was very much higher throughout the whole task compared to the other groups which indicates that their interactions with the poor interface were having an impact on their physiological measures.

Psychophysiological measures of heart rate and skin conductance levels are acknowledged that they respond to arousal and mental workload challenges and increase in a linear manner with increasing task difficulty (Kramer 2001, Brookhuis and De Waard 2001 and Reimer and Mehler 2011). Increases in heart rate have also been associated with better human performance. For example, Mehler *et al.* (2008), compared middle-to-late aged drivers, after grouping them (after completion of a driving task and cognitive task) into either a heart rate acceleration group (increase by > 2bpm) or heart rate no acceleration group (< 2bpm). Findings from that study indicated that drivers with an increased heart rate also performed better in the driving task. These results link to the arousal theory, discussed earlier, (see 1.2.6) where a level of arousal (increased heart rate) is required for optimum performance (Yerkes and Dodson 1908, Coughlin, Reimer and Mehler 2009). A proposal why there was no increase in heart rate with the non-acceleration participants has been suggested that this group were already operating at a capacity and when a second workload task was presented, they had no additional resources left to invest in the new task (Mehler *et al.* 2008).

It is also important to remember that heart rate is controlled by both the sympathetic and parasympathetic branches of the ANS and heart rate may decrease by a reduction in the sympathetic activity. But heart rate may also be decreased by an increase in parasympathetic activity. This perhaps outlines the complexities of the interplay between the two branches, and as commented by

Mehler *et al.* (2008) it may have been beneficial to assess electrodermal activity at the same time as heart rate. As EDA is solely innervated by sympathetic activity, it may have provided additional information relating to the two groups of middle-age drivers (Mehler *et al.* 2008). In other studies, (Mehler *et al.* 2013, Mehler *et al.* 2009, Mehler 2010), heart rate has been the preferred measurement in similar studies, as it has been shown to be more sensitive to cognitive workload tasks. However, in this study it was not only the effect that workload had on the participant's heart rate, it was how the increased workload influenced the participant's interactions with the interface.

Participant group's skin conductance level was measured throughout the experiment and mean values were calculated for baseline, primary and interface tasks. The participant's tonic skin conductance levels increased from the baseline reading to the primary and then an additional increase to the interface tasks. The three academic groups showed an incremental increase from baseline to primary and then to interface, showing that as the participants were interacting with the interface and at the same time attending to the primary task their arousal levels increased. The automotive group that was using the good interface but had no additional tasks to complete had a high baseline measurement which did not increase as perhaps would have been expected when they were given the interface task to complete. However, when discussing this result in context with their heart rate data, it is clear that the two measures mirror each other, as during baseline this group of participants had a higher than expected baseline.

All five groups, independent of the interface and tertiary task they were using showed an increase in their skin conductance level. As the task became more complex their skin conductance level increased. This finding would also impact on the group using the poor interface as their performance in the tracking task and tertiary task would be impaired due to psychophysiological stress.

Electrodermal activity, has frequently been used as a sensitive index of emotional processing and cognitive workload (Braithwaite et al. 2013, Brookhuis and De Waard 2011). Son and Park (2011) demonstrated that using short window length periods of 10 seconds, 20 seconds and 30 seconds, SCL was sensitive to a cognitive task (n-back) but not able to distinguish between the higher workload challenge (2-back). This may be due to the participant reaching a threshold and being unable to invest in the combined activities of driving a car and completing the cognitive challenge task (Mehler *et al.* 2009).

Picard et al. (2016) in an emotion review has proposed a Multiple Arousal Theory, whereby different sides of the body may respond differently. With this in mind, it may be relevant to measure two sites on both sides of the body. As a proponent of EDA, Picard has created wearable sensors that can be worn for long periods of time on wrists and ankles. By studying EDA over time, different associations are likely to be made. A study in 2012, has linked EDA activity to the postictal period of a seizure. An increase in EDA activity has correlated with the duration of EEG suppression, where there are intense periods of activity followed by periods of no activity (Poh et al. 2012). It has been proposed that due to skin and neural tissue being closely entwined, research into the mapping of brain regions to skin may highlight valuable information (Poh *et al.* 2012). In the future it may be feasible to examine the EDA activity for more complex processing research, which is relevant and of interest in automotive studies where processing and attention are key factors. As EDA data is considerably easier to obtain than EEG data, it is a promising area for this type of study.

The final physiological measure that was recorded was the HRV indices. As was seen in Study Three, the groups LF/HF ratio decreased, and it as not until they had interacted with several interface tasks that there was an increase seen. The two time-domain indices were all similar across groups, and are known to reflect the parasympathetic activity. But perhaps of more interest were the HF power and the LF power which produced large percentage changes as the participants moved through the primary and interface tasks. All groups recorded a reduction in the LF power but the large magnitude changes came from the group of participants that had interacted with the poor interface; with a reduction of 90% as observed from their baseline reading to the interface tasks. This group of participants also had the largest HF power reductions when they interacted and attended to the tertiary tasks. As the HF power reflects the parasympathetic branch of the ANS, this suggests that the group using the poor interface is having a negative experience during the trial.

However, what was perhaps a little surprising was when the participants were given an additional cognitive task to complete (tertiary task) there was no further decrease seen. This finding has also been seen in previous studies. For example, Stuiver *et al.* (2012) adopted a short-segment approach, whereby periods of 30 seconds were assessed. The short time windows reflected the simulated emergency call task that the participants had to make to dispatch an ambulance. Results from that study revealed that reductions in HF and LF bands were seen during the interaction task, as similarly seen in Study Three, when the participants were given an additional level of stress in a difficult task, there was no further decrease in variability. This would infer that participants were already engaged to a greater degree during the task and they did not invest in any additional cognitive effort during the additional task. Stuiver *et al.* (2012) proposed that as their participant's HRV levels had been reduced greatly from the first interface task, and were already low with no additional decrease in HRV levels, this lack of response may indicate that the participants were reaching a workload ceiling. A study by Mehler *et al.* (2011), where drivers had an increasing difficult auditory task to complete (an n-back task), the HRV indices also reduced as they completed the first level of difficulty but, similar to the studies mentioned and to Study Three findings, during the additional challenge, the HRV indices did not reduce further.

To summarise the findings of the physiological indices and how they impact on the current findings, it is important to view them as a collective groups of results rather than in isolation. The range of HRV indices showed a decline when the participant groups interacted with the tasks and especially the group using the poor interface. By viewing Tables 13, 14, 15 and 16 there was a decrease in the power values as the participants interacted with the interface from the primary task.

The physiological results indicated that for all participants taking part in the study they showed an increase in arousal and mental workload, which has been demonstrated in many automotive studies, Mehler *et al.* 2009, Mehler, Reimer and Wang 2011, Reimer *et al.* 2010.). The three biosignals all increased as the participants were introduced to the study, but as the study progressed the participants began to become habitualised over a period of time. By using HRV metrics another level of information and detail can be gathered, which builds a more informed picture of the interaction and experiences that the participants were having when they interacted with the interface. In Study Three the findings that were able to be achieved in a laboratory environment have also been seen in a 'real' driving situation (Mehler, Reimer and Wang 2011), where the HRV indices decreased as the participant were involved in tasks. Although all the participant groups had a decrease in HRV, which could be linked to an increase in mental workload, the group that were using the poor interface had a far greater decrease in their HRV powers. From the other findings, such as visual tracking errors, response times and subjective scales, one could suggest that the interactions the participants were having with the poor interface was causing them to be very frustrated and it was this frustration that was being detected by the HRV metrics.

Results from the third study highlighted that when using a poor interface, emotions, user experience and cognitive processes are all affected and can have a cascade effect, which in a 'world' environment could lead to serious consequences. When a participant is faced with an unattractive interface, their perceived usability of the interface is reduced, which leads to a difficulty in forming a mental model of the system (Norman 2013). By not forming a proper mental model, an individual may not be able to perform or navigate successfully through the intended tasks of the system as they wished. The third study also demonstrated that by using too many similar modalities of information processing tasks (visual in this case), an individual would struggle with resource allocation as a bottleneck in the information processing would occur. This is in-line with the Multiple Resource Theory (MRT) as postulated by Wickens (1984), whereby cognitive demand along one modality results in an increase in mental workload and an associated difficulty in completing the task. As a result of this the participant is more likely to become overloaded and consequently becomes more frustrated in interacting with the interface. This will inevitably lead to an increasing likelihood of human error (Wickens *et al.* 1998). The automotive research community are always seeking to find new ways to lessen the distraction form the interface and allow the driver to maintain their focus on the road. However, as earlier discussed this is a difficult balance for designers and engineers to achieve as customers want to see the latest technologies and gadgets they have in their home environment, to be available in their car.

The close connection between emotion, cognitive processes and the autonomic nervous system, enable certain physiological measurements to reflect these processes and provide an insight into these complex human states. By selecting HRV frequency indices, the findings in study three highlight that physiological measurements that reflect the autonomic nervous system may afford us the ability to observe such phenomenon. As participants attended to the interface tasks we see that their HRV levels decreased.

Additional HRV indices in Study Three were calculated to add another layer of information associated with both the autonomic nervous system and the affective system. Although HRV provides detail on the heart dynamics it is the association with the autonomic system and how that association is linked to the affective system and provokes emotions to different stimuli. Neural studies, including functional magnetic resonance imaging (fMRI) and positron emission tomography (PET) have shown an association between the vagal activity (HF HRV) and different neural structures. Critchley *et al.* (2003) and Gianaros *et al.* (2004) investigated cognitive tasks that were emotionally stressful and in both studies found different areas of the brain were activated and correlated with HRV activity. Neumann *et al.* (2006) found an inverse relationship between resting LF/HF ratio and dorsal activity in the anterior cingulate cortex (ACC) during a go/no-go task. The above studies highlight the relevance of vagal activity to neural processing of cognitive and emotional stimuli.

5.1.9 Limitations

When planning a complex study that includes many different components, the aim is always to produce a study that has a robust methodology, is valid and is repeatable. It is important to reflect on these key elements after the study has finished as improvements for future studies is important. The key limitations in Study Three were, firstly the number of groups and the size of the groups. Running physiological studies can be problematic as individual variation, especially in the cardiovascular indices can be large, and therefore if sample numbers are not sufficient, this can impact on the significant results, that were perhaps lacking in Study Three. Although sample size can always be larger, the allocation of groups in this study contributed to this problem. Secondly, it is worth considering that participants across the two groups (Academic versus Automotive) involved a change in laboratory environment, due to the different locations required to set-up the study. Another consideration would be that by trying to tease out elements of user experience, such as frustration, the study design had to include workload elements, which may have had an overlap into the performance and physiological measures.

5.1.10 Conclusion

Participants in Study Three engaged in an experiment that assessed their user experience when they interacted with a HMI system. By using HRV and other physiological indices, it was hoped that an additional layer of information could provide an interesting insight into their physiological state while they were participating in the experiment. The important link, which should not be forgotten, is the relationship that the selected physiological measures have with the ANS, and it is this relationship that plays an important part in emotions and how an individual perceives their experiences. It was also important in the study's wider context that the continued use of subjective measures and self-report scales be used and studied. In the evaluation process of HMIs self-report measures are used frequently. The above studies highlight the relevance of vagal activity to neural processing of cognitive and emotional stimuli. Study Three has enabled emotional and workload behaviours to be subjectively and objectively measured. By selecting physiological indices that hold a strong connection with the autonomic nervous system, which in turn has a key role in emotion and cognition, an insight into this complex process has been gained.

6 Chapter Six: General Discussion

In order to maintain competitive advantage, automotive manufacturers need to ensure that the HMI systems they introduce into their next generation automobiles are highly desirable. This would mean that the visual display would not only have a great deal of aesthetic appeal but also possess a high degree of usability. The usability of an in-car system will just as likely turn a potential customer away from purchasing future products from the manufacturer, more so if the experience it evokes leads to frustration and higher workload. It is therefore, important to assess not only how effectively an individual interacts with the automotive display, but also the emotional response involved in such interaction. As previously stated automotive manufacturers have traditionally tended to use subjective assessments to investigate these issues, while the potential of adopting the discussed physiological measures could potentially offer a more robust (and objective) means of evaluation and feedback.

The studies conducted in this thesis were designed to test the research hypothesis that physiological measures could be utilised (in some cases alongside subjective assessment) as providing feedback to the design of automotive HMIs. In order to address this, it was important to establish which, if any, physiology indices were able to be used to provide assessment of HMI usability. Thus in Study One the use of the HRV metric was used alongside validated stimuli that was proven to evoke an emotional response. The ANEW noun list was shown to provoke an emotional response in terms of physiological measurements and the nature of methodological design was also examined. The rationale behind the design of this study was enhanced for Study Two, where the move from presenting words was replaced with visual stimuli, as this was thought to be closer to the sort of cognitive processing associated with visual displays. Again, a validated stimuli set was used to present participants with differing emotional saliency. At the end of these two studies it was clear that physiological sequelae were able to be identified when stimuli (with varying levels of emotional content) could trigger a physiological response. This allowed the design of the final study, Study Three, whereby a number of automotive designs were tested within a realistically-framed task environment. Arguably Studies One and Two are not representative of an in-car environment, and thus the final study used the previous studies as a foundation for asking whether usability issues (such as frustration) could be identified within an environment where participants were asked to interact with different automotive HMIs (infotainment displays) akin to what they would normally do when they were in a car.

Examining the literature in this area, it was evident that user experience, emotions and cognitive processes are highly interconnected when considering automotive HMIs. Due to the nature of the driving task, with safety aspects being central to any in-car system, all of the above processes have a high priority. It was evident from the literature that the face of automotive development is rapidly changing with creative opportunistic technology companies seeing the automotive industry, and in particular, the next generation in-car systems as a target for their next technology designs. There are already some examples where gaming creators are beginning to think about how their interactive technologies could be adapted for the in-car environment (Sohdi 2013). With the growing focus on producing autonomous cars, with both the driver and their passengers being more connected, it will be interesting to see if these gaming designers see the car environment as a lucrative opportunity. The major impact that autonomous and connected cars will have in the near future will not only impact on the usability aspect, but of equal importance will be the trust and safety element related to interacting with the system (Richards and Stedmon 2016).

In Study Three it was the accumulation of findings that highlighted the difficulties that participants were having during interactions with an intentionally poorly designed interface. Opinions on the aesthetics of the interface formed quickly and their feelings relating to the usability and user experience were also poor. Research outside the automotive HMI domain has also demonstrated that when aesthetics were deemed poor, the perception and actual usability qualities of a device or system were also regarded as low. Automatic Teller Machines (ATMs) that were identical in function, but for which the displays were perceived as being more or less attractive then users perceived that the more attractive display would also be easier to operate than the unattractive one. (Kurosu and Kadhimura 1995).

For automotive companies the goal is simple, design vehicles that customers desire. Unfortunately, although true, this simplistic statement is harder to achieve. The automotive industry is highly competitive and as the in-car HMI systems are recognised as a key differentiator when a customer is purchasing a car, the value of producing a novel, exciting and intuitive HMI is extremely crucial for the automotive manufacturers. For designers and engineers this is a difficult balance as often creative and futuristic interfaces that may be produced are often not transferable to the environment inside a car, where safety of the driver, passengers and pedestrians are always the first consideration. It is important to remember that any distraction from the primary task of driving entails a risk element.

To ensure that a new in-car HMI system is desired, designers and engineers need to make sure that the person using the system has a positive experience when they interact with the HMI. There are many different components that contribute to this positive experience, such as aesthetics and usability of the system. If a person connects to an object or system due to its attractive design, the person will perceive that the usability of the object or system will be good. A later study by Lavie, Oron-Gilad and Meyer (2011) performed a study to explore perceived aesthetics and usability with in-vehicle navigation maps and supported the connections between these two components. However, this study also concluded that usability evaluations were not always linked to the user's actual performance and indicated justification for running objectives measures at the same time. This is an important finding as it highlights the need for attention when using subjective assessments of usability (Lavie, Oron-Gilad and Meyer 2011). At times our perception of performance can be misleading, depending on complex factors relating to our experience and indeed how our mental models are constructed (Johnson-Laird 1983).

When considering the use of different metrics for evaluating HMI systems within the automotive industry it is vital that the data is a true reflection of human performance, rather than a subjective perception that may be formed as a basis of past experience (residing as a schema within the individual's mental model). The studies outlined in this research are not meant as an attack on subjective assessment or neither a statement that detecting changes in physiology is the single answer when investigating usability or user preference for one display over another. Rather, these studies have shown how both objective and subjective metrics can be used together in order to explore (and in some instances validate) user perceptions of displays that possess different aesthetic qualities.

Driving a vehicle is the most universal application of mechanical technology but the days of a clutch and throttle as the sole and dominant knowledge requirements are at an end. We move now into a new world of advanced technology. Designers and engineers must base these developments on a clear understanding of user response and interactions, this is essential for safety and thereby direct the lanes for morally correct and progressive investment.

6.1.1 Future Research

There is a great opportunity for future research within the automotive industry to adopt the application of physiological measurements to explore and determine drivers' internal body states that have an important impact on the way a driver interacts with the automotive HMI. By building a clearer picture of an individual's physiological parameters during HMI interaction evaluation studies, it should be conceivable that analysis of physiological data could provide the automotive design and engineering teams a real insight into how human-system interactions impact on the driver and ultimately how that driver behaves whilst driving.

HRV metrics were used in this thesis to develop an initial understanding of how the autonomic nervous system behaves and captures the experiences that the participants were having during the usability study. It would be therefore interesting and worthy to develop additional studies that focus wholly on the user experience when interacting with an HMI without having the distraction of workload elements within the study. Due to the high degree of individual variability in HRV metrics, it would be beneficial to run larger scale, repeatable studies.

As the era of autonomous cars is approaching, there is an urgency for automotive manufacturers to understand the complexities of the human-system interaction, and physiological monitoring can assist in bridging the gap between objective and subjective evaluation. This will be critical when the feedback and data captured from a user evaluation trial is being used to steer the design of the next generation HMI systems.

