SOUND QUALITY EVALUATIONS USING INTERACTIVE SIMULATION

SEBASTIANO D. GIUDICE

INNOVATION REPORT

THE UNIVERSITY OF WARWICK
SCHOOL OF ENGINEERING
WMG

JANUARY 2009

Abstract

Sound Quality engineering (SQE) is a discipline that should be embedded within Noise Vibration and Harshness (NVH) engineering. Its purpose is to tailor and enhance a vehicle's sound in order for it to meet and / or exceed the customers' expectations of the car and of the brand. NVH engineers need to use the opinions of customers, key decision makers in the organisation and those of their colleagues to set objective NVH targets for new vehicles. Their opinions are captured through jury evaluations.

Interactive simulation is a new approach to presenting sounds to a jury. It enables assessors to evaluate existing and concept cars generated in real time and before the manufacture of physical prototypes. This document summarises the research carried out for the EngD research programme. Its aim was to develop methodological approaches for setting up evaluations using interactive simulation, optimise its data capture and analysis capabilities and provide insight, to NVH engineers, into how assessors evaluate sound quality.

The first stage of the research involved benchmarking how the interactive simulation tool was being used by its developers, and how NVH targets are set within an OEM. This provided the foundation upon which to build the new methodology. The benchmark was compared against methodological approaches used in experimental psychology and in other sensory perception practices. This identified that principles for the design of evaluations had not been considered and appropriate statistical analysis techniques were not being implemented. Therefore it was not possible to ensure if the differences observed in the results were significant or not. It also became apparent that as each assessor was free to drive vehicles however they wished, the NVH engineers would not be able to link the assessor's subjective impressions with the acoustical stimuli used to form an opinion. This was due to the lack of observational methods that could be applied. In addition, the use of this form of interactivity was novel compared to the approaches available before the introduction of the simulation. Therefore it was not yet understood how it could influence the outcome of the evaluations.

An iterative approach was adopted to develop both tools and methods. Following the benchmarking stages, experimental design principles were implemented and a structured briefing method was formulated for the first time. These contributed to the overall methodologies and were also used to ensure the studies conducted as part of this research programme were free from experimental biases. This stage was followed by identifying a statistical analysis framework which can be used to study the assessors' subjective impressions. These contributions enable the NVH engineer to understand if the differences observed between sounds are significant or not.

The next phase was to understand how to best capture subjective opinions. Recommendations for this were based on the purpose of the evaluation. For the benchmarking stages of the NVH target setting process, many cars are often evaluated and therefore the duration of the evaluation can be a concern. However, the desired level of accuracy of the results is not as high as it is for the validations stages, upon which key engineering decisions are based, and also fewer cars are evaluated. Taken together it was possible to recommend the use of scaled interfaces for the benchmarking stages and the paired comparison method for the validation stages. The former takes less time to complete than the latter, which is however more accurate. The data capture capabilities were further enhanced through the development of a driver observation module; this enabled the assessor's assessment strategies to be recorded. The data collected provided insight into how assessors drove with the aid of new visualisation and analysis techniques developed.

Given the availability of these new methods and tools it was then possible to use the simulator to observe the influence of interactivity on the outcome of the evaluation. This demonstrated that assessment strategies can vary depending on the question asked to the assessor. The study indicated that assessors associated the refinement character of a car with driving conditions other than those presented to them in traditional evaluations.

Through the research and developments of this EngD programme, NVH engineers can now observe with confidence if differences between sounds are significant and they can see and hear how the assessor made decisions. Therefore they can now link subjective impressions formed with the stimuli evaluated by the assessor, hence an effective way of using the simulator has been proposed.

Acknowledgement.

I would like to thank a number of people for the help and support that they have given me throughout the EngD programme.

First of all I would like to thank WMG for providing me with the opportunity to be part of this research programme. I would like to thank in particular my mentor Professor Paul Jennings, for his guidance and encouragement, Dr Andy McGordon for taking the time to review my work and Mr Diego Sánchez-Repila, for his support in developing the tools presented in this document. I also would like to acknowledge the UK Engineering and Physical Sciences Research Council for supporting the research through the Warwick Innovative Manufacturing Research Centre and the EngD programme.