References

Aasman, J., Mulder, G., and Mulder, L. J. M. (1987) 'Operator effort and the measurement of heart rate variability'. *Human Factors* 29 (2), 161-170

Abercrombie, H., Chambers, A.S., Greischar, L., and Monticelli, R. M. (2008) 'Orienting, emotion, and memory: Phasic and tonic variation in heart rate predicts memory for emotional pictures in men'. *Neurobiology of Learning and Memory* 90(4), 644-650

Akselrod, S., Gordon, D., Ubel, F. A., Shannon, D. C., Barger, A. C., and Cohen, R. J. (1981) 'Power spectrum analysis of heart rate fluctuations: a quantitative probe of beat-to-beat cardiovascular control'. *Science* 213(4504), 220–222

Allen, J., Armstrong, J., and Roddie, I. (1973) 'The regional distribution of emotional sweating in man'. *Journal of Physiology* 235 (3), 749-759

Amditis, A., Andreone, L., Pagle, K., Markkula, G., Deregibus, E., Rue, M., Bellotti, F., Engelsberg, A., Brouwer, R., Peters, B., and De Gloria, A. (2010) 'Towards the Automotive HMI of the Future'. *IEEE Transactions on Intelligent Transportation Systems* 11(3), 1-12

Amditis, A., and Polychronopoulos, A. (2004*) Communication and interaction strategies in automotive adaptive interfaces. '*Conference on Systems, Man and Cybernetics'*.* held 10-13 October *2004 at* The Hague, Netherlands: Springer-Verlag

Amrhein, P.C., McDaniel, M.A., and Waddill, P. (2002) 'Revisiting the picture-superiority effect in symbolic comparisons: do pictures provide privileged access?' *Journal of Experimental Psychology: Learning Memory and Cognition 28(5), 843-857*

Anderson, J.R. (2005) *Cognitive Psychology and its Implications.* San Francisco: W H Freeman

Appel, M.L., Berger, R.D., Saul, J. P., Smith, J.M., Cohen, R.J. (1989) 'Beat to beat variability in cardiovascular variables: noise or music?'. *Journal of the American College of Cardiology* 14(5) ,1139-1148

Apple (2014) *Apple Rolls Out CarPlay Giving Drivers a Smarter, Safer & More Fun Way to Use iPhone in the Car* [online] available from <http://www.apple.com/uk/pr/library/2014/03/03Apple-Rolls-Out-CarPlay-Giving-Drivers-a-Smarter-Safer-More-Fun-Way-to-Use-iPhone-in-the-Car.html> [5 May 2016]

Arnell, K.M., Killman, K.V., and Fijavz, D. (2007) 'Blinded by emotion: target misses follow attention capture by arousing distractors in RSVP'. *Emotion* 7(3), 465-477

Autocar, (2016) *Autocar First for car news and reviews* [online] available from <http://www.autocar.co.uk/> [15 May 2016]

Baddeley, A. (2012) 'Working memory, theories models and controversy'. *The Annual Review of Psychology 63* ,1-29

Bargas-Avila, K.A., and Hornbæk, K. (2011) *Old Wine in New Bottles or Novel Challenges? A Critical Analysis of Empirical Studies of User Experience.* 'Conference on Human factors in computing systems'. held 7-12 May 2011 at Vancover BC

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., and Plumb, I. (2001) 'The Reading the Mind in the Eyes" test revised version: a study with normal adults, and adults with Asperger syndrome or high-functioning autism'*. Journal of Child Psychology and Psychiatry* 42(2), 241–251

Ben-Bassat, T., Meyer, J., and Tractinsky, N. (2006) 'Economic and subjective measures of the perceived value of interface aesthetics and usability'. *Transactions on Computer-Human Interaction* 13(2),210-234

Berntson, G.G., Quigley, K.S., and Lozano, D. (2007) 'Cardiovascular Psychophysiology'. in *Handbook of psychophysiology.* ed. by Cacioppo, J., Tassinary, L., and Bernstons, G. Cambridge: Cambridge University Press, 182-210

Berntson, G.G., and Cacioppo, J. T. (2004) 'Heart rate variability: Stress and psychiatric conditions'.in *Dynamic Electrocardiography*. ed.by Malik, M., and Camm, A.J. New York: Blackwell Futura, 57-64

Berntson, G.G., and Cacioppo, J.T., Eckberg, D.L., Grossman, P., Kaufmann, P.G., Malik, M., Nagaraja, H.N., Porges, S.W., Saul, J.P., Stone, P.H., and Van-derMolen, M.W. (1997) ‘Heart rate variability: Origins, Methods, and Interpretive caveats’. *Psychophysiology* 34(6), 623-648

Bigger, J. T., Fleiss, J., Steinman, R. C., Rolnitzky, L. M., Kleiger, R. E., Rottman, J. N. (1992) ‘Frequency domain measures of heart period variability and mortality after myocardial infarction’. *Circulation* 85, 164–171

Billman, G. E. (2013) ‘The effect of heart rate on the heart rate variability response to autonomic interventions’. *Frontiers in Physiology* 4 ,222

Biopac (2016) BIOPAC Systems, Inc [online] available from < http://www.biopac.com/> [08 June 2016]

Birbaumer, N and Ohman, A. (1993) *The structure of emotion: Psychophysiological, cognitive and clinical aspects.* Toronto: Hogrefe and Huber

Birch, L.S., Duncan, J.M. and Franklin, C. (2012) ‘Overweight and reduced heart rate variability in British Children: an exploratory study’. *Preventive Medicine* 55(5), 430-432

Boduroglu, A., Shah, P., and Nisbett, R. (2010) ‘Cultural Differences in Allocation of Attention in Visual Information Processing’. *Journal of Cross Cultural Psychology* 40(3), 349-360

Bonner, M.A., and Wilson, G.F. (2001) 'Heart Rate Measures of Flight Test and Evaluation'. *The International Journal of Aviation psychology* 12(1), 63-77

Boucsein, W. (2012) *Electrodermal Activity*. 2nd edn. New York: Springer

Bradley, M. M., and Lang, P. J. (2007a) 'Affective Norms for English Text (ANET): Affective ratings of text and instruction manual'. *Technical Report No. D-1* The Centre for Research in Psychophysiology University of Florida: Gainesville

Bradley, M., and Lang, P. (2007b) 'The International Affective Digitized Sounds (2nd Edition; IADS-2): Affective Ratings of Sounds and Instruction Manual'. *Technical Report No. B-3* The Centre for Research in Psychophysiology University of Florida: Gainesville

Bradley, M. M., and Lang, P. J. (1999a) 'International affective digitized sounds (IADS): Stimuli, instruction manual and affective ratings'. *Technical Report No B-2* The Centre for Research in Psychophysiology, University of Florida: Gainesville

Bradley, M.M., and Lang, P.J. (1999b) 'Affective Norms for English Words (ANEW): Instruction manual and affective ratings'. *Technical Report C-1* The Centre for Research in Psychophysiology, University of Florida: Gainesville

Bradley, M. M., and Lang, P. J. (1994) 'Measuring emotion: The self-assessment manikin and the semantic differential'. *Journal of Behavioural Therapy and Experimental Psychiatry* 25(1), 49-59.

Braithwaite, J., Watson, D., Jones, R., and Rowe, M. (2013) ‘A guide for analysing electro dermal activity (EDA) & skin conductance responses (SCRs) for psychological experiments’. *Psychophysiology* 49 (1) ,1017-1034

Broadbent, D. (1958) *Perception and Communication.* London: Pergamon Press

Brook, R. (1986) ‘A Robust Layered Control System for a Mobile Robot’. *IEEE Journal of Robotics and Automation* 2(1), 14-23

Brooke, J. (1986) *System usability scale (SUS): a quick-and-dirty method of system evaluation user information*. Reading, UK: Digital Equipment Co Ltd

Brookhuis, K. A., and De Waard, D. (2001) ‘Assessment of drivers' workload: Performance and subjective and physiological indexes’. in Stress, Workload, and Fatigue.ed.by Hancock, P.A., and Desmond, P.A. Mahwah, NJ: Lawrence Erlbaum Associates, 321-333

Brookhuis, K.A., De Vries, G., and De Waard, D. (1991). ‘The effects of mobile telephoning on driving performance’. *Accident Analysis and Prevention* 23(4), 309-316

Brown, T.M., and Fee, E. (2002) ‘Walter Bradford Cannon: Pioneer Physiologist of human emotions’. *American Journal of Public Health* 92(10), 1594-1595

Brown, J. D., and Huffman, W.J. (1972) ‘Psychophysical measures of drivers under actual driving conditions’. *Journal* of *Safety Research* 4(4), 172-178

Bublatzky, F., Flaisch, T., Stockburger, J., Schmälzle, R., and Schupp, H.T. (2010) ‘The interaction of anticipatory anxiety and emotional picture processing: an event-related brain potential study’. *Psychophysiology* 47(4), 687-696

Burnett, G., Crundall, E., Large, D., and Lawson, G. (2013) *A study of Unidirectional swipe gestures on In-Vehicle Touch Screens*. ‘Proceedings of the 5th International Conference on Automotive User Interfaces and Interactive Vehicular Applications’. held 27-30 October 2013 at Eindhoven, The Netherlands. New York: ACM Press

Burnout Protector (2016) *BURNOUT Risk-Test* [online] available from <https://www.burnoutprotector.com/> [08 December 2016]

Cahill, L., Prins, B., Weber, M., and McGaugh, J.L. (1994) ‘β-Adrenergic activation and memory for emotional events’. *Nature* 371(6499),702–704

Carsten, O.M.J., and Nilsson, L. (2001) ‘Safety Assessment of Driver Assistance Systems’. *European Journal of Transport and Infrastructure Research* 1(3), 225-243

Choi, I., Choi, J., and Norenzayan, A. (2004) ‘Culture and Decisions’. In *Handbook of Judgment and Decision Making*. ed.by Koehler, D.J., and Harvey, N. Oxford, UK: Blackwell, 504-524

Cisco (2013) Cisco *Customer Experience Research, Automotive Industry Global Data* [online] available from <http://www.cisco.com/c/dam/en_us/about/ac79/docs/ccer_report_manufacturing.pdf> [5 May 2016]

Clark, J. M., and Paivio, A. (1991) 'Dual coding theory and education. *Educational Psychology Review* 3(3), 149-210

Coughlin, J.F., Reimer, B., and Mehler, B. (2009) 'Driver Wellness, Safety & the Development of an AwareCar'. *MIT, Centre for Transportation and Logistics*, 1-15

Craik, K. (1943) *The Nature of Explanation*. Cambridge: Cambridge University Press

Critchley, H.D., Josephs, O., O'Doherty, J., Zanini, S., Dewar, B.K., Mathias, C.J., Cipolotti, L., Shallice, T., and Dolan, R.J. (2003) 'Human cingulate cortex and autonomic cardiovascular control: converging neuroimaging and clinical evidence'. *Brain* 126(10) ,2139-2152

Cunningham, J., and Klein, B. (2007) *Textbook of Veterinary Physiology.* 4th edn. Missouri: Saunders Elsevier

Damasio, A. R. (1994) *Descartes' Error: Emotion Reason and the Human Brain.* New York: GP Putnam's Sons

Dawson, M., Schell, A., and Filion, D. (2007) '7 The Electrodermal System'. in *Handbook of psychophysiology.* ed. by Cacioppo, J., Tassinary, L., and Bernstons, G. Cambridge: Cambridge University Press, 159-181

De Rivecourt, M., Kuperus, M. N., Post, W. J., and Mulder, L. J. M. (2008) 'Cardiovascular and eye activity measures as indices for momentary changes in mental effort during simulated flight.' *Ergonomics* 51(9), 1295-1319

De Warrd, D., Steyvers, F. J. J. M., Brookhuis, K.A. (2004) 'How much visual road information is needed to drive safely and comfortably?'. Safety Science 42(7), 639-655

Deutsch, A., and Deutsch, D. (1963) 'Attention: Some theoretical considerations'. *Psychological review* 70(1), 80-90

Dolcos, F., Kragel, P., Wang, L., and McCarthy, G. (2006) 'Role of the inferior frontal cortex in coping with distracting emotions'. *NeuroReport* 17(15),1591–1594

Duschek, S., Heiss, H., Werner, N., and Reyes, P.G.A., (2009) 'Modulations of autonomic cardiovascular control following acute alpha-adrenergic treatment in chronic hypotension'. *Hypertension Research* 32(11), 938–943

Eckberg, D.L. (1997) 'Sympathovagal balance: a critical appraisal.' *Circulation* 98(23), 3224-3232

Eggemeier, F. T., and Wilson, G.F. (1991) 'Performance-based and subjective assessment of workload in multi-task environments'. in *Multiple-task Performance*. ed. by Damos, D.L. London: Taylor and Francis, 207-216

Ellis, R. A. (1967) 'Eccrine, sebaceous and apocrine glands'. in Ultrastructure of normal and abnormal skin. ed.by Zelickson, A.S. Philadelphia: Lea and Febiger,132-162

Embrey, D., Blackett, C., Marsden, P., and Peachey, J. (2006) *Development of a Human Cognitive Workload Assessment Tool: MCA Final Report [online] Lancashire: Dalton, Human Reliability Associates. available from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.138.4767&rep=rep1&type=pdf> [15 May 2016]*

Endsley, M.R. (1995) 'Toward a theory of situation awareness in dynamic systems'. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 37(1), 32-64

Endsley, M.R. (1988) *Situation awareness Global Assessment Technique (SAGAT).* 'Proceedings of the National Aerospace and Electronics Conference (NAECON)' held 23-27 May 1988 at Dayton, OH . New York: IEEE ,789-795

English, B.A., and Jones, C. K. (2012) 'Cholinergic Neurotransmission'. in *Primer on the Autonomic Nervous System*. 3rd edn. ed. by Robertson, D., Biaggioni, I., Burnstock, G., Low, P.A., and Paton, J.F.R. London: Elsevier, 71-74

Engstrom, J., Johansson, E., and Ostlund, J. (2005) 'Effects of visual and cognitive load in real stimulate motorway driving'. *Transportation Research Part F: Traffic Psychology and Behaviour* 8(2), 97-120

Ernest, G. (2014) *Heart Rate Variability*. New York: Springer

Forlizzi, J., and Battarbee, K. (2004) *Understanding experience in interactive systems*. 'DIS 04 *Proceedings of the 5th Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques'.* held 1-4 August 2004 at Cambridge MA. New York: ACM Press

Forsstrom, J., Heinonen, E., Valimaki, I., and Antila, K. (1986) 'Effects of haemodialysis on heart rate variability in chronic renal failure'. *Scandinavian Journal of Clinical and Laboratory Investigation* 46(7), 665-70

Fox, S.I. (2011) *Human Physiology*. New York: McGraw-Hill International

Foyle, D.C., Dowell, S.R., and Hooey, B.L. (2001) *Cognitive tunnelling inhead-display (HUD) Superimposed symbology: Effects of information location*. 'Proceedings of the 11th International symposium on Aviation Psychology'. held 5- 8 March 2001 at Columbus, Ohio: Ohio State University Department of Aviation and Aeronautical Engineering

Franchini, K.G., and Cowley, A.W. (2012) 'Autonomic control of cardiac function'. In *Primer on the autonomic nervous system*. 2nd edn. ed.by Roberson, D. Amsterdam/Boston: Elsevier/AP,134-138

Frenneaux, M.P. (2004) 'Autonomic changes in patients with heart failure and in post-myocardial infarction patients'. *Heart* 90 (11), 1248-55

Frischen, A., Eastwood, J.D., and Smilek, D. (2008) 'Visual search for faces with emotional expressions. *Psychological Bulletin* 134(5),662-672

Furness, J. (2006) 'The organisation of the autonomic nervous system: peripheral connections'. *Autonomic NeuroScience: basic & clinical* 130(1-2), 1-5

Gable, T.M., Walker, B.N., Moses, H.R., and Chitloor, R.D. (2013) *Advanced auditory cues on mobile phones help keep drivers' eyes on the road*. 'Conference on Automotive User Interfaces and Interactive Vehicular Applications'. held 27-30 October 2013 at Eindhoven, The Netherlands. New York: ACM Press

Geisler, F.C.M, Vennewald, N., Kubiak, T., and Weber, H. (2010) 'The impact of heart rate variability on subjective well-being is mediated by emotion regulation'. *Personality and Individual Differences* 49, 723-728

Gentner, D., and Stevens, A.L. (1983) *Mental Models*. Hillsdale, NJ: Lawrence Erlbaum Associates

Gianaros, P.J., Van Der Veen, F., and Jennings, J., (2004) Regional cerebral blood flow correlates with heart period and high-frequency heart period variability during working memory tasks: implications for the cortical and subcortical regulation of cardiac autonomic activity. Psychophysiology 41(4) ,521-530

Gibson, J. (1966) *The Senses Considered as Perceptual Systems.* Oxford: Houghton Mifflin

Gilbert, C.D. (2013) 'The Constructive Nature of Visual Processing'. in *Principles of Neural Science'*. 5th edn. ed.by Kandel, E.R., Schwartz, J.H., Jessell, T.M., Siegelbaum, S.A., and Hudspeth, A.J. New York: McGraw-Hill, 556-576

Gillie, B. L., Vasey, M. W., and Thayer, J. F. (2014) 'Heart rate variability predicts control over memory retrieval'. *Psychological Science.* 25(2), 458-465

Goldberger, J., Challapalli, S., Tung, R., Parker, M., and Kadish, A. (2001) 'Relationship of heart rate variability to parasympathetic effect'.*Circulation 103*(15), 1977-1983

Goldstein, D.S., and Kopin, I.J. (2007) ‘Evolution of concepts of stress’. *Stress* 10(2), 109-120

Gray, S., Zanre, E., and Gray, S. (2014) ‘Fuzzy Cognitive Maps as Representations of Models and Group Beliefs’. in *Fuzzy Cognitive Maps for Applied Sciences and Engineering.* ed. by Papageorgiou, E. Berlin Heidelberg: Springer, 29-48

Green, P*.* (2013) *Standard definitions for driving measures and statistics: overview and status of recommended practice.* ‘Conference on Automotive User Interfaces and Interactive Vehicular Applications’. held 28-30 October 2013 at Eindhoven, Netherlands. New York: AMC Press

Gregory, R. (1970) *The Intelligent Eye*. New York: McGraw-Hill

Gross, J.J., Feldman-Barrett, L. (2011) ‘*Emotion Generation and Emotion Regulation: One or Two Depends on Your Point of View’. Emotion Review 3*(1), 8-16

Haarmann, B., Boucsein, W., and Schaefer, F. (2009) ‘Combining electrodermal responses and cardiovascular measures for probing adaptive automation during simulated flight’. *Applied Ergonomics* 40 (6), 1026-1040

Hackenberg, L., Bongartz, S., Hartle, C., Leiber, Baumgarten, T., and Sison, J.A. (ed) (2013) *International evaluation of NLU benefits in the domain of in-vehicle speech dialog systems*. 'Conference on Automotive User Interfaces and Interactive Vehicular Applications'. held 28-30 October 2013 at Eindhoven, Netherlands. New York: AMC Press

Hagemann, D., Waldstein, S.R., and Thayer, J.F. (2003) 'Central and autonomic nervous system integration in emotion'. *Brain and Cognition* 52(1), 79-87

Hart, S.G., Lowell, E.S. (2008) 'Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research'. *Advances in psychology* 52(0), 139-183

Hasin, D.S., Stinson, F.S., Ogburn, E., and Grant, B.F. (2007) 'Prevalence, correlates, disability , and comorbidity of DSM-IV alcohol abuse and dependence in the United States: Results from the national Epidemiologic survey on Alcohol and related conditions'. *Archives of General Psychiatry* 64(7), 830-842

Hassenzahl, M., and Tractinsky, N. (2006) 'User experience – a research agenda'. *Behaviour and Information Technology* 25 (2), 91–97

Hassenzahl, M. (2005) 'Hedonic, emotional, and experiential perspectives on product quality'. in *Encyclopaedia of Human Computer Interaction*. ed. by Ghauri, C. Hershey, P.A: Idea Group Reference, 266-272

Healey, J. A., and Picard, R. W. (2005) 'Detecting stress during real-world driving tasks using physiological sensors.' *IEEE Transactions on Intelligent Transportation Systems* 6 (2) ,156-166

Healey, J. (2000) *Wearable and automotive systems for affect recognition from physiology*. Unpublished Ph.D. dissertation. Massachusetts: Massachusetts Institute of Technology

Hellard, P., Guimaraes, F., Avalos, M., Houel, N., Hausswirth, C., and Toussaint, J.F. (2011) 'Modelling the association between HR variability and illness in elite swimmers'. *Medicine and Science in Sports and Exercise*, *American College of Sports Medicine* 43(6),1063-70

Henry, I. (2015) 'The future of UK automotive manufacturing in 2025 and beyond'. *The Society of Motor Manufactures and Traders* ,1-23

Hilburn, B., Jorna, P., Byrne, E., and Parasuraman, R. (1997) 'The effect of adaptive air traffic control (ATC) decision aiding on controller mental workload'. In *Human-automation interaction: Research and practice*. ed. by Mouloua, M., and Koonce, J. Hillsdale: CRC Press, 84-91

Holmes, E.A., Geddes, J.R., Colom, F., and Goodwin, G.M. (2008) 'Mental imagery as an emotional amplifier: Application to bipolar disorder'. *Behaviour Research and Therapy* 46(12), 1251-1258

Hon, E.H., Lee, S.T. (1963) 'Electronic evaluation of fetal heart rate'. American Journal of Obstetrics and Gynaecology 15 (87) 814-826

Horn, J. and Swanson, I. (2013) 'The autonomic nervous system and the hypothalamus'. in *Principles of Neural Science'*. 5th edn. ed.by Kandel, E.R., Schwartz, J.H., Jessell, T.M., Siegelbaum, S.A., and Hudspeth, A.J. New York: McGraw-Hill, 1056-1078

Hsia, J., Larson. J.C., Ockene, J.K., Gloria E Sarto, G.S., Allison, M.M., Hendrix, S.L., Robinson, J.G., LaCroix, A.Z., and Manson, J.E. (2009) 'Resting heart rate as a low tech predictor of coronary events'. *Women's Health Initiative Research Group BMJ* 338 ,(b)219

Huang, H.H., Chan, H.L, Lin, P.L., Wu, C.P., Huang, C.H. (1997) 'Time frequency spectral analysis of heart rate variability during induction of general anaesthesia'. *British Journal of Anaesthesia* 79(6),754-758

Hubel, D., and Wiesel, T. (1959) 'Receptive fields of single neurons in the cat's striate cortex'. *The Journal of Physiology* 148 (3), 574–591

Huey, B., and Wickens, C. (1993) *Workload Transition Implications for Individual and Team Performance.* Washington DC: National Academy Press

Hunn, B.P., and Camacho, M.J. (1999) 'Pilot performance and anxiety in a high-risk flight test environment'. *Proceeding of the Human Factors and Ergonomics Society Annual Meeting* 43 *(*1), 26-30

Hyman, S.E., and Cohen, J.D. (2013) 'Disorders of thought and volition: Schizophrenia'. in *Principles of Neural Science'*. 5th edn. ed.by Kandel, E.R., Schwartz, J.H., Jessell, T.M., Siegelbaum, S.A., and Hudspeth, A.J. New York: McGraw-Hill, 1402-1424

Ihssen, N., and Keil, A (2009) 'The costs and benefits of processing emotional stimuli during rapid serial visual presentation'. *Cognition and Emotion* 23 (2),296–326

Iversen, J.A., Tencer, H.L., Lany, J., and Goldin-Meadow, S. (2000b) 'The relation between gesture and speech in congenitally blind and sighted language-learners'. *Journal of Nonverbal Behaviour* 24(2),105-130

Iverson, S., Iverson, L., and Saper, C.B. (2000a) 'The autonomic nervous system and the hypothalamus'. in *Principles of Neural Science'.* 4th edn. ed.by Kandel, E.R., Schwartz, J.H., Jessell, T.M., Siegelbaum, S.A., and Hudspeth, A.J. New York: McGraw-Hill, 960-981

Izard, C.E. (2006) *Experts' definitions of emotion and their ratings of its components and characteristics.* Unpublished manuscript. Newark, DE: University of Delaware

Jaguar Land Rover Limited, 'F – TYPE' [online] available from <F_TYPE_16MY_MB_UKE_LOW_RES_tcm91-201473.pdf> [10th May 2015]

Jennings, J.R., Kamarck, T., Stewart, C., Eddy, M., and Johnson, P. (1992) 'Alternate cardiovascular baseline assessment techniques: vanilla or resting baseline'. *Psychophysiology* 29(6), 742-750

Jennings, J.R., and Hall, S.W., (1980) 'Recall, Recognition and rate: Memory and the heart.' *Psychophysiology* 17(1), 37-46

Johnson-Laird, P. (2013) 'Mental models and cognitive change'. *Journal of Cognitive Psychology* 25(2), 131-138

Johnson, L.C., and Lubin, A. (1966) 'Spontaneous electro dermal activity during walking and sleeping'. *Psychophysiology* 3(1), 8-17

Johnston, E., and Olson, L. (2015) *The Feeling Brain: The Biology and Psychology of Emotions.* New York: W.W. Norton & Company

Johnson-Laird, P.N. (1983) *Mental Models: Towards a cognitive Science of language, Inference and consciousness.* Cambridge, MA: Harvard University Press

Jung, R.M., and Willumeit, H.P.(ed) (2000) *Objective evaluation of the complexity of usage for car infotainment systems*.' Conference on Information Technology for Balanced Automation Systems in Manufacture and Transportation: Advanced Network Enterprises, Virtual Organizations, Balanced Automation, and Systems Integration'. held 27-29 September 2000 at Deventer, The Netherlands: AMC Press

Just, M.A., Carpenter, P.A. (1985) 'Cognitive Coordinate Systems: Accounts of Mental Rotation and Individual Differences in Spatial Ability'. *Psychological Review* 92(2),137-171

Kahneman, D. (1973) *Attention and Effort Book*. New Jersey: Prentice-Hall

Kamath, V., Watanabe, M.A., and Upton, A.R.M. (2013) *Heart Rate Variability (HRV) Signal Analysis: Clinical Applications*. Boca Raton, FL: CRC Press Taylor & Francis Group

Kamath, M.V., Fallen, E.L. (1993) 'Power spectral analysis of heart rate variability: A noninvasive signature of cardiac autonomic function'. *Critical Reviews in Biomedical Engineering* 21(3), 428-434

Kantowitz, B., and Knight, J. (1976) 'Testing Tapping Timesharing, II: Auditory Secondary Task'. *Acta Psychologica* 40 (5), 343-36

Karapanos, E., Zimmerman, J., Forlizzi, J. and Martens, J. (2009) *User Experience Over Time. An Initial Framework*. *'Conference on Human Factors in Computing Systems*. ed. by Mackay, W., Brewster, S., and Bodker, S. Boston: ACM Press, 729 -738

Katsis, C.D., Katertsidis, N.S., and Fotiadis, D.I. (2011) 'An integrated system based on physiological signals for the assessment of affective states in patients with anxiety disorders'. *Biomedical Signal Processing and Control* 6(3), 261-268

Keenan, D Grossman, P. (2006) 'Adaptive Filtering of Heart Rate Signals for an improved measure of Cardiac Autonomic Control'. *International Journal of Information and Signal Processing* 2 (1), 52-58

Kemp, A.H., and Quintana, D.A. (2013) 'The relationship between mental and physical health: Insight from the study rate variability'. *International Journal of Psychophysiology* 89(3), 288-296

Kemp, A.H., Quintana, D.S., Felmingham, K.L., Matthews ,S., and Jelinek, H.F. (2012) 'Depression, comorbid anxiety disorders and heart rate variability in physically healthy, unmediated patients : implications for cardiovascular risk'. *PLos ONE* 7(2), E30777

Kemp, A.H., Quintana, D.S., Gray, M.A., Felmingham, K.L., Brown, K., and Gatt, J.M. (2010) 'Impact of depression and antidepressant treatment on heart rate variability: a review and meta-analysis'. *Biological Psychiatry* 67(11), 1067-1074

Kensinger, E.A., and Schacter, D.W (2006) 'Processing emotional pictures and words: Effects of valence and arousal'. *Cognitive, Affective and Behavioural Neuroscience* 6 (2), 110-126

Kensinger, E.A., and Corkin, S. (2003) ‘Memory enhancement for emotional words: Are emotional words more vividly remembered than neutral words?’. *Memory & Cognition* 31(8),1169-1180

Kern, M. (2014)*This ingenious touchscreen UI should be in every car* [online] available from <http://www.wired.com/2014/02/watch-ingenious-touchscreen-UI-inside-every-car> [15 May 2016]

Khandoker, A.H., Karmakar, C., Brennan, M., Voss, A., and Palaniswami, M. (2013) *Poincare Plot Methods for Heart Rate Variability Analysis*. New York: Springer

Kleiger, R.E., Miller, J.P., Bigger, J.T., Moss, A.J.(1987) ‘Decreased heart rate variability and its association with increased mortality after acute myocardial infarction’. *American Journal of Cardiology* 59(4) ,256-262

Kleinginna, P. R., and Kleinginna, A. M. (1981) ‘A categorized list of emotion definitions, with suggestions for a consensual definition’. *Motivation and emotion* 5(4), 345–379

Kopin, I.J., Eisenhofer G., and Goldstein, D. (1988) ‘Sympathoadrenal medullary system and stress’. In *Mechanisms of Physical and Emotional Stress: Advances in experimental medicine and biology 245*. ed.by Chrouso, L.D., Loriaux, D., and Gold, P.W. New York: Plenum Press, 11-23

Kramer, A. E. (1991) ‘Physiological metrics of mental workload: a review of recent progress’. In *Multiple-task performance*. ed.by Damos, D.L. London: Taylor & Francis ,279-328

Kramer, D.A. (1990) ‘Conceptualizing wisdom: the primacy of affect-cognition relations’. In *Wisdom: Its nature, origins and development* .ed.by Sternberg ,R. Cambridge: Cambridge University Press, 279- 309

Krants, J.H. (2014) Cognition Laboratory Experiments [online] available from <http://psych.hanover.edu/JavaTest/CLE/Cognition/Cognition.html> [15 May 2016]

Kreibig, S.D. (2010) ‘Autonomic nervous system activity in emotion: A review’. *Biological Psychology* 84(3), 394-421

Kuno, Y. (1956) *Human Perspiration*. Springfield, IL: Charles C. Thomas Publisher

Kurosu, M., and Kashimura, K. (1995) *Apparent usability vs. inherent usability: experimental analysis on the determinants of the apparent usability*. ‘CHI ’95 Proceeding Conference Companion on Human Factors in Computing Systems’. Held 7-11 May 1995 at Denver. Colorado. New York: ACM ,292-293

Kuusela, T. (2013) ‘Methodological Aspects of Heart Rate Variability Analysis’. in *Heart Rate Variability (HRV) Signal Analysis*. ed.by Kamath, M.V., Watanabe, M.A., and Upton, A.R.M. London: Taylor and Francis Group, 9-40

Kuwabara, M., and Smith, L.B. (2016) ‘Cultural differences in visual object recognition in 3-year-old children’. *Journal of Experimental Child Psychology* 147(2016) , 22-38

Lanfranchi, P.A., and Somers, V.K., (2002) ‘Arterial baroreflex function and cardiovascular variability: interactions and implications’. *American Journal of Physiology Regulatory, integrative and comparative physiology* 283 (4) ,815–826

Lang, P.J., Bradley, M.M., and Cuthbert, B.N. (2008) 'International affective picture system (IAPS): Affective ratings of pictures and instruction manual.' *Technical Report A-8* The Centre for Research in Psychophysiology, University of Florida: Gainesville, FL.

Lang, J.P., Bradley, M.M., and Cuthbert, B.N. (1998) 'Emotion, Motivation and anxiety: Brain Mechanisms and Psychophysiology'. *Society of Biological Psychiatry* 44(12), 1248-1263

Lang, P.J., Bradley, M.M., and Cuthbert, B.N. (1997) 'International Affective Picture System (IAPS): Technical manual and affective ratings'. *NIMH Center for the Study of Emotion and Attention.*

Lang, P. J. (1980) 'Behavioural treatment and bio behavioral assessment: computer applications'. In *Technology in mental health care delivery system.* ed. by Sidowski, J.B., Johnson, J. H., and Williams, T. A. Norwood, NJ: Ablex, 119-137

Lavie, T., Oron-Gilad, T., and Meyer, J. (2011) 'Aesthetics and usability of in-vehicle navigation displays'. *International Journal of Human-Computer Studies* 69(1),80-99

Lavie, N. (1995) 'Perceptual load as necessary condition for selective attention'. *Journal of Experimental psychology Human perception and performance* 21(3), 451-466

Law, E.L.C., Roto, V., Hassenzahl, M., Vermeeren, A.P.OS., and Kort, J. (2009) *Understanding, scoping and defining user experience: a survey approach*. 'Conference on Human Factors in Computing Systems'. held 4-9 April 2009 at Boston. New York: ACM Press

Leap Motion technology (2014) Leap Motion [online] available from <https://www.leapmotion.com/about-us> [5 May 1016]

LeDoux, J.E., and Damasio, A.R. (2013) 'Emotions and feelings'. In *Principles of Neural Science.* ed. by Kandel, E.R., Schwartz, J.H., Jessell, T.M., Siegelbaum, S.A., and Hudspeth, A.J. New York: McGraw-Hill, 1079–1094

LeDoux, J. (1998) *The Emotional Brain: The Mysterious Underpinnings of Emotional Life*. New York: Simon and Schuster

Lenneman, J.K., and Backs, R.W. (2010) *Enhancing assessment of in-vehicle technology attention demands with cardiac measures*. 'Conference on automotive user interfaces and interactive vehicular applications'. held 11-12 November 2010 at Pittsburgh Pennsylvania. New York: ACM

Levenson, R.W. (2011) 'Basic Emotion Questions'. *Emotion Review* 3(4), 379-386

Levenson, R.W., Soto, J., Pole, N. (2007) 'Emotion, biology, and culture'. In *Handbook of cultural psychology*. ed. by Kitayama, S., and Cohen, D. New York: Guilford Press, 780–796

Lindgaard, G., Fernandes, G., Dudek, C., and Brown, J. (2006) 'Attention web designers: You have 50 milliseconds to make a good first impression!'. *Behaviour & Information Technology* 25 (2), 115-126

Lindholm, E., and Cheatham, C.M. (1983) 'Autonomic activity and workload during learning of a simulated aircraft carrier landing task'. *Aviation, Space and Environmental Medicine* 54(5), 435-439

Lorig, T.S. (2007) ‘The Respiratory System’. In *Handbook of Psychophysiology*. 3rd edn. Ed.by Cacioppo, J.T., Tassinary, L.G., and Berntson, G.G. New York: Cambridge University Press, 231-244

Lykken, D., and Venables, P. (1971) ‘Direct measurement of skin conductance: a proposal for standardization’. *Psychophysiology* 8 (5), 656–672

Maffiolo, V., and Ochs, M. (2012) ‘The Role of Emotions in Human-Machine Interaction’. in *Emotion-Oriented Systems*. ed. by Pelachaud, C. New Jersey: John Wiley & Sons, 215-246

Malik, M., Farrell., Cripps, T., Camm, A.J. (1989) ‘Heart rate variability in relation to prognosis after myocardial infarction: Selection of optimal processing techniques’. *European Heart Journal* 10(12),1060-1074

Malliani, A.,Pagani, M., Lombardi, F., and Cerutti, S.(1991) ‘Cardiovascular neural regulation explored in the frequency domain’. *Circulation* 84(2),482-492

Malpas, S. (2010) ‘Sympathetic nervous system overactivity and its role in the development of cardiovascular disease’. *Physiological Reviews* 90 (2), 513-557

Mandryk, R. L., and Inkpen, K. M. (2004) *Physiological indicators for the evaluation of co-located collaborative play*. ‘Conference on Computer Supported Cooperative Work’. held 6 -10 November 2004 at Chicago Illinois. New York: ACM Press 102-111

Marr, D. (1982) Vison. San Francisco: W.H. Freeman and Co

Martini, F.H. (2006) Fundamentals of Anatomy & Physiology. 7th edn. San Francisco: Pearson, Benjamin Cummings

Mauss, I.B., and Robinson, M.D. (2009) 'Measures of emotion: A review' *Cognition and Emotion* 23(2), 209-237

McCraty, R., Akinson, M., Tiller, W.A., Rein, G., and Watkins, A.D. (1995) 'The effects of emotions on short-term power spectrum analysis of heart rate variability'. *American Journal of Cardiology* 76(14),1089-1093

Mehler, B., Reimer ,B., and Coughlin, J.F. (2012) 'Sensitivity of Physiological Measures for Detecting Systematic Variations in Cognitive Demand from a Working Memory Task: an on-Road Study Across Three Age Groups'. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 54(3), 396-412

Mehler, B., Reimer, B., and Dusek, J.A., (2011a) 'MIT AgeLab Delayed Digit Recall Task'. *MIT AgeLab White Paper*: Cambridge, MA: Massachusetts Institute of Technology, 1-23

Mehler, B., Reimer, B., and Wang, Y. (2011b) 'A comparison of heart rate and heart rate variability indices in distinguishing single task driving and driving under secondary cognitive workload'. *Proceedings of the 6th International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design*'. held 27-30 June 2011 at Lake Tahoe, California. Cambridge, Massachusetts: MIT AgeLab & New England University Transportation Centre