I would like to thank Mr Roger Williams, and Mr Mark Allman-ward from Sound Evaluations Ltd, for their continual involvement in the research. Without them I would have not been able to complete the doctorate. I would also like to thank Dr Garry Dunne from J&LR for providing much needed insight into the processes implemented within their NVH team.

I would like to thank my parents for encouraging me to take on new challenges, and for providing me the opportunities that were not available to them. Finally I would like to thank my wife Rachael for showing her complete support, consideration and recognition of the interruption the programme has brought to our lives. I will make it up to you.

Declaration

I confirm that the work contained in this document is my own unless otherwise stated

Sebastiano Giudice, January 2009

Contents Page

List of Figures

List of Tables

Abbreviations and Acronyms

- 2GWOT: 2nd Gear Wide Open Throttle
- DOM: Driver Observation Module
- EngD: Engineering Doctorate
- J&LR: Jaguar LandRover
- NVH: Noise Vibration Harshness
- SQE: Sound Quality Engineering
- SVT: Sound & Vibration Ltd.

1. INTRODUCTION TO THE INNOVATION REPORT

1.1. Introduction

The purpose of this document is to summarise the key findings of a research programme carried out as part of an Engineering Doctorate (EngD) at the University of Warwick. The research focused on the development of methodological approaches for sound quality evaluations using interactive simulation.

Within the automotive industry, Sound Quality Engineering (SQE) is the discipline of tailoring and enhancing a vehicle's sound in order to communicate and reinforce brand qualities. It requires the capture of subjective opinions, through evaluations, in order to set objective Noise Vibration and Harshness (NVH) targets for use within the vehicle development programme. Conventional subjective evaluation approaches include the assessment of sound in laboratory-style environments known as listening rooms, or by evaluating vehicles on-road.

Within a listening room, the sound of vehicles is represented by predetermined driving conditions. For the Jaguar brand, the sound of vehicles at 2nd Gear Wide Open Throttle (2GWOT) was assessed. The sounds are recordings of existing vehicles but the acoustic properties of concept vehicles can be simulated by applying filters. This approach is therefore suited for understanding which acoustic features influence perception without building physical prototypes. The evaluation is structured, and therefore the test conditions are repeatable for all assessors. A limitation however, is that assessors cannot choose the driving conditions to evaluate. Furthermore, from an engineering perspective, resolution of issues at this full load does not imply equivalent resolution at part load conditions.

During on-road evaluations, assessors can evaluate any driving conditions. However to evaluate concept vehicles, a physical prototype has to be designed and built. If at this stage, the acoustic character of the vehicle is not appropriate, then engineering changes have to take place, leading to delays in the vehicle development programme.

An interactive acoustic simulation tool was developed by Sound & Vibration Technology Ltd. (SVT) to bridge the gap between the two evaluation environments. Using this approach, assessors drive through a virtual environment, whilst assessing existing and concept vehicles, which are accurately generated in real-time in response to a driver's actions. The simulation tool can be experienced through two environments, a full vehicle simulator and a desktop simulator.

The full vehicle simulator comprises of a real vehicle fixed into positioned in front of a screen onto which the virtual environment (VE) is projected. The assessors hear the simulated sound through headphones and feel vibration through the seat, floor panel and steering wheel. They can control the perceived motion of the vehicle through the VE using the steering wheel and foot pedals within the car. For the desktop simulator, the assessor sits at a desk, in front of a PC monitor onto which the VE is displayed. They control the vehicle using the steering wheel and foot pedals typically used with computer games. They hear sound through headphones but vibration is not present.

This new tool however required new methodologies to cater for sources of variability and optimise its use within NVH target setting processes. The purpose of the research was to propose methods for setting up and conducting subjective evaluations using interactive simulation and provide tools to interpret the new forms of data available. To achieve this, it was necessary to propose approaches that would complement and enhance Jaguar & LandRover's (J&LR) existing NVH target setting processes. It was also necessary to take into consideration principles of experimental psychology and how the assessor perceives the use of the simulation tool.