Mehler, B., Reimer, B., and Coughlin, J.F. (2010) *Physiological reactivity to graded levels of cognitive workload across three age groups: An on-road evaluation*. 'Proceedings of the 54th Annual Meeting of the Human Factors and Ergonomics society'. held 27-1 September 2010 at San Francisco. Cambridge Massachusetts: MIT AgeLab & New England University Transport Centre

Mehler, B., Reimer, B., Coughlin, J.F., and Dusek, J.A. (2009) 'The impact of incremental increases in cognitive workload on physiological arousal and performance in young adult drivers'. *Transportation Research Record* 2138(12), 6-12

Mehrabian, A., and Russell, J. (1974) 'Distinguishing anger and anxiety in terms of emotional response factors'. *Journal of Consulting and Clinical Psychology* 42 (1), 79-83

Meister, M., and Tessier-Lavigne, M. (2013) 'Low-level Visual Processing: The Retina'. in *Principles of Neural Science'*. 5th edn. ed.by Kandel, E.R., Schwartz, J.H., Jessell, T.M., Siegelbaum, S.A., and Hudspeth, A.J. New York: McGraw-Hill, 577-600

Memon, A., and Young, M. (1997) 'Desperately seeking evidence: The recovered memory debate'. *Legal and Criminological Psychology* 2 (2), 131-154

Millington, P.F., and Wilkinson, R. (1983) *Skin*. New York: Cambridge University Press

Mitsopoulos-Rubens, E., Trotter, M.J. and Lenne, M.G. (2011) 'Usability evaluation as part of iterative design of an in-vehicle information system'. *IET Intelligent Transport Systems* 5 (2), 112-119

Mohrman, D.E., and Heller, L. J. (2014) *Cardiovascular Physiology*. 8th edn. New York: McGraw-Hill

Montagna, W., and Parakkal, P. (1974) *The Structure and Functioning of Skin*. New York: Academic Press

Montano, N., Ruscone, T.G., Porta, A., Lombardi, F., Pagani, M., and Malliani, A. (1994) 'Power spectrum analysis of heart rate variability to assess the changes in sympathovagal balance during graded orthostatic tilt'. *Circulation* 90(4) ,1826-1831

Moray, N. (1979) 'Models and Measures of Mental Workload.' In *Mental Workload its Theory and Measurement*.ed. by Moray, M. New York: Plenum Press, 13-21

Moshagen, M., and Thielsch, M. (2012) 'A short version of the visual aesthetics of websites inventory'. *Behaviour & Information Technology* 32 (12), 1305-1311

Mousavi, S.Y., Low, R., and Sweller, J. (1995) 'Reducing Cognitive load by mixing auditory and visual presentation modes'*. Journal of Educational Psychology* 87(2), 319-334

Mulder, L. J. (1992) 'Measurement and analysis methods of heart rate and respiration for use in applied environments'. *Biological Psychology* 34(2-3), 205-236

Myers, A., and Hansen, C. (2011) *Experimental Psychology*. 7th edn. USA: Wadsworth

Neumann, S.A., Brown, S.M., Ferrell, R.E., Flory, JD., Manuck, S.B, and Hariri,A.R. (2006) ‘Human choline transporter gene variation I associated with corticolimbic reactivity and autonomic – cholinergic function’. *Biological Psychiatry* 60(10), 1155-1162

Neumann, E., and Blanton, R. (1970) ‘The Early History of Electrodermal Research’. *Psychophysiology* 6 (4), 453-475

NHS choices (2014) *Your Health ,Your Choices* [online] available from <http://www.nhs.uk/pages/home.aspx> [15 May 2016]

Nickholson, M., Hoye, R., and Houlihan, B. (2010) *Participation in sport* Routledge: London

Nielsen Norman Group (2016)*The Definition of User Experience* [online] available from <https://www.nngroup.com/articles/definition-user-experience/> [15 May 2016]

Norman, D.A. (2013) *The Design of Everyday Things: Revised and Expanded.* New York: The MIT press

Norman, D.A. (2004) *Emotional Design : Why we love (or hate) everyday things*. New York: Basic Books

Nunan, D., Sandercock, G. R., and Brodie, D.A. (2010) ‘A quantitative systematic review of normal values for short-term heart rate variability in healthy adults’. *Pacing and Clinical Electrophysiology* 33(11), 1407- 1417

Oatley, K., Jenkins, J.M. (1996) *Understanding emotions*. Oxford: Blackwell Publishing

O'Carroll, R.E., Drysdale, E., Cahill, L., Shajahan, P., and Ebmeier, K.P. (1999) 'Stimulation of the noradrenergic system enhances and blockade reduces memory for emotional material in man'. *Psychological Medicine* 29 (5), 1083-8

O'Donnell, R.D., and Eggemeier, F.T. (1986) 'Workload Assessment Methodology'. in *Handbook of Perception and Human Performance*. ed.by Boff, K.., Kaufman, L., and Thomas, J. New York: Wiley Interscience, 421-429

Open Automotive Alliance (2015) *Introducing the Open Automotive Alliance* [Online] available from <http://www.openautoalliance.net/#about> [5 May 2016]

Osgood, C. (1952) 'The nature and measurement of meaning'. *Psychological Bulletin* 49(3), 172-237

Owens, J., Mclaughlin, S., and Sudweeks, J. (2011) 'Driver Performance while text messaging using handheld and in-vehicle systems'. *Accident Analysis and Prevention* 43 (3), 939-947

Paas, F. (1992) 'Training Strategies for attaining transfer of problem solving skill in statistics: A cognitive load approach'. *Journal of Educational Psychology* 84(4),429-434

Pahlm, O., Sörnmo, L. (1984) 'Software QRS detection in ambulatory monitoring a review'. *Medical & Biological Engineering & Computing* 22(4), 289–297

Paivio, A. (2007) *Mind and its evolution: A dual coding theoretical approach*. Mahwah, NJ: Erlbaum

Pashier, H. (1999) *The psychology of Attention.* Massachusetts institute of technology: Cambridge

Peirce, C.S. (1932) *Collected Papers of Charles Sanders Peirce, Volumes I and II: Principles of Philosophy and Elements of Logic*. Cambridge : MA Harvard University Press

Peirce, J.W. (2009) 'Generating stimuli for neuroscience using PsychoPy'. *Frontiers in Neuroinformatics* 2(10),1-8

Pharmacology Online (2011) *Neurotransmitter chemistry of the Autonomic Nervous system* [online] available from <http://pharmacology-online.blogspot.co.uk/2011/04/neurotransmitter-chemistry-of-autonomic.html> {04 January 2017]

Picard, R., Fedor, S., and Ayzenberg, Y. (2015) 'Multiple Arousal Theory and Daily Life Electrodermal Activity Asymmetry'. *Emotion Review* 8 (1), 62-75

Picard, R.W. (2003) 'Affective computing: challenges'. *International Journal of Human-Computer Studies*, 59(1-2),55-64

Picard, R.W., Vyzas, E., and Healey, J. (2001) 'Toward machine emotional intelligence: Analysis of affective physiological state'. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 23(10), 1175-1191

Piferi, R.L., Kline, K.A., Younger, J., and Lawler, K.A. (2000) 'An alternative approach for achieving cardiovascular baseline: Viewing an aquatic video'. *International Journal of Psychophysiology* 37(2), 207-217

Pinkus, H. (1970) ‘The direction of growth of human epidermis’. *British Journal of Dermatology* 83(5), 556–564

Poh, M.Z., Loddenkemper, T., Reinsberger, C., Swenson, N.C., Goyal, S., Madsen, J.R., and Picard, R.W. (2012) ‘Autonomic changes with seizures correlate with postictal EEG suppression.’ *Neurology* 78(23),1868–1876

Poh, M.-Z., Swenson, N. C., and Picard, R. W. (2010) ‘A wearable sensor for unobtrusive, long-term assessment of electrodermal activity’. *IEEE Transactions on Biomedical Engineering* 57(5), 1243–1252

Porges, S. (2003) ‘Social engagement and attachment: a phylogenetic perspective’. *Annals of the New York Academy of Sciences* 1008(1), 31-47

Prietsch, V., Knoepke, U., and Obladen, M. (1994) ‘Continuous monitoring of heart rate variability in preterm infants’. *Early Human Development* 37 (2), 117 – 131

Quintana, D., Guastella, A.J., Outhred, T., Hickie, I.B., and Kemp, A.H. (2012) ‘Heart rate variability is associated with emotion recognition: Direct evidence for a relationship between the autonomic nervous system and social cognition’. *International Journal of Psychophysiology* 86(2),168-172

Raj, S.R., Brouillard, D., Simpson, C.S., Hopman, W.M., and Abdollah, H. (2000) ‘Dysautonomia among patients with fibromyalgia: A non-invasive assessment’. *Journal of Rheumatology* 27(11), 2660 - 2665

Reagan, I., and Kidd, D. (2013) ‘Using Hierarchical Task Analysis to Compare Four Vehicle Manufacturers’ Infotainment Systems.’ *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*’. Held 30-4 September 2013 at Hilton San Diego. San Diego, California: Ashgate Publishing

Reimer, B., Mehler, B., Dobres, J., and Coughlin, J. F. (2013) *The effects of a production level “voice-command” interface on driver behaviour: reported workload, physiology, visual attention, and driving performance*. Cambridge: MIT AgeLab

Reimer, B., Mehler, B., Coughlin, J.F., Roy, N., and Dusek, J.A. (2011)’The impact of a naturalistic hands-free cellular phone task on heart rate and simulated driving performance in two age groups’. *Transportation Research Part F: Traffic Psychology and Behaviour* 14(1),13-25

Reimer, B., Mehler, B., Wang, Y., and Coughlin, J.F.(2010) *The impact of systematic variation of cognitive demand on drivers visual attention across multiple age groups*. ‘Proceedings of the54th Annual Meeting of the Human Factors and Ergonomics society’. held 27-1 September 2010 at San Francisco Cambridge Massachusetts: MIT AgeLab & New England University Transport Centre

Reimer, B., Mehler, B., Coughlin, J.F., Godfrey, K.M., and Tan, C. (2009) *An on road assessment of the impact of cognitive workload on physiological arousal in young adult drivers.* ‘Proceedings of the 1st International Conference on Automotive User Interfaces and Interactive Vehicular Applications’. held 21-22 September 2009 at Essen Germany. New York: ACM Press

Reyes, M., and Lee, J. (2008) 'Effects of cognitive load presence and duration on driver eye movements and event detection performance'. *Transportation Research Part F: Traffic Psychology and Behaviour* 11(6), 391-402

Richards, D., and Stedmon, A. (2016) 'To delegate or not to delegate :a review of control frameworks for autonomous cars'. *Applied Ergonomics* 53(B) ,383-388

Richter, E.A., Jensen, P., Kiens, B., and Kristiansen, S. (1998) 'Sarcolemma glucose transport and GLUT -4 translocation during exercise are diminished by endurance training'. *American Journal of Physiology* 274, 89-95

Riener, A., Ferscha, A., and Aly, M. (2009) *Heart on the road: HRV analysis for monitoring a driver's affective state.* 'Proceedings of the 1st International Conference on Automotive User Interfaces and Interactive Vehicular Applications'. held 21-22 September 2009 at Essen Germany. New York: ACM

Riener, A., Ferscha, F., Bachmair, F., Hagmuller, P., Lemme, A., Muttenhaler, D., Puhringer, D., Rogner, H., Tappe, A., and Weger, F. (2013) *Standardization of the In-Car Gesture Interaction Space.'*Conference on Automotive User Interfaces and Interactive Vehicular Applications'. held 27-30 October 2013 at Eindhoven, Netherlands. New York: ACM Press

Rigas, G., Goletsis, Y., Bougia, P., and Fotiadis, D. I. (2011)'Towards Driver's State Recognition on Real Driving Conditions'. *International Journal of Vehicular Technology* ,1687-5702

Roscoe, A. H. (1992) ‘Assessing pilot workload: Why measure heart rate, HRV, and respiration?’. *Biological Psychology* 34(2-3), 259–288

Roscoe, A.H (1979) ‘Handling qualities, workload and heart rate’. *AGARD-AG-*246, 83-91

Rumelin, S., and Butz, A. (ed) (2013) *How to make large touch screens usable while driving.* ‘Conference on Automotive User Interfaces and Interactive Vehicular Applications’. held 28-30 October 2013 at Eindhoven, Netherlands. New York: AMC Press

Russell, J. A., and Fehr, B. (1994) ‘Fuzzy concepts in a fuzzy hierarchy: Varieties of anger’. *Journal of Personality and Social Psychology* 67(2), 186-205

Sano, A., Picard, R., and Stickgold, R. (2014) ‘Quantitative analysis of wrist electrodermal activity during sleep’*. International Journal of Psychophysiology* 94 (3), 382-389

Sato, K., Kang, W., Saga, K., and Sato, K. (1989) ‘Biology of sweat glands and their disorders. I. Normal sweat gland function’. *Journal of the American Academy of Dermatology* 20 (4), 537-563

Sauro, J. (2011) Measuring Usability with the System Usability Scale (SUS) [online] available from <http://www.measuringu.com/sus.php> [08 June 2016]

Schaefer, F., and Boucsein, W. (2000) ‘Comparison of electro dermal constant voltage and constant current recording techniques using phase angle between alternating voltage and current’. *Psychophysiology* 37(1), 85-91

Scherer, K. (2001) ‘Appraisal considered as a process of multilevel sequential checking’. In *Appraisal Processes in Emotion: Theory, Methods, Research*. ed.by Scherer, K., Schorr, A., and Johnstone, T. New York: Oxford University Press, 120-144

Schipke, J.D., Pelzer, M., and Arnold, G. (1999) ‘Effect of respiration rate on short-term heart rate variability’. *Journal of Clinical and Basic Cardiology* 2(1), 92–95

Schmidt, K.E., Liu, Y., and Sridharan, S. (2009) ‘Webpage aesthetics, performance and usability: design variables and their effects’. *Ergonomics* 52(6) ,631-643

Schneegass, S., Pfleging, B., Broy, N., Heinrich ,F., and Schmidt, A. (2013) A data set of real world driving to assess driver workload. ‘Conference on Automotive User Interfaces and Interactive Vehicular Applications’. held 28-30 October 2013 at Eindhoven, Netherlands. New York: AMC Press

Scrimali, T. (2012) *Neuroscience-Based Cognitive Therapy: New Methods for Assessment, Treatment and Self-Regulation*. Chichester: Wiley

Shields, S., MacDowell, K., Fairchild, S., and Campbell, M. (1987) ‘Is mediation of sweating cholinergic, adrenergic, or both? A comment on the literature’. *Psychophysiology* 24 (3), 312-319

Shirley, D., Greenwood, A., Bottcher, C., and Cao, Z. (2009) *Driver Distraction Test Rig for HMI Studies*. ‘Conference on Systems, Man and Cybernetics’. held 11-14 October 2009 at Hyatt Regency Riverwalk. San Antonio, Texas: IEEE

Sinnreich, R., Kark, J.D., Friedlander, Y., Sapoznikov, D., and Luria, M.H. (1998) ‘Five-minute recording of heart rate variability for population studies: repeatability and age-sex characteristics’. *Heart* 80(2), 156-162

Smith, T., Vardhan, H., and Malet, B. (2014) Are we there yet? Thoughts on in-car HMI ustwo [online] London*: ustwo. available from* <*http://cdn.ustwo.com/automotive/AreWeThereYet_V1.0.pdf*> [15 May 2016]

Sodhi, S., Poupyrev, I., Glisson M., and Israr, A. (2013) ‘AIREAL: Interactive Tactile Experiences in free air’. *ACM Transactions on Graphics* 32 (4), 25-26

Son, J., and Park, M. (2011) *Estimating Cognitive load complexity using performance and physiological data in a driving simulator.* ‘Conference on Automotive User Interfaces and Interactive Vehicular Applications’. held 30-2 November at Salzburg, Austria: University of Salzburg

Son, J., Reimer, B., Mehler, B., Pohlmeyer, A. E., Godfrey, K.M., Orszulak, J., Long, J., Kim, M.H., Lee, Y.T. and Coughlin, J.F. (2010) ‘Age and cross-cultural comparison of drivers’ cognitive workload and performance in simulated urban driving’. *International Journal of Automotive Technology* 11(4), 533-539

Sonderegger, A., and Sauer, J. (2010) ‘The influence of design aesthetics in usability testing: Effects on user performance and perceived usability’. *Applied Ergonomics* 41(3), 403-410

Sternberg, R.J. (2006) ‘The Nature of Creativity’. *Creativity Research Journal* 18(1), 89-98

Storm, H. (2001) ‘Development of emotional sweating in preterms measured by skin conductance changes’. *Early Human Development* 62 (2), 149-158

Stuiver, A., De Waard, D., Brookhuis, K.A., Dijksterhuis, C., Lewis-Evans, B., and Mulder, L.J.M. (2012) ‘Short – term cardiovascular responses to changing task demands’. *International Journal of Psychophysiology* 85(2), 153-160

Suunto (2011) *Suunto T6d User Guide* [online] available from <http://ns.suunto.com/Manuals/t6d/Userguides/Suunto_t6d_UG_EN.pdf> [15 May 2016]

Swette, R., May, K.R., Gable, T. M., and Walker, B.N. (2013) Comparing three novel multimodal touch interfaces for infotainment menus. ‘Conference on Automotive User Interfaces and Interactive Vehicular Applications’. held 28-30 October 2013 at Eindhoven, Netherlands. New York: AMC: Press

Sztajzel, J. (2004). ‘Heart rate variability: a non invasive electrocardiographic method.’ *Swiss Medical Weekly* 134 (0), 515-522

Tarvainen, M.P., Niskanen, J.P., Lipponen, J.A., Ranta-aho, P.O., and Karjalainen, P.A. (2014) ‘Kubios HRV – Heart Rate Variability Analysis Software’. *Computer Methods and Programs in Biomedicine* 113(1), 210-220

Task Force. (1996) ‘Heart rate variability: standards of measurement, physiological interpretation and clinical use. Task force of the European society of cardiology and the North American Society of pacing and electrophysiology’. *European Heart Journal* 17, 354-381

Tellegen, A., Watson, D., and Clark, L.A. (1999) ‘On the Dimensional and Hierarchical Structure of Affect’. *Psychological Science* 10(4), 297-303

Tesla Model 3’s (2016) *Welcome to the interior of your self-driving car* [online] available from <http://www.theverge.com/2016/4/1/11348960/tesla-model-3-interior-touchscreen-self-driving-theory-autopilot> [04 January2017]