1.2. Aims

Aims for the research programme were formulated with the project partners J&LR and SVT. J&LR were interested in the inclusion of the simulator into their existing NVH target setting processes. They would need to know how to use interactive simulation in sound quality evaluations, and how to optimise the collection and use of data. SVT would also be interested in a methodology, and the development and exploitation of new opportunities for the technology underpinning the simulator. This would involve introducing new methods to capture how people evaluate sounds using the simulator.

The aims were therefore as follows:

1. Propose guidelines on how to set up and conduct sound quality evaluation in the simulator.
2. Develop methods and tools which most effectively exploit interactivity in sound quality evaluations.

The ability to generate existing and concept vehicles, and drive them interactively, in the simulator, before the manufacture of physical prototypes, is innovative. Using this

approach, new ways of conducting sound quality evaluations, and using the data collected, have been developed.

1.3. The approach

An iterative approach to developing a methodology for the simulator was conducted. Each stage of the research facilitated the next.

The research began by benchmarking how the simulator was being used by the developers. Through this, it became evident that a rigorous approach to setting up evaluations had not been adopted. The benchmarking stage was followed with a comparison of the original methodology against methods from experimental psychology, specifically how generic subjective evaluations should be set up and conducted. This provided the insight necessary to begin setting up the evaluations, which provided a greater understanding of how people perceive the simulator-based evaluations. This stage was followed by understanding methods to capture subjective opinions. It was then possible to use the processes identified to further research how to improve the exploitation of the interactive aspect of the simulator.

Throughout the research programme, contributions were made to the developers, who incorporated the approaches proposed into the simulation software.

To support the literature, discussions were held with psychologists, experienced NVH engineers and sound quality practitioners, from J&LR, SVT, LMS International Ltd. (a service provider of NVh simulation software) and General Motors, and also statisticians and signal processing engineers. In addition the conventional methods for sound quality evaluations were experienced. This was necessary as the literature related to interactive simulation was limited.

1.4. The EngD Portfolio

During the course of the research, 8 submissions were completed. Figure 1 summarises how to review the EngD portfolio content. The first document that should be read is submission 1, this is an introduction to sound quality, the interactive NVH simulator, and the need for best practice guidelines. Also within this submission is a review of the five research areas that needed to be addressed: The assessors, the interactivity, the context; the data capture methods and the corresponding analysis techniques.

The second submission that should be read is submission 2. This documents the design and commissioning of the simulator testing facility at the University of Warwick. The

purpose of this document is twofold. Firstly it discusses how the NVH simulator used in this research is built and what the facility looks like, and reasons for the facility's appearance. Second it discusses the author's personal experience of project managing this stage of the project.

Submission 2 should be followed by submission 3. Within this document, NVH target setting processes are discussed. These were investigated, as potential guidelines had to be integrated into existing processes. Also reported, is how the simulator was being used by its developers before this research programme. In doing this, it was evident that rigorous experimental designs were not being applied, and therefore sources of experimental error were being potentially introduced. Recommendations were made to avoid these from occurring in the future.

Submission 4 is the next document to be read. This document focuses on the assessors' perception of the evaluation experience with the simulator. A methodology for the simulator would rely on the integration of people into its processes. How they interact with the simulator, and how they perceive the experience could dictate potential technological improvements which enhance their well-being, confidence and enjoyment of the evaluation. This is important if they are to focus on the task of evaluating sounds and be willing to return for future evaluations. The outcomes of this stage of the research identified that simulator sickness, a form of nausea induced from exposure to virtual environments, was affecting 10% of assessors. Recommendations, by the authour, were made to minimise this. Also during this study, an attempt was made to observe driver behaviour, to further understand the interactive evaluation. The method implemented did not facilitate the capture of the assessors driving strategy. This prompted the development of the simulator software and development of new techniques for observational studies. These developments are discussed in submission 5, which should be read next.

Submission 5 reports on the development of a Driver Observation Module (DOM). This was used to collect data from the simulator's performance model during the evaluation and capture how the assessors used the data capture interface to record their preferences. Use of this type of data has not been found in published literature therefore new visualisation approaches were developed to identify how assessors drive when evaluating sound quality. Following the development of these techniques it was then possible to explore the influences of interactivity and optimise its use. This is discussed in Submission 8.