Thaler, S., and Wenzel, T. (2012) ‘Capacitive Touch Technology Opens the Door to a New Generation of Automotive User Interfaces’. *Atmel 9(1), 1-5*

Thayer, J.F., Lane, R.D. (2000) ‘A model of neurovisceral integration in emotion regulation and dysregulation’. *Journal of Affective Disorders* 61(3),201-216

Thayer, R. (1989) *The biopsychology of mood and arousal*. New York: Oxford University Press

The British Psychological Society (2016) *Promoting excellence in psychology* [online] available from <http://www.bps.org.uk/> [15 May 2016]

The Society of Motor Manufacturers and Traders (2016) SMMT Driving the motor industry [online] available from <http://www.smmt.co.uk/> [08 June 2016]

Tokay, D. (ed.) (2004) *Translation as a Metaphor in Academic Writing*. ‘Conference on International Writing Centres’. held 3-5 April 2003 at Ankara University. Istanbul: Sabanci University Press

Tosa, N., and Nakatsu, R. (1996) *Life-like communication agent - emotion sensing character "MIC" and feeling session character "MUSE". ‘*Conference on Multimedia computing and systems’. held 17-23 June at Hiroshima, Japan: IEEE Computer Society Press

Tractinsky, N., Cokhavi, A., Kirschenbaum, M., and Sharfi, T. (2006) ‘Evaluating the consistency of immediate aesthetic perceptions of web pages’. *International Journal of Human-Computer Studies* 64 (11), 1071–1083

Tractinsky, N., Shoval-Katz, A., and Ikar, D. (2000) ‘What is beautiful is usable’. *Interacting with Computers* 13 (2), 127–145

Tractinsky, N. (1997*) Aesthetics and apparent usability: empirically assessing cultural and methodological issues*. ‘CHI 97 Conference Proceedings’. held 22-27 March 1977 at Atlanta, USA. New York: ACM ,115–122

Treisman, A. (1996) ‘The binding problem’. *Current opinion in Neurobiology* 6 (2), 171-178

Treisman, A. (1982) ‘Perceptual Grouping and Attention in Visual Search for Features and for Objects’. *Journal of Experimental Psychology: Human Perception and Performance* 8 (2), 194-214

Treisman, A., and Gelade, G. (1980) ‘A feature integration theory of attention’. *Cognitive Psychology* 12 (1), 97-136

Underwood, G., Champham, P., Bowden, K., and Crundall, D. (2002) ‘Visual search while driving: skill and awareness during inspection of the scene’. *Transportation Research Part F: Traffic Psychology and Behaviour* 5(2), 87-97

Ustwo Auto (2016) *The near future of In-car HMI* [online] available from <https://ustwo.com/blog/the-near-future-of-in-car-hmi> [15 May 2016]

Valenza, G., Citi, L., Lanata, A., Pasquale, S., and Barbieri, R. (2014) ‘Revealing Real-Time Emotional Responses: a Personalized Assessment based on Heartbeat Dynamics’ *Scientific Reports* 4 (4998),1-13

Venables, P., and Christie, M. (1980) ‘Electrodermal activity’. in *Techniques in Psychophysiology.* ed. by Martin, I., and Venables, P. New York: John Wiley, 2-67

Vyzas, E., and Picard, R. W. (1999) *Online and Offline Recognition of Emotion Expression from Physiological Data*. ‘Workshop on Emotion-Based Agent Architectures. ‘Conference on Autonomous Agents’. Held on May 1st 1999 at Seattle, WA. Cambridge, MA: MIT Media Laboratory

Wahab, A. (2012) *Heart Rate Variability and QT Dispersion in Acute Myocardial Infarction*. German: Lambert Academic Publishing

Walker, G.H., Stanton, N.A., and Salmon, P.M. (2015) *Human Factors in Automotive Engineering and Technology*. Farnham, UK: Ashgate Publishing Limited

Wallin, G. (1981) 'Sympathetic nerve activity underlying electro dermal and cardiovascular reactions in man'. *Psychophysiology* 18 (4), 470-476

Weippert, M., Kumar, M., Kreuzfeld, S., Arndt, D., Rieger, A., and Stoll, R.(2010) 'Comparison of three mobile devices for measuring R-R intervals and heart rate variability: Polar S810i, Suunto t6 and an ambulatory ECG system'. *European Journal of Applied Physiology* 109 (4), 779-786

Wickens, C.D. (2002) 'Multiple resources and performance prediction'. *Theoretical issues in ergonomics science* 3 (2), 159-177

Wickens, C., Mavor, A., Parasuraman, R., and McGee, J. (1998) *The future of air traffic control: Human operators and automation*. Washington D.C: National Academies Press

Wickens, C. D. (1976) 'The effects of divided attention on information processing in tracking'. *Journal of Experimental Psychology: Human Perception and Performance* 2 (1), 1–13

Widmaier, E., Raff, H., and Strang, K. (2006) *Vander's Human Physiology*. Boston: McGraw-Hill

Wiggins, B. (2013) *Bradley Wiggins: My Hour*. London: Vintage Digital

Williams, D.P., Cash, C., Rankin, C., Bernardi, A., Koenig, J., and Thayer, J. (2015) 'Resting heart rate variability predicts self-reporting difficulties in emotion regulation: a focus on different facets of emotion regulation'. *Frontiers in Psychology* 10(6) ,261

Wilson, G. F. (2002) ‘Psychophysiological test methods and procedures’. In *Handbook of human Factors testing and Evaluation*. ed.by Charlton, S.G., and O’Brien, T.G. Mahwah, NJ: Lawrence Erlbaum Associates ,127-156

Wolf, M. M., Varigos, G. A., Hunt, D., and Sloman, J. G. (1978) ‘Sinus arrhythmia in acute myocardial infarction’. *Medical journal of Australia* 2(2), 52–53

Wordle (2014) *Create your own* [online] available from <http://www.wordle.net/> [15 May 2016]

Van Dooren, M., De Vries, J.J., and Janssen, J.H. (2012) ‘Emotional sweating across the body: Comparing 16 different skin conductance measurement locations’. *Physiology & behavior* 106 (2), 298-304

Van Roon, A. M., Mulder, L. J. M., Althaus, M., and Mulder, G. (2004) ‘Introducing a baroreflex model for studying cardiovascular effects of mental workload’. *Psychophysiology* 41(6), 961–981

Van Schaik, P., and Ling, J. (2009) ‘The role of context in perceptions of the asthetics of web pages over time’. *‘International Journal of Human-Computer Studies* 67(1), 79-89

Yang, C., Ji, L., Chen, W., and Fu, X. (2014) *Positive Affective Learning Improves Memory.* ‘*Engineering Psychology and Cognitive Ergonomics* 11th International Conference, Part of HCI International 2014’. Held 22-27 June 2014 at Heraklion, Crete, Greece. Switzerland: Springer International Publishing, 293-300

Yasuma, F. (2004) ‘Respiratory sinus arrhythmia: why does the heartbeat synchronize with respiratory rhythm?’. *Chest* 125(2), 683-690

Yerkes, R., and Dodson, J. (1908) 'The relation of strength of stimulus to rapidity of habit-formation'. *Journal of comparative neurology and psychology* 18 (5), 459-482

Zhao, S., Brumby, D.P., Chignell, M.H., and Salvucci, S.G. (2013) 'Shared Input Multimodal Mobile Interfaces: Interaction Modality Effects on Menu Selection in Single-Task and Dual-Task Environments'. *Interacting with Computers* 25(5), 386-403

7 Appendices

Appendix 1: Automotive User Interface 2013 Conference Paper

Appendix 1

Exploring Heart Rate Variability for Automotive Human Interface Evaluation

Jane Furness
Coventry University
furness4@uni.coventry.ac.uk

ABSTRACT

In-vehicle information systems have seen an era of continuous development within the industry and they are acknowledged as an important differentiator when purchasing a car. The challenge for designers and engineers is to produce systems, which are helpful, novel, exciting, safe and user friendly. In this extremely competitive market the analysis of customer response is vital and should include customers' affective responses. This paper outlines current research exploring the relationship between heart rate variability and affective responses with the aim of providing a psychophysiological model of user responses that could be applied throughout the design cycle of human machine interface (HMI) systems.

Keywords

Human machine interface (HMI), Heart rate variability (HRV), Affective responses.

1. INTRODUCTION

In the last decade there has been a surge of new technologies in vehicle information systems (IVS) and advanced driver assistance systems (ADAS), which have been found to be desirable by customers and thus influential in their decision making when choosing a specific car. [1] These systems now deploy a wide range of applications such as climate control, satellite navigation, adaptive cruise control, intelligent speed adaption, telecommunications and entertainment systems. The challenge for the industry is to design systems, which while being novel and exciting remain user friendly and supportive of both the driving task and automotive context. This challenge is being met by a collaborative partnership across different attributes within automotive companies. Of primary importance in these designs is the risk factor to the user when interacting with in-vehicle interfaces. [4] It is therefore essential that the design and implementation of information systems ensures that safety is not compromised by aesthetic and complex novel design features or merely driven by the technologists without appreciating the user needs. [11]

In the present economic climate each automotive manufacturer must listen attentively to any feedback from current and potential customers to gain advantages over competitors. An area that has not been fully explored is customers' affective responses. These responses must be examined as they could provide valuable user affective feedback, that could then be applied to the design cycle with the objective of improving the user experience with the HMI systems. [13] A difficulty arises however in attaining quantifiable measures to interpret these responses. This has led to a need to find quantifiable measures to asses these subjective responses.

AutomotiveUI'13, October 28--30 2013, Eindhoven, Netherlands.
Adjunct Proceedings

2. AFFECTIVE RESPONSES

Research in human emotion has notable difficulties mainly due to a lack of consensus between scientific theorists regarding what emotions are, who can experience them and methods to elucidate them. There is agreement, however, that emotions are a group of psychological states that include subjective experience, expressive behaviour and physiological responses. Osgood's (1952) seminal work in semantic differential where the variance in emotional assessments are attributed to three major dimensions (valence, arousal and dominance) led Bradley & Lang (1997;2008) to develop a range of standardised emotionally evocative pictures, sounds and words that researchers can apply to test dimensions of emotions in a diverse range in studies. [12,3,10]

Affective responses are frequently evaluated using a range of self-rating scales but due to the subjective nature of this type of evaluation it has often been difficult to extract meaningful data. To obtain objective methods to evaluate emotions, the autonomic nervous system (ANS) and its relationship with these emotions has been researched over a number of years. Current research suggests that there is a considerable ANS response to emotion. [8]

Measurement of ANS activity however can be problematic, as methods may be invasive, require large immobile equipment and individual intrinsic physiological factors may obscure the data. Physiological measures are frequently applied when assessing autonomic nervous activity, with cardiovascular indices being a widespread choice. With the advancement in mobile devices, cardiovascular metrics offer *real time*, non-invasive and reliable measurements. With their ease of recording, their use has been accepted in a wide range of research and clinical fields.

3. AUTONOMIC NERVOUS SYSTEM (ANS)

The ANS is usually independent of conscious control and innervates smooth muscle, cardiac muscle and glandular tissue. The ANS plays an influencing factor in the cardiovascular, respiratory, digestive, urinary and reproductive systems. The two the parasympathetic nervous system (PNS). The majority of interplay between these two branches normally has opposing effects, where one causes excitement and the other causes inhibition. [5] The SNS increases the body's general arousal state and primes it in readiness for an emergency situation that may arise; this is generally referred to as the 'fight or flight' response. Stimulation of the SNS causes an increase in respiration, heart rate, blood pressure, metabolic rate and activation of sweat glands with a heightened mental alertness. The PNS is referred to as the restorative system as the general nature of this system is to conserve energy and encourage sedentary demands such as digestion and absorption of food. [5]

3.1. Cardiovascular Measures

Cardiovascular measurements have been widely used as an ANS measure in clinical, laboratory and field studies and have a wide range of applications. Heart rate has proved to be a reliable unobtrusive measure that is easy to collect and by far the most frequently applied cardiovascular index used in aviation and research studies. [16] Although heart rate continues to be used in a wide range of studies and clinical applications, the emergence of heart rate variability (HRV) to evaluate the balance between the sympathetic and parasympathetic divisions of the ANS has accelerated research in this area. Advancements in computer technology and signal processing algorithms have greatly increased the understanding of the electrical heart rhythms that are recorded on an electrocardiogram (ECG). [6]

3.1.1. Heart Rate Variability

Heart rate variability (HRV) describes the beat-to-beat variability of the heart rate. HRV measures are determined by the R-R waves on an electrocardiogram, which are recognised to reflect the fluctuations in the atrioventricular conduction activity (see Figure 1).

Figure 1: R-R Interval time series

The heart rate signal can be decomposed by power spectral analysis into specific frequency bands, to which both the SNS and PNS contribute. In short-term recordings the spectral power is divided into three frequency bands: high frequency (HF) 0.15-0.40Hz, low frequency (LF) 0.04-0.15 Hz and very low frequency (VLF) 0-0.04. The high frequency correlates with fluctuations that are due to the efferent parasympathetic nervous system activity. Fluctuations in the low frequency band are more complex to determine and although they are linked to efferent sympathetic nerve activity, there are also efferent parasympathetic nerve activity influences on fluctuations seen in this frequency band. The origins of the fluctuations in the very low frequency band are yet undetermined. [9]

The prospect of being able to quantitatively assess the ANS facilitated new research and publications on HRV, spanning over a diverse group of research areas. Although progress and development has been achieved in methodology techniques and analytical processes, there is as yet no definitive standardisation for measurements and processing, such is the complexity of this field of study. In 1996 the European Society of Cardiology and the North American Society of Pacing and Electrophysiology addressed issues relating to measurement methodologies to provide direction and guidelines to researchers from all fields to bring some homogeneity to publications. [15] Both time and frequency domain analyses are frequently employed in HRV studies. A frequency domain measurement of interest is the LF/HF ratio, which has been attributed to reflect the sympatho-vagal balance and quantifies the relationship between the sympathetic and parasympathetic nerve activities. [7] As HRV is influenced by sympathetic and parasympathetic nerve activity, it makes it a promising measure to explore, as it may have the potential to elucidate physiological responses that participants may experience when exposed to an emotional stimulus.

To obtain a RR interval time series for HRV analysis an ECG can be assessed or a device that produces RR data used. Although these non-invasive devices are relatively straight forward to use, biosignals can be difficult to analyse due to internal or external influences that produce instability in the signal. Movement and poor contact with the sensors can cause movement and noise artifacts. Other issues are missed or ectopic beats and signal digitization (time and amplitude) which can disturb the signal and produce inaccurate measurements. [9]

Questions arise in the preparation of this research proposal in particular reference to the use of HRV as a quantifiable measure of affective user response to automotive human interfaces. The two questions stated below have been the basis for the feasibility study that is described in section 4.

Q1: Can HRV detect emotional responses ?
Q2: Can HRV differentiate between different emotional dimensions ?

1. STUDY AIMS AND DESIGN

Before commencement of the testing of automotive human interface systems, it was imperative that the methodology and equipment be rigorously tested and therefore initial studies are focussed on these pre-conditions. The initial study also needed to assess the sensitivity of HRV to emotional responses.

Coventry University Ethics Committee granted approval for a feasibility study. Resting base line measurements were taken for all physiological parameters prior to any psychological stimuli. Participants were assessed on their emotional responses using a validated list of emotive words. The words had been categorised into high, neutral and low groupings based on their mean rated dimensions of valence, arousal and dominance. [3] Each participant viewed 20 words that were randomly delivered to counterbalance any order of effects. Each word was displayed for a set period of time and participants were instructed to view the word for the entire display time. Immediately after each word had been viewed the participants were asked to complete an emotional subjective rating questionnaire. This was followed by participants viewing a normalising object (a dot presented on the screen) before the next word was viewed. Concurrent physiological assessments were measured to assess the modulations in the LF/HF ratio when the participants were presented with different emotional word strings.

4.1. Equipment and Materials

A *Suunto t6c* heart rate chest monitor and wrist watch was fitted to the participants and an *Omron RX3* blood pressure monitor was placed on each participant's wrist. Both devices were fitted and used in accordance with the manufacturer's instructions.

Sixty words selected from the Affective Norms for English Words (ANEW) and a subjective emotional rating questionnaire was used (see Table 1). [3] The Self Assessment Manikin (SAM) was used in this study. [2] SAM is a set of icons that enable participants to associate an affective response with a visual expression of emotion (as illustrated in Figure 2).

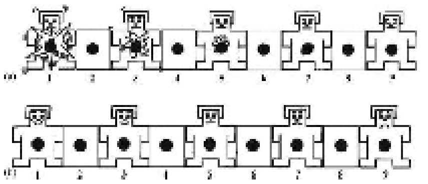


Figure 2. SAM icons for arousal and valence dimension

Table 1. ANEW selected words prior to random selection

HIGH Group	NEUTRAL Group	LOW Group
win	tool	loneliness
orgasm	black	failure
cash	lesbian	unhappy
thrill	razor	sad
kiss	news	sick
fun	market	gloom
graduate	ship	paralysis
victory	avenue	depressed
sexy	ketchup	depression
ecstasy	name	lonely
leader	hat	coward
desire	nursery	inferior
promotion	doll	fatigued
admired	hit	poverty
excitement	office	dreary
fame	rough	loser
engaged	stove	infection
erotic	concentrate	obesity
surprised	cannon	illness
adventure	medicine	deformed

4.1.1. Methodology

Twenty-one Coventry University students were recruited for the study (11 males and 10 females). The mean age of the participants was 21 years (SD 0.89). Participants were asked to refrain from caffeine drinks and nicotine four hours prior to the study. Before commencement of the trial, the participants were given instructions regarding the physiological equipment and procedure. Informed consent sheets were signed by each participant. Resting base line measurements with the duration of five minutes were taken prior to the ANEW words being presented.

Participants were presented with the selected word strings on a 40 x 31cm screen in a PowerPoint format. Each word appeared for six seconds during which time the participant had been instructed to concentrate on the word during its presentation. The word was then removed from the screen and the participant completed a pen and paper SAM to rate their subjective reactions to the words over the three dimensions of valence, arousal and dominance. [3] Participants had fifteen seconds to complete the SAM rating scale before returning their attention to the screen where a normalising object (dot) was displayed for 10 seconds. This sequence was repeated until all 20 words had been displayed.

1. DATA ANALYSIS AND RESULTS

The HRV data was first processed through *Suunto Training Manager* (2.3.0) then exported to *Kubios HRV (2.1)*, which is an advanced tool for studying heart rate variability. [14] *Kubios* calculates a wide range of time and frequency variables using the R-R interval data. Table 2 illustrates measures calculated from the baseline recordings.

Table 2. Mean HR, BP, RR & LF/HF ratio baseline measures (±SD)

Group	HR (bpm)	Systolic BP (mmHg)	Diastolic BP (mmHg)	Mean RR (ms)	LF/ HF ratio
Students	80.40 (9.3)	125 (11.5)	78 (10.8)	768.22 (85.0)	1.13 (0.97)

To assess the sympatho-vagal balance, which is reflected by the fractional distribution of power across the frequencies the LF/HF ratio was calculated during the ANEW presentation. [7] For this study fast Fourier transform analysis was used to determine autonomic modulation. Calculated data was inputted into excel and both excel and statistical analysis was undertaken via SPSS.

The LF/HF ratio data assessed during the ANEW stimuli was analysed for the Test of Normality using the Shapiro-Wilk's test. Only three participant's data was normally distributed which was processed through SPSS using an ANOVA model to assess the difference in the LF/HF ratio against the three word groups (high, neutral and low), rated across the emotional dimensions of valence, arousal and dominance. There was no significant difference found in the LF/HF ratio between the ANEW high, neutral and low word groupings. The non-normally distributed data was processed using Kruskal-Wallis's non-parametric test and indicated that there was a significant difference in the LF/HF ratio between the word groups (high, neutral and low). To establish where these differences occurred a post hoc analysis using a Pairwise comparison was carried out.

The LF/HF measures were significantly different between the ANEW high words and neutral words ($p<0.001$). LF/HF measures between ANEW Neutral and Low words indicated $p<0.01$. There

Appendix 1

was no significant difference between the ANEW High and Low words ($p>0.05$).

1. FUTURE WORK

The first stage in the current study was exploratory research into the feasibility of HRV as a suitable method to assess affective states. The initial results indicated that emotional responses were detected by the biosignals of the heart rhythm. The feasibility study also assessed and validated the equipment to determine whether data could be extracted from the recordings with meaningful interpretation of affective responses. The running of the initial study enabled the methodology to be tested and re-designed to ensure rigorous future testing.

Ethical approval has now been granted for a further study which has already commenced. The focus will be solely on HRV measures in isolation. 40 participants are to be recruited for this study. Participants will view 21 words from the ANEW system, which will be followed by a response and recognition task, where 60 words will be presented in a random format and the participants have to select an appropriate key if they recognize the words from the previous task. Participants will then view 21 images from the International Affective Picture System (IAPS) before completing the response and recognition task using 60 images. [10] The addition of using psychological imagery stimuli, will allow for comparisons to be made against the ANEW stimuli. This will be more in line with the aim of the research being used to assess visual presentation on automotive interfaces. In addition to a direct comparison the introduction of imagery stimuli, will be the second step of an incremental pathway construct, in which future trials will incorporate automotive interface visual images that can be assessed using HRV metrics and subjective self assessment scales.

By understanding the cognitive construct in the processing of visual displays there is an opportunity to map both psychological and physiological measures and to apply a theoretical structure to automotive interface design which would offer quantifiable objective results parallel with self subjective evaluation.

The mapping of both objective and subjective metrics will provide a powerful tool to automotive manufacturers in allowing them to access a scientific and quantifiable process that facilitates a better understanding of their customers' responses. This level of information would be essential in giving them important guidance within the early stages of design evaluation in terms of usability and user satisfaction.

2. QUESTIONS / AND ISSUES

- The nature of requiring supervisors from different domains and trying to keep the correct balance of physiology and cognitive psychology within the study.
- How to manage equipment problems and ways to ensure that participants are confident in fitting the equipment correctly.
- Due to the different methods of analysing HRV, it can be difficult to know which method is most appropriate for study.
- Can HRV be mapped reliably against self assessment rating scales ?

3. ACKNOWLEDGEMENTS

The author would like to thank her Coventry University Academic Supervisors; Dr. Helen Maddock, Dr. Dale Richards, Dr Graham Shelton-Rayner and also Sebastian Paszkowicz from Jaguar Land Rover for their support with this study.

4. REFERENCES

[1] Amditis, A., Andreone, L., Pagle, K., Markkula, G.,Deregibus, E., Rue, M., Bellotti, F., Engelsberg, A.,Brouwer, R., Peters, B., and De Gloria, A . 2010 Towards the automotive HMI of the future: Overview of the ADE-Integrated Project Results. *IEEE Transactions on Intelligent Transportation Systems*. 11, 3 (Sept. (2010),567-578.

[2] Bradley, M.M., and Lang, P.J.1994. Measuring emotion: The self–assessment manikin and the semantic differential. J.Behav.Ther.and Exp.Psychiat.25,49-59.

[3] Bradley, M.M., and Lang, P.J. 1999. *Affective Norms for English Words (ANEW): Instruction manual andaffective ratings. Technical Report* The Centre for Research in psychophysiology, University of Florida.

[4] Carsten, O.M.J., and Nilsson, L. 2001. Safety Assessment of Driver Assistance Systems, *European Journal of Transport and Infrastructure Research*. 1, 225-243.

[5] Fox, S.I. 2011. *Human Physiology*. McGraw-Hill International, New York.

[6] Kamath, V., Watanabe, M.A., and Upton, A.R.M. 2013. *Heart Rate Variability (HRV) Signal Analysis*. Taylor and Francis Group, London.

[7] Keenan, D. and Grossman, P. 2006. Adaptive Filtering of Heart Rate Signals for an *International Journal of Information and Communication Engineering*. 2 .

[8] Kreibig, S.D. 2010 Autonomic nervous system activity in emotion: A review. *Biological Psychology* 84, 394-421.

[9] Kuusela, T. 2013. Methodological Aspects of Heart Rate Variability Analysis. in Kamath,M.V., Watanabe, M.A., and Upton, A.R.M ed. *Heart Rate Variability (HRV) Signal Analysis*. Taylor and Francis Group, London, 9-40.

[10] Lang, P.J., Bradley, M.M., and Cuthbert, B.N. 2008. *International affective picture system (IAPS): Affective ratings of pictures and instruction manual. Technical Report A-8*.University of Florida, Gainesville, FL.

[11] Lavie, T., Oron-Gilad, T. and Meyer, J. 2010. Aesthetics and usability of in-vehicle navigation displays. *International Journal of Human-Computer Studies*, 69, 80-99.

[12] Osgood, C.1952. The nature and measurement of meaning. *Psychological Bulletin*. 49, 172-237.

[13] Picard, R.W., Vyzas, E., and Healey, J. 2001. Toward machine emotional intelligence: Analysis of affective physiological state. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 23,10,1175-1191.

[14] Tarvainen, M.P., Niskanen, J.P., Lipponen, J.A., Ranta-aho, P.O., and Karjalainen, P.A. 2013. Kubios HRV - HeartRate Variability Analysis Software, *Computer Methods and Programs in Biomedicine*. http://dx.doi.org/10.1016/j.cmpb.2013.07.024.

[15] Taskforce .1996. Heart rate variability: standards of measurement, physiological interpretation and clinical use.*Task force of the European society of cardiology and the North American Society of pacing and electrophysiology*. 93, 1043– 65.

Appendix 2: HCI International Conference Paper 2013

Autonomous control in military logistics vehicles: Trust and safety analysis

Nicole Gempton, Stefanos Skalistis, Jane Furness, Siraj Shaikh and Dobrila Petrovic

Faculty of Engineering and Computing, Coventry University, Coventry, CV1 5FB, UK

{n.gempton;stefanos.skalistis;furness4;s.shaikh;d.petrovic} @coventry.ac.uk

Abstract. Ground vehicles are increasingly designed to incorporate autonomous control for better performance, control and efficiency. Such control is particularly critical for military logistics vehicles where drivers are carrying sensitive loads through potentially threatening routes. It is imperative therefore to evaluate what role does autonomy play to help safety, and whether drivers trust autonomous control. In this paper we investigate the use of semi-autonomous vehicles used for military logistics and carry out human factors analysis to reflect on trust and safety issues that emerge from the driving of such vehicles.

Keywords: Military, Semi-autonomous vehicles, Logistics, Human Factors

Introduction

Human failure is often a cause of accidents. Increasing the level of automation while useful in many cases, does not necessarily reduce the number of human failure related accidents. For such automation to be successful the human user must be aware of the automation and react to it appropriately. In some cases it is not possible to fully automate the desired behaviour and the system has to rely on humans exhibiting the right behaviour. Examples of such systems include Unmanned Aerial

Vehicle (UAV) guidance [1], health care especially patient safety [2] and computer security [3].

Enhancing the driver experience of ground vehicles through increasing autonomy has been of interest for well over a decade now. Reduction in driver stress, freeing up limited attentional resources and improving road safety have been the major goals of this effort. However, autonomy brings with it a variety of other challenges that potentially risk road safety [4]. This could be due to sensor limitations, system design faults, error inducing design, or inadequate driver training; these certainly are some of the lessons learned from the introduction of autonomy in the aviation domain.

Adaptive cruise control (ACC) is an example of one mechanism introduced to provide safe distance control from a vehicle in front: once engaged, the vehicle operates in a typical cruise controlled fashion with the added feature of sensing the vehicle in front to adapt speed if it slows down or speeds up ensuring a minimal safe distance at

adfa, p. 1, 2011.

all times. Studies have demonstrated that such autonomy has the potential of causing delayed driver reaction [5], awkward handover and mode confusion, with up to a third of drivers having forgotten at some stage whether ACC was engaged or otherwise [6]. This has serious road safety risks and raises a question whether design of such mechanisms would ultimately be detrimental to the intended goal. In addition to the time on task effects, road conditions and terrain also significantly affect driver experience, and contribute to fatigue [7]; difficult terrains require more frequent driver interventions [8] in semi-autonomous vehicles.

In this paper we investigate the use of semi-autonomous vehicles used for military logistics and carry out human factors analysis to reflect on trust and safety issues that emerge from the driving of such vehicles. Section 2 describes some of the problems associated with this research. Section 3 describes our methodology in relevant detail. Section 4 presents the results of our experimental analysis. Section 5 presents a brief conclusion to the paper.

Motivation

There are potential economic, health and safety benefits of semi-autonomous vehicles in various industrial applications. Although the level of automation in mining is more advanced than many other domains, human oversight and control is still necessary given various factors such as legacy equipment, interoperability of hardware, and the ability to handle unforeseen circumstances. It is essential to use virtual engineering environments to model the vehicle and environment which can then be used to train drivers [9]. In addition to the known challenges, such as mode error where the driver cannot recall what state the system is in, there are particular challenges posed by semi-autonomous vehicles including

- handover between manual and automated control during a task [10, 11], which is critical as the driver needs to be able to judge when to reclaim control or otherwise,

- inadequate feedback from the vehicle to the driver [12], with the consequence that the system fails on drivers' expectations during a task and ultimately maximum benefit of the technology is not derived, and
- a fundamental change of task for the driver as their role changes from monitoring the situation to monitoring the situation and automation [6].

Most of the work done so far in this area has addressed such challenges in isolation and at a high abstract level [13-17], has studied vehicle sensor data [11], driver feedback [6] in a real or simulated environment, or performed physiological assessments [18,19]. The latter two strands of work entirely focus on driver perception and experience, borrowing from separate traditions of cognitive and physiological science. Our approach in this research is to conduct experiments involving master drivers (who are professionally trained to drive such vehicles) and analyse physiological and reaction time measures to assess how autonomy affects driver experience.

Experimental Setup

This section describes the methodology adopted. Section 3.1 describes the experimental design followed by Section 3.2 which gives the details of the virtual driving simulator implemented to carry out the experiments.

Experimental Design

The overall purpose is to assess the impact of autonomous control for drivers tasked with driving military logistics vehicles. Typical journeys are undertaken in convoys through hazardous and life threatening enemy territory. Such convoys could include a large number of vehicles, traveling over large distances at a slow speed, and may take up to 36 hours to complete a mission. The drivers are expected to keep an optimal distance between the vehicles. Autonomous control (in terms of cruise and lateral control) is expected to enhance convoy performance by maintaining an optimal speed, reducing fuel consumption, engine and brake wear, and reduce driver fatigue and cognitive load. This is particularly critical given drivers of driving through such journeys are likely to pose a difficult terrain, low visibility, high noise and roadside obstacles.

Experiments were designed as part of a virtual simulator where the drivers were asked to drive through a 3 hour journey and follow a vehicle in front as part of a convoy. Some experiments were designed to allow drivers to have manual control throughout the journey, whereas others were designed to (uniformly) incur periods of autonomous control (when control was explicitly taken over from the drivers) varying from 1 minute to up to 10 minutes. The journeys were designed to simulate ascending and descending routes, short periods of poor visibility and loud (bang) noises. Three drivers took part in a total of six experiments. The drivers had varying levels (7 to 24 years) of driving experience.

The total set of data collected from the experiments is given below

- Time (since start of experiment) in seconds
- Wheel input in terms analog turn of wheel
- Pedal input in terms of analog press of pedal
- Vehicle speed in metres per second
- Distance measured as the length of gap between two vehicles
- Time taken by the driver to attempt a stroop test
- Heart Rate (HR) in heart beats per minute (collected every 2 seconds)

The primary task performance measure is the reaction time of the drivers measured separately for cases of lateral track error and inter-vehicle gap in all scenario runs. This allows us to measure how quickly the driver is able to safely return the vehicle to the middle of the lane or within a safe distance of the vehicle in front. The distance travelled over the course of the experiment (3 hours in total for all experiments) is also evaluated.

Secondary task performance measures are influenced by the demands placed on the driver by the primary task of driving the vehicle (indicating the driver's spare cognitive capacity). The secondary task we used was a stroop test, as the test includes an implicit series of cognitive processes, including perception, attentional allocation, decision-making, and a motor response used to assess the drivers' cognitive load during the driving/monitoring task. The stroop test is one of the most widely used examples to study attention and cognitive control [20]. Our implementation of the stroop test used a body of text to pose a question (to judge colour-matching) displayed at a fixed location on the screen, occurring at regular intervals of 6 minutes asking the drivers to respond within three seconds (via paddles on the steering wheel).

To capture physiological responses, heart rate (HR) is a frequently used cardiovascular measure of mental workload in complex task environments [21]. Related to this is stress which is essentially a physiological response to the mental, emotional, or physical challenges that we encounter [22]. The drivers were asked to wear a HR monitor for 10 mins before commencing to get a baseline of their personal HR. A Garmin FR70 wireless HR monitor belt was used to capture readings at a 2s interval.

Driving Simulator

To develop a virtual driving simulator, the game engine Unity3D [23] was used to provide realism along with rapid development. The game was designed to simulate the cockpit of a heavy load vehicle following another similar vehicle at all times. As shown in Figure 1, a terrain containing a clear path was created. Both vehicles (being driven and followed) are similar in dimensions and capabilities. The vehicle interface provided the driver speed and temperature readings, along with warning signs to indicate proximity to the vehicle in front and autonomous control.

Fig. 1. Driver vehicle simulator screenshot

The vehicle shown in front was simulated to be driven autonomously and a similar controller was implemented for the vehicle being driven in front. The autonomous controller aimed to keep an average target speed of 40 Km/h using only 50% of the throttle in 0 degrees inclination/declination. While the percentage of throttle varied according to the degrees of inclination/declination, the average speed is maintained at 40Km/h at all times. In terms of lane alignment, it drove towards the centre of the lane predicting the vehicle's position 1 second ahead. The simulated terrain is a 4x4 (Km) circular plane containing hills and valleys. The driving lane runs throughout the terrain, and is 7m wide and approximately 50 Km long.

One challenge was to ensure that all data collected was optimized in terms of computational and memory usage. To address this, unnecessary rendering was avoided using thread-locking mechanisms. In order to achieve autonomous behaviour, in terms of lane alignment, KD-tree data structure was used keeping the computational effort as low as possible. The data structure provided the capability, given an arbitrary point, of finding the nearest stored point. This was used on the central points of the lane in conjunction with the vehicle's position. Throughout the experiments the drivers' input was monitored including the values generated by the wheel and pedals that the driver used to interface with the simulator. All data was collected at least every 200ms (except stroop tests, which were regularly scheduled).

To measure the deviation from the centre of the lane, a tolerance zone was defined to allow the driver to maneuver. From the centre (point) of the vehicle, deviation was defined as the centre of the lane and the projection of the centre of the vehicle on the 3D plane that defined the lane. The tolerance zone was defined as 30% of the total lane that is 15% left and right from the centre of the lane.

The physical properties of the vehicles are similar to a Tatra 8x8 vehicle and took into account dimensions, wheels (powered and steerable), engine torque curve, gear ratios and differential ratio.

Finally, a variety of development specific technologies were used to achieve a realistic simulation. The lane was developed using the Path tool [24] that allows the creation of an arbitrary path, in the form of single-columned triangular meshes, through the terrain. The autonomous system was developed using the UnitySteer library that provides an extensible framework to model low-level behaviours which can be combined to provide a sophisticated high-level behaviour. The low-visibility incident was implemented using the particle system and using textures from assets provided by Unity. All of the aforementioned tools are either provided by Unity3D or can be found on the Unity Store.

Analysis

We divide our attention to two different aspects of the results achieved from our experiments, including trust as an attribute of drivers' level of reliance and faith in autonomous control of the vehicle, and the impact of autonomous control on safety.

Of the total four experiments where drivers were asked to sit through a period of manual and autonomous driving, three out of four experiments showed that drivers had a slight increase in their average HR through periods of autonomous control. Figure 2 shows the average HR for the four experiments along with the baseline HR.

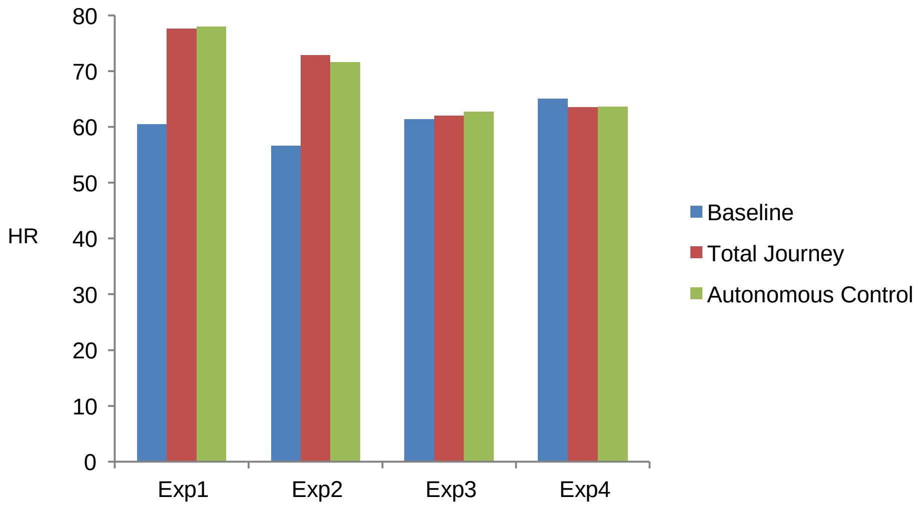


Fig. 2. The readings include baseline HR for drivers, along with their average HR measure for the total journey and autonomous control for the four mixed-mode experiments.