Submission 6 however, is the next submission to be read. This concerns the introduction of analysis techniques for subjective impressions formed during the evaluation. These had not been previously implemented by either J&LR or SVT. These analysis techniques facilitate decisions by NVH engineers, by associating a statistical significance to the results of evaluations. The benefit of this is enhanced confidence is knowing whether or not the observed differences between sounds are due to random variation or not. This is the last submission which provides guidelines on how to design sound quality evaluations using the simulator. The research conducted up to this point enabled further investigations into interactivity. This is discussed in submission 8.

Submission 8 discusses the influence of interactivity on the outcomes of the evaluation. Here it was found that stimuli used in the traditional evaluations, did not appropriately represent the attributes assessed. The tools developed and reported in submission 5 were applied, to demonstrate how to make the most of the driver behaviour data. This document shows how the new forms of data collected which can be used to understand how people make decisions.

Submission 7 and 7.1 are the final submissions to be read. These are a summary of the published material which resulted from this research project. At the time of writing this report the author contributed to the content of 5 international conference papers and 1 journal paper. Four of the conference papers concern sound quality evaluations using interactive simulation. For the remaining conference paper the author contributed towards research into how to use lexicon data from interviews to select inputs for artificial neural networks. This was used to predict the outcomes of sound quality evaluations using a listening room. For the journal paper, the author of this report contributed towards specifying the research areas that needed to be addressed by the project.

Also included within the EngD portfolio is a personal profile. This summarises how the author has developed in the EngD's prescribed competences. Evidence for each of the competences is provided either through one of the 16 post module assignments completed, or reference is made to the project work conducted and reported within the submissions. Figure 1 summarises how to review the EngD portfolio content.