While there are increases in the first and third experiment, the increase in the fourth experiment is negligible. When demonstrated for individual drivers, the impact of autonomous control on individual HR remains inconclusive.

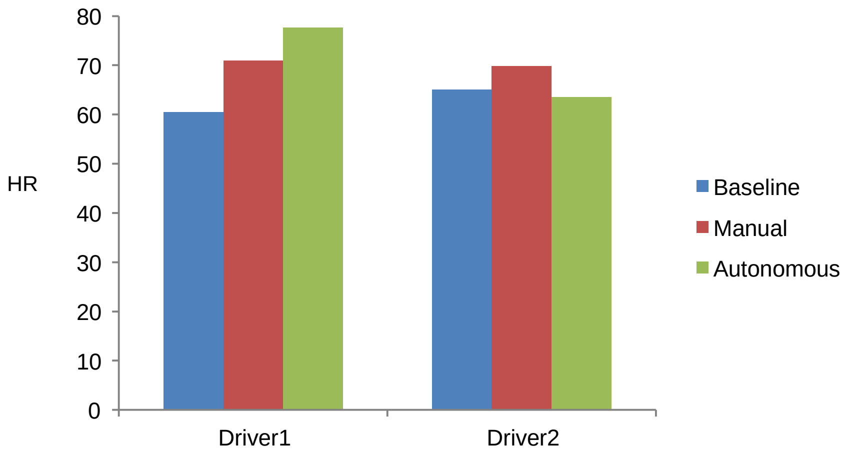


Fig. 3. Baseline HR for two drivers, along with their average HR measure for their manual (only) experiment and HR reading through autonomous for the mixed-mode experiments.

Figure 3 above shows the difference in HR readings for the two drivers who sat through a pair of manual and mixed-mode experiments.

One aspect of this relationship worthy of note is the drivers' experience over a period of time of driving with autonomy. For the four mixed-mode experiments, the average HR readings were observed longitudinally for the autonomous only periods and the trends are shown in Figure 4 below.

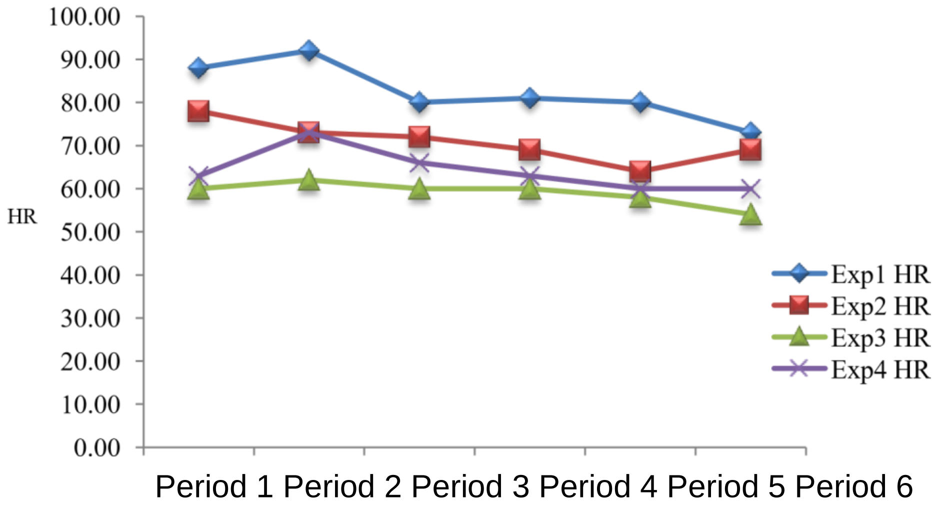


Fig. 4. There were six autonomous only periods during mixed-mode experiments. This graph shows the average HR readings longitudinally over the course of the experiment.

The four experiments show a notable decline in average HR readings as the experiment progress (except one experiment where in the last period the HR goes up). This reflects on drivers' increasing comfort in autonomous control as their experience of it progresses. Similar observation is found in the literature where drivers' trust in the vehicle autonomy is found to be improved over time [25-27].

An important part of safety is to ensure that any opportunity for driver error is avoided during the period of control transition. To assess how the drivers are affected in our study during periods of autonomous to manual handover, a handover period was identified to measure for HR. This included the period of 10 seconds before and after the actual handover (a total of 20 seconds) reflecting on how the HR readings change throughout this period. Figure 5 illustrates four experiments, of which three demonstrate an increase in average HR over the transition period. The maximum HR recorded in these three experiments is significantly higher than the average, demonstrating the high potential to cause errors during handover.

Moreover, the type of autonomy also influences the ease with which the operator could reclaim control. In a study of a shooting and planning style game, automation systems that removed the operator from task implementation, allowing the operator to plan ahead, proved the most disruptive when reclaiming control after automation failure [28]. The study shows that it was optimal to keep the operator in the loop with the system assisting in performing manual actions; removing the operator from manual control is detrimental to reclaiming control when needed [28].

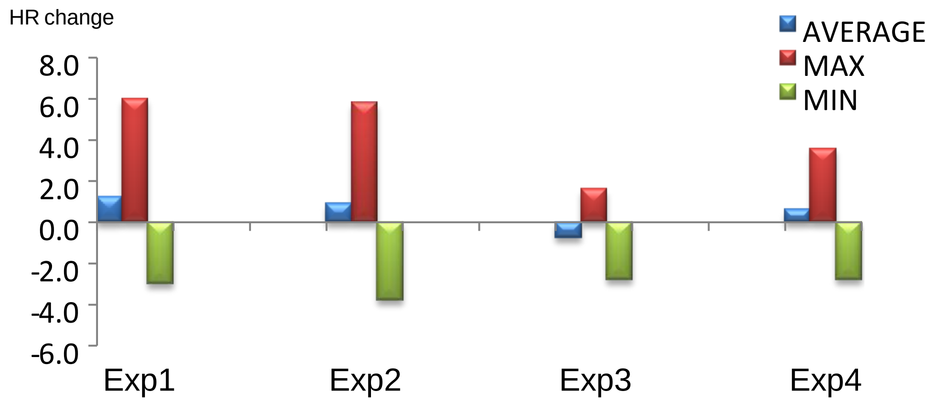


Fig. 5. For the four mixed-mode experiments average, high and low HR change is recorded throughout the handover period (10 seconds before and after the handover).

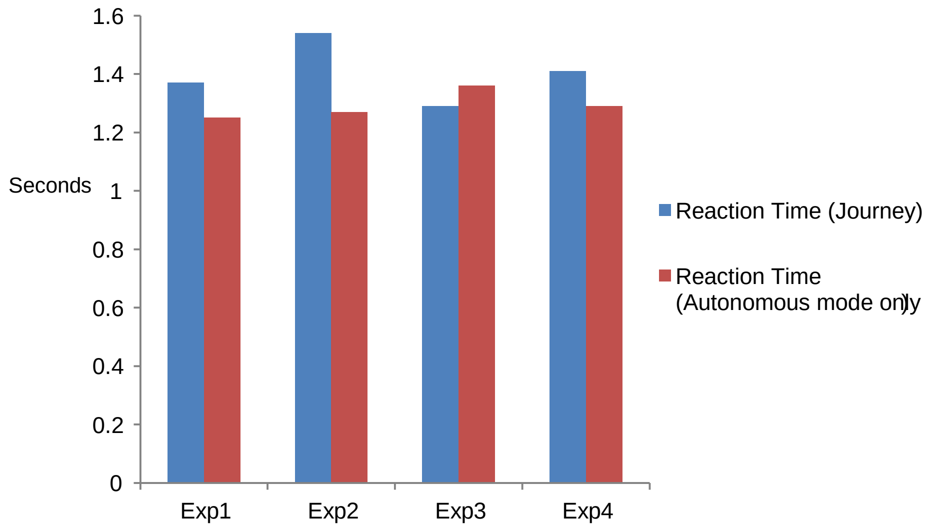


Fig. 6. In the four mixed-mode experiments, stroop tests were scheduled at regular intervals (during both manual and autonomous modes). The above readings show the reaction time for each experiment, both an average for the total journey and average reaction times for test conducted during autonomous mode only.

One factor that directly corresponds to safety is drivers' reaction time, which is critical if they are to respond to unexpected situations. Reaction time depends on the ability to process information within the driving environment, interpret that information to choose an action, and then react to the situation. Other studies show that when the driver role changes to a monitoring task, operator vigilance is reduced leading to loss of situational awareness and potential skill decay [29]. Our experiments however demonstrate a notable improvement in reaction time when drivers are driving through autonomous control mode. In Figure 6 above, three out of four drivers show a notable improvement over their average journey reaction times.

Conclusion

The use of actual master drivers from the military domain with field experience provides for valuable insight into how autonomy could play a significant role in establishing trust and improving on safety aspects of semi-autonomous vehicles. This work contributes to the field with empirical evidence obtained on the basis of carefully planned experiments. Further work is planned to use the empirical data to derive drive behavioural models for autonomous vehicles.

Acknowledgements

This work has been funded by the Ministry of Defence (UK) under the Unmanned Distribution Capability (UDC) initiative.

References

1. Chen, J.Y.C.: UAV-guided Navigation for Ground Robot Tele-operation in a Military Reconnaissance Environment. Ergonomics. 53(8), 940–950 (2010)
2. Carayon, P., ed.: Human Factors and Ergonomics in Health Care and Patient Safety. CRC Press (2011)
3. Cranor, L.: A framework for reasoning about the human in the loop. In: Proceedindgs of the 1st Conference on Usability, Psychology and Security, pp.1-15, USENIX (2008)
4. Stanton, N.A., Marsden, P.: From fly-by-wire to drive-by-wire: safety implications of automation in vehicles. Safety Science. 24, 35-49 (1996)
5. Vollrath, M., Schleicher, S., Gelau, C.: The influence of cruise control and adaptive cruise control on driving behaviour a driving simulator study. Accident Analysis & Prevention. 43(3), 1134-1139 (2011)
6. Larsson, A.F.: Driver usage and understanding of adaptive cruise control. Applied Ergonomics. 43(3), 501-506 (2012)
7. Oron-Gilad, T., Ronen, A.: Road characteristics and driver fatigue: A simulator study. Traffic Injury Prevention. 8(3), 281-289 (2007)
8. Scholtz, J., Antonishek, B., Young, J.: Operator interventions in autonomous off-road driving: effects of terrain. In: International Conference on Systems, Man and Cybernetics. Volume 3, pp. 2797-2802, IEEE (2004)
9. Moore, P., Ujvari, S., Pu, J., Lundgren, J.O., Xie, C.: Intelligent semi-autonomous vehicles in materials handling. Mechatronics. 9(7), 881-892 (1999)
10. Larsson, A.F.: Issues in Reclaiming Control from Advanced Driver Assistance Systems. In: Eu. Conf. on Human Centred Design for Intelligent Trans. Sys. Vol 2, pp. 557-564 (2010)
11. Viti, F., Hoogendoorn, S., Alkim, T., Bootsma, G.: Driving behavior interaction with acc: results from a field operational test in the netherlands. In: Intelligent Vehicles Symposium, pp. 745-750, IEEE (2008)
12. Norman, D., ed.: The Design of Future Things: Author of The Design of Everyday Things. Basic Books (2007)
13. Rushby, J.: Using model checking to help discover mode confusions and other automation surprises. Reliability Engineering & System Safety, 75(2), 167-177 (2002)
14. Curzon, P., Ruksenas, R., Blandford, A.: An approach to formal verification of human computer interaction. Formal Aspects of Computing. 19(4), 513–550 (2007)

15. Shaikh, S.A., Krishnan, P., Cerone, A.: A Formal Approach to Human Error Recovery. In Preproceedings of the 2nd Int. Workshop on Formal Methods for Interactive Systems (FMIS 2007), pp. 101-115, Technical Report RR-07-08, Queen Mary, Univ. of London
16. Cerone, A.: Closure and Attention Activation in Human Automatic Behaviour: A Framework for the Formal Analysis of Interactive Systems. In: Proceedings of the Fourth International Workshop on Formal Methods for Interactive Systems (FMIS), Electronic Communications of the EASST (ECEASST) Volume 45 (2011)
17. Shaikh, S.A., Krishnan, P: Framework for Analysing Driver Interactions with SemiAutonomous Vehicles. Electronic Proceedings in Theoretical Computer Science (EPTCS). Vol 105, 85–99.
18. Shelton-Rayner, G.K.: Quantifying exposure to psychological and physiological stress and automotive design. PhD thesis, Coventry University (2009)
19. Shelton-Rayner, G.K., Mian, R., Chandler, S., Robertson, D., Macdonald, D.W.: Quantitative physiological assessment of stress via altered immune functioning following interaction with differing automotive interface technologies. International Journal of HumanComputer Interaction. 27(9), 900-919 (2011)
20. Gwizdka, J.: Using Stroop task to assess cognitive load. In; Proc. of the 28th European Conf. on Cognitive Ergonomics, pp. 219-222 (2010)
21. Nikolova, R., Collins, S.: Effect of Workload and Stress on Operator Functional State. NATO Science Series Sub Series I Life and Behavioural Sciences. 355, 303-312 (2003)
22. Sun, F-T., Kuo, C., Cheng, H-T., Buthpitiya, S., Collins, P. and Griss, M.: Activity-Aware Mental Stress Detection Using Physiological Sensors. Mobile Computing, Applications, and Services, 282-301 (2012)
23. Unity – Game Engine. http://www.unity3d.com
24. Road/Path Tool. http://www.sixtimesnothing.com/road-path-tool/
25. Stanton, N.A., Young, M.S.: Driver behaviour with adaptive cruise control. Ergonomics 48(10), 1294-1313 (2005)
26. Kazi, T.A., Stanton,N.A., Walker, G.H., Young, M.S.: Designer driving: drivers' conceptual models and level of trust in adaptive cruise control. International Journal of Vehicle Design, 45(3), 339-360 (2007)
27. Dickie, D.A., Boyle, L.N.: Drivers' Understanding of Adaptive Cruise Control Limitations. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting. 53 (23), 1806-1810, SAGE Publications, (2009)
28. Endsley, M.R.: Level of automation effects on performance, situation awareness and workload in a dynamic control task. Ergonomics. 42(3), 462-492 (1999)
29. Endsley, M. R., Kiris, E.O.: The out-of-the-loop performance problem and level of control in automation. Human Factors: The Journal of the Human Factors and Ergonomics Society. 37(2), 381-394 (1995)

Appendix 3: Human Factors Conference Paper 2015

Appendix 3

Exploring physiological markers to assess user engagement with an automotive touch screen graphical interface (GUI)

Jane Furness, Coventry University

In-vehicle information systems have seen an era of continuous development within the automotive industry, providing a key differentiator for buyers when purchasing a car. This presents a challenge for designers and engineers in producing the next generation systems, which are helpful, novel, exciting, safe and user friendly. Therefore, the usability of any new visual display has an implicit cost in terms of the perceived aesthetic perception and associated user experience. Achieving the next engaging automotive infotainment system not only has to address the user requirements but also has to incorporate established safety standards whilst considering new interaction technologies.

For an automotive human machine interface (HMI) evaluation a triad of physiological, subjective and performance-based workload measurements are often employed to provide relevant and valuable data for product evaluation. However, there is also a growing interest and appreciation that determining real-time quantitative metrics to drivers' affective responses is worthy of exploration in providing valuable user affective feedback, that could then be applied to the design cycle with the objective of improving the user experience with the infotainment system.

The aim of this study is to employ both objective and subjective metrics to assess user engagement during interactions with an automotive infotainment HMI. By mapping both physiological and self-report scales it is hoped that a greater understanding of users' responses will be elucidated and such information may provide important guidance within the early stages of in-vehicle design evaluation in terms of usability and user satisfaction.

This is the third study in a research project using physiological measures to assess affective responses. The two previous studies focussed on determining whether heart rate variability (HRV) could detect emotional states when participants were exposed to emotional images and words. Heart rate variability provides an opportunity to study the autonomic nervous system which is an important factor when assessing affective states. This study will now use a range of physiological measures to detect emotional responses and validate self-report scales when participants engage with a touch screen HMI.

The study design will take a between subjects approach where each participant will interact with one of two interfaces, an experiment matrix will be followed to counter balance any order effects. The HMI was designed to simulate a range of touch screen

Appendix 4: Visual Aesthetics Questionnaire

Appendix 4

Date: ☐ Participant I/D: ☐

Visual Aesthetics of Graphical User Interface

The below questionnaire has 4 items with 7 response options.

Instructions: For each of the following statements, mark one box that best describes your reactions to the interface you have been using today.

Simplicity of Interface

1. Everything goes together on this interface.

Strongly Disagree Strongly Agree

☐ ☐ ☐ ☐ ☐ ☐ ☐

Diversity of Interface

2. The layout of the interface is pleasantly varied.

☐ ☐ ☐ ☐ ☐ ☐ ☐

Colourfulness of Interface

3. The colour composition is attractive.

☐ ☐ ☐ ☐ ☐ ☐ ☐

Craftsmanship of Interface

4. The layout appears professionally designed.

☐ ☐ ☐ ☐ ☐ ☐ ☐

Appendix 5: System Usability Scale

Appendix 5

Date: ☐ Participant I/D: ☐

System Usability Scale

The SUS is a 10 item questionnaire with 5 response options.

Instructions: For each of the following statements, mark one box that best describes your reactions to the interface you have been using today.

1. I think that I would like to use this interface frequently.

Strongly Disagree				Strongly Agree
☐	☐	☐	☐	☐

2. I found the interface unnecessarily complex.

☐ ☐ ☐ ☐ ☐

3. I thought this interface was easy to use.

☐ ☐ ☐ ☐ ☐

4. I think that I would need the support of a technical person to be able to use this interface.

☐ ☐ ☐ ☐ ☐

5. I found the various functions in this interface were well integrated.

☐ ☐ ☐ ☐ ☐

Appendix 6: Participant Information Sheet

Appendix 6

Participant information sheet

Study title: To assess user engagement when interacting with a touch screen graphical user interface.

Purpose of the project:

This is the third study in a research project which is examining physiological measures which will validate subjective self-report scales and assist in the evaluation process on automotive in-vehicle interfaces. The aim of this study is to assess user engagement when interacting with a touch screen graphical user interface. Physiological measures and self-report scales will be recorded and examined for correlations between the two measures. Results will also be examined to determine any differences in user engagement when interacting with different touch screens.

Why have I been approached ?

For the purpose of the study I need to recruit a number of healthy adults to participant.

Do I have to participate in this trial ?

No, participation is entirely voluntary.

What will happen to me if I take part in the study ?

You will be asked to attend the trial at a specific time and location within Coventry University HLS Faculty. Each trial will last approximately 60 minutes. You will be asked to refrain from caffeine drinks and nicotine for a period of four hours before commencement of the trial. An informed consent form will require signing and a short questionnaire to be completed before starting the trial.

Before commencement of the study, full instructions will be given to participants on the tasks. Demonstration and assistance in placing the physiological sensors will also be given where applicable. Baseline recordings will take place before any tasks commence.

The trial will assess your engagement with a touch screen automotive interface while you are carrying out a visual primary task (to simulate driving processes). During the primary task, a series of tasks, associated with an in-vehicle infotainment system will be performed. Whilst engaging with the interface an additional visual cognitive task will also be introduced to increase workload. Physiological measures will be recorded; these will consist of heart rate, heart rate variability, electrodermal activity and respiration. Self-report scales will also be completed at the end of the study. A video recording will run throughout the trial and will aid in the evaluation of user interaction with the interface.

What are the possible disadvantages and risks of taking part ?

The physiological monitoring equipment may be slightly uncomfortable to wear and in some instances the sensor contact gel may cause a little skin irritation.

What are the benefits of taking part ?

Apart from the opportunity to participate in a post-graduate research project you will experience new in-car display technology.

Appendix 7: Informed Consent Form

Appendix 7

Informed Consent Form

Participant reference code: ....

Study title: To assess user engagement when interacting with a touch screen graphical user interface.

Overview: The aim of this study is to assess participants' engagement when interacting with a touch screen graphical user interface. A primary visual attention tracking task and a cognitive task will be employed to increase workload during interface engagement. Physiological measures will be recorded and self report rating scales completed. A video camera will be used to capture the interactions with the touch screen. Data will be examined to determine if correlations exist between physiological and self-report scales. Differences in user engagement between different interfaces will be examined.

	Please initial
1. I confirm that I have read and understood the participant information sheet for the above study and have had the opportunity to ask questions.	
2. I understand that my participation is voluntary and that I am free to withdraw at any time without giving a reason.	
3. I understand that all the information I provide will be treated in confidence.	
4. I understand that I also have the right to change my mind about participating in the study for a short period after the study has concluded (two weeks from my last session).	
5. I agree to take part in the research project.	

Name of participant: ....

Signature of participant: ....

Date: ....

Name of witness: ....

Signature of witness:..

Name of Researcher:..

Appendix 8: IAPS Codes

Appendix 8

IAPS Image Numbers for Study Three

1280 / 1440 / 1463 / 1710 / 1811 / 2115 / 2191

2205 / 2208 / 2271 / 2590 / 2799 / 2800 / 3102

3230 / 5621 / 5623 / 5629 / 5825 / 5830 / 5833

6242 / 6800 / 6836 / 7000 / 7038 / 7041 / 7052 /

7081 / 7092 / 7137 / 7179 / 7300 / 7405 / 7502 /

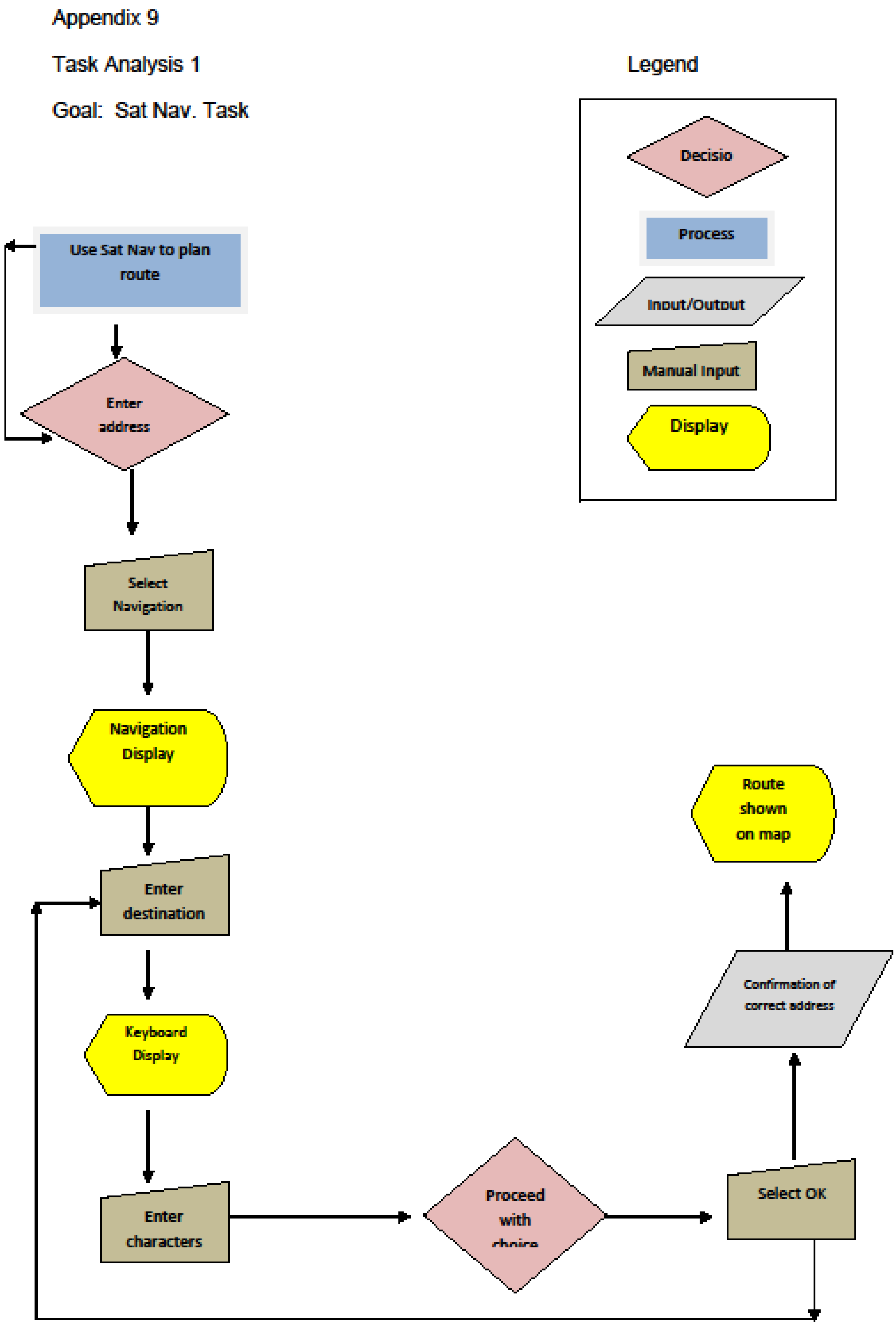
7546 / 7595 / 7920 / 7950 / 8010 / 8080 / 8170 /

8190 / 8370 / 8470 / 8492 / 8501 / 8531 / 9040 /

9075 / 9140 / 9220 / 9330 / 9331 / 9332 / 9417 /

9421 / 9560 / 9922 / 9927

Appendix 9: Task Analysis 1: satellite navigation

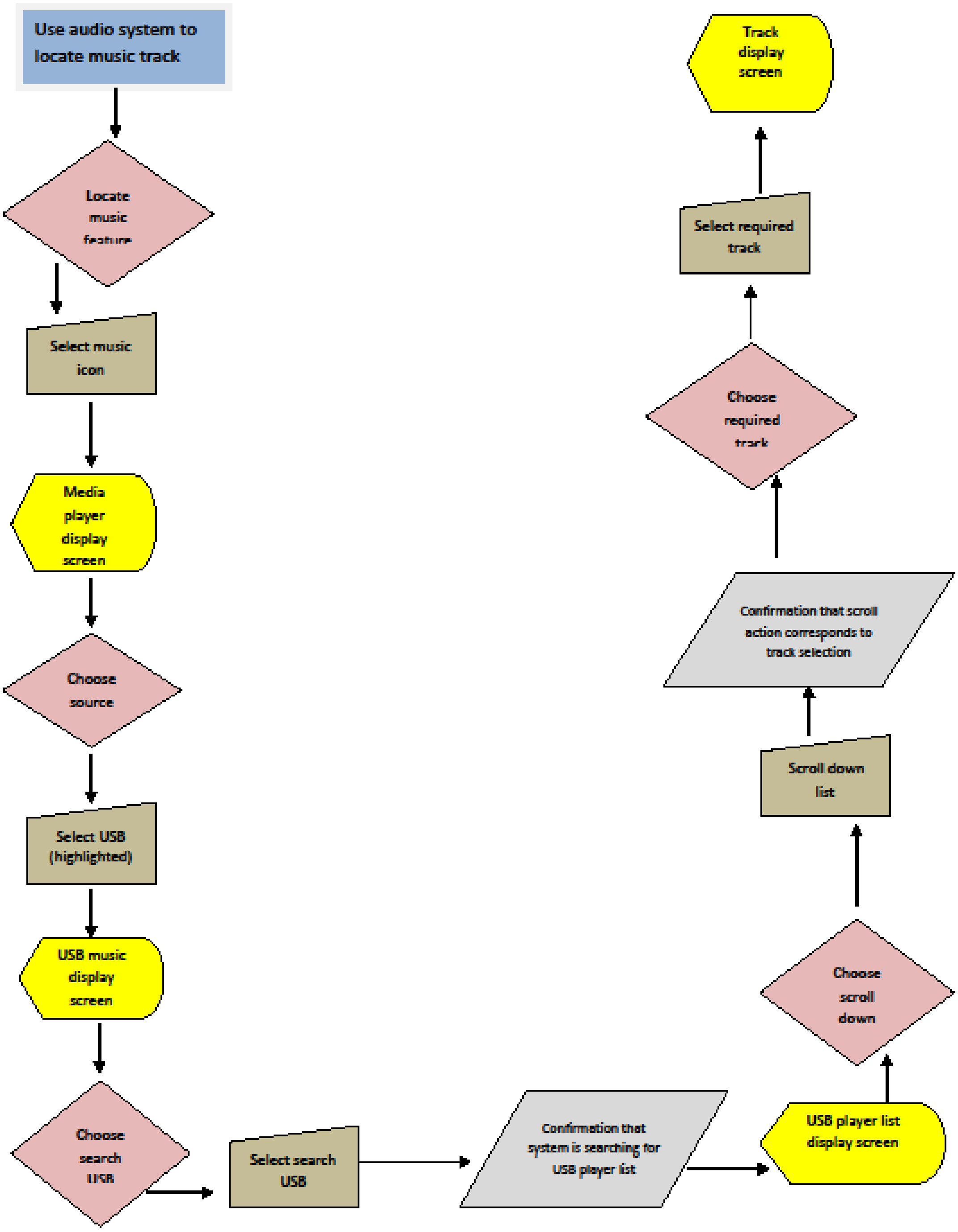


Appendix 10: Task Analysis 2: USB play list

Appendix 10

Task Analysis 2

Goal: Select a track from USB play list

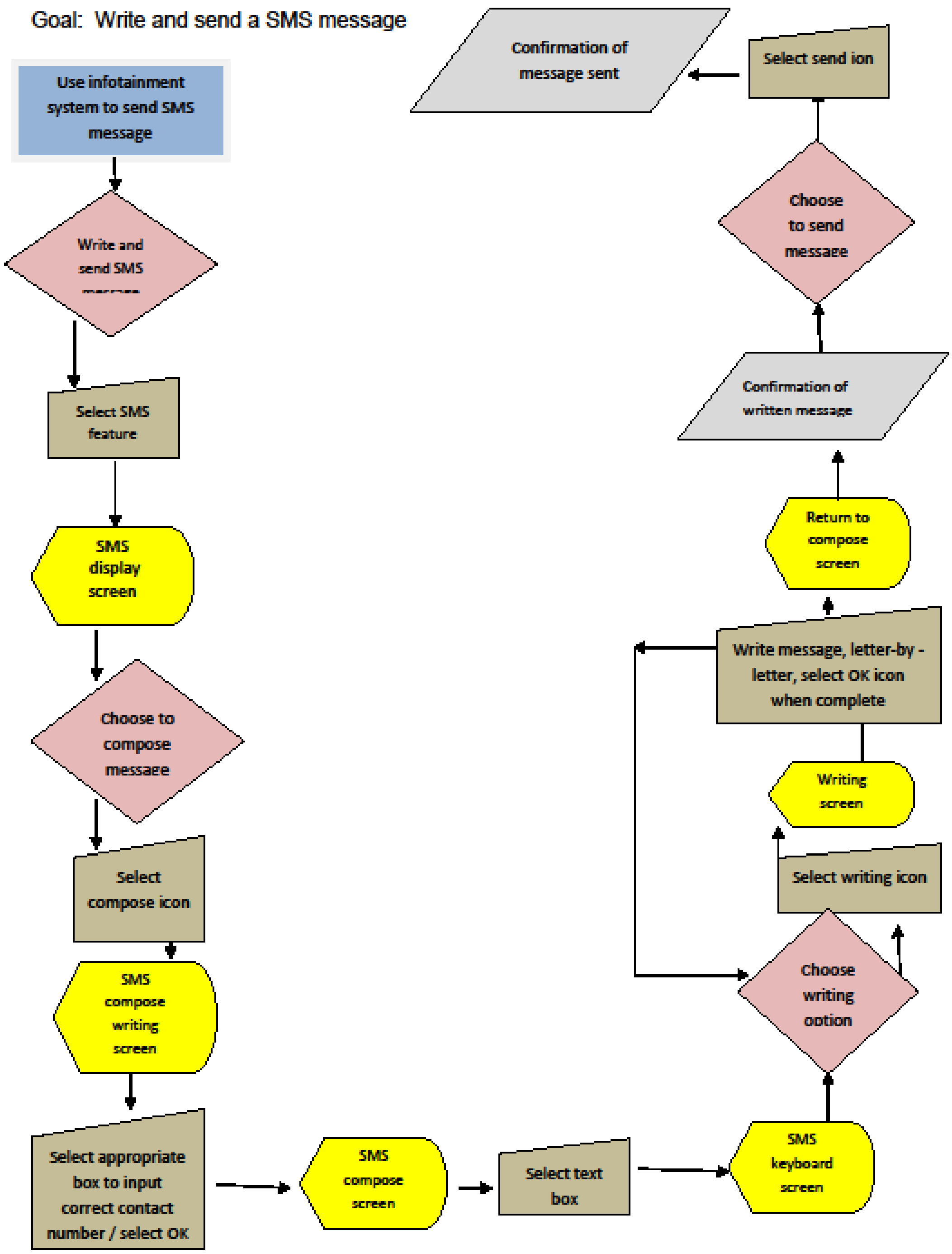


Appendix 11: Task Analysis 3: SMS message

Appendix 11

Task Analysis 3

Goal: Write and send a SMS message

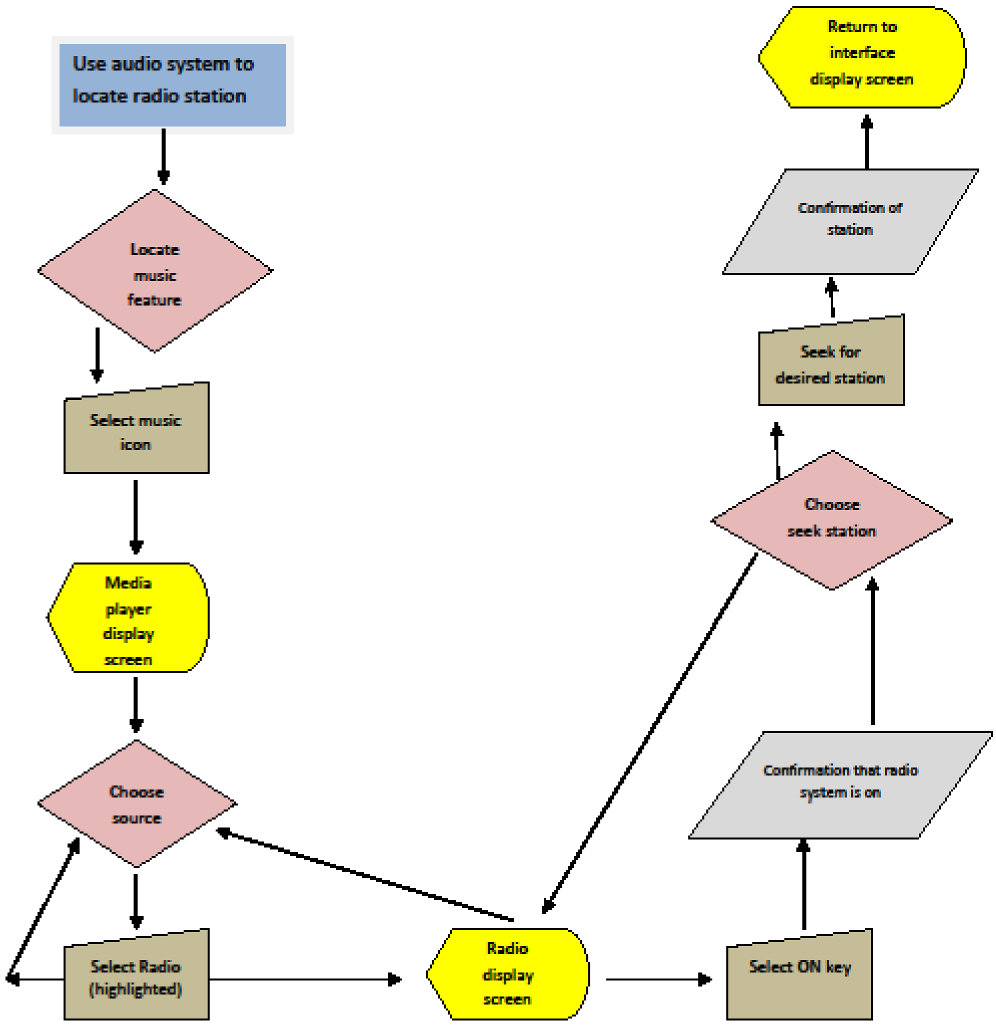


Appendix 12: Task Analysis 4: radio station

Appendix 12

Task Analysis 4

Goal: Select a radio station



Appendix 13: Task Analysis 5: telecommunication