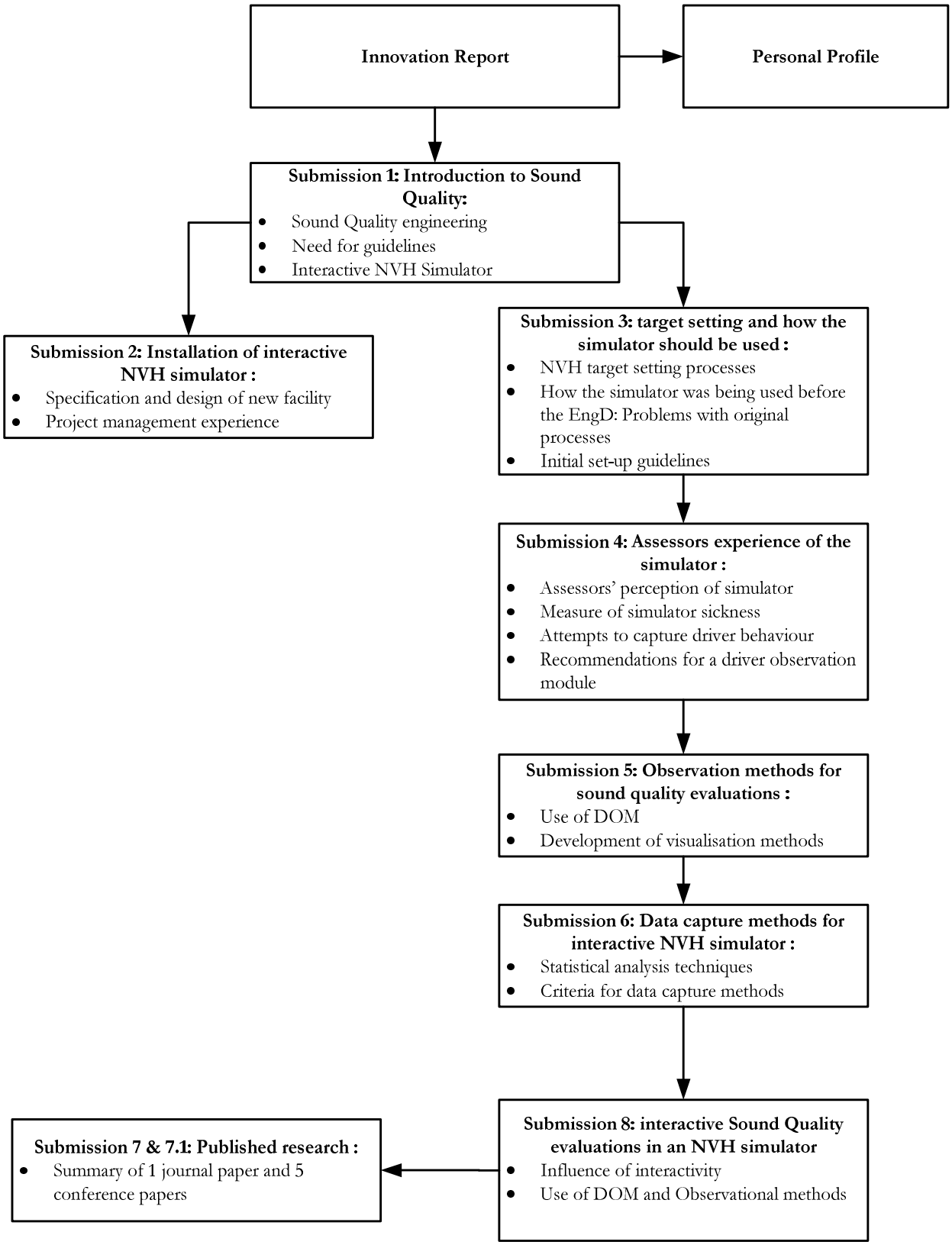

Figure 1: How to read the EngD portfolio

1.5. Structure of this Innovation Report

Chapter 1 has already introduced the report and the structure of the EngD submissions.

Chapter 2 is an introduction to sound quality. Here a definition for sound quality is given and its value is discussed. Sound Quality Engineering (SQE) processes are also discussed, including target setting and target sound generation. This is followed by an introduction to jury evaluation methods and evaluation environments. This includes the introduction to interactive simulation and the problems associated with the methodologies used before this research.

An integral part of the evaluation process is the human assessor. Chapter 3 explores how humans make decisions, and how traditional evaluation environments do not cater for their inherent decision making processes. The background information underpinning the topics discussed in chapters 2 and 3 was needed to identify the research areas that would need to be considered within this project. It also presents the specific research questions identified through further discussions with SVT and J&LR.

Chapter 4 presents the research concerning the assessors' perception of the simulator and the task of assessing sound within it. Chapter 5 reviews how to set up evaluations. It provides future sound quality practitioners guidance on how to set up and run evaluations in the simulator, thus ensuring the outcome of the evaluation is free from experimental errors.

Chapter 6 reports on the research concerning influence and exploitation of the interactivity aspect of the sound quality evaluations. Shown here was a novel method to capture assessor behaviour and how to use this information to provide NVH engineers with a deeper understanding of how assessors make decisions.

Chapter 7 is the discussion of how the research conducted over the registration period has changed the way in which the simulator developed as a product and how it should be used in industry. Chapter 9 is a summary of the guidelines developed.

Chapter 9 focuses on the future work. The recommendations made are being followed up in APPRAISE project (Appropriate Product Representations for Assessment in Structured Evaluations) a £600,000 project funded by WIRMC, with J&LR and Bruel and Kjaer Ltd (a leading manufacturer and distributor of NVH solutions) as partners. The final chapter concludes by summarising the major impact of this research programme.

2. SOUND QUALITY ENGINEERING

2.1. Sound Quality

Intense competition within the automotive industry has led to manufacturers exploring many approaches to make their product more appealing to prospective customers. The design of the components that make up the car should communicate product quality and reinforce the manufacturer's brand image. Understanding customer opinion early in the development programme and presenting this in a suitable format can guide designers, engineers and key decision makers to choose between concept designs with confidence, such that the end product will achieve the optimum impact on the customer.

The customers' opinion of a car can be influenced by their perception of stimuli. The feel of the materials used in the cabin and the comfort of the seat are only a few of the aspects of a car which can contribute towards an overall feel of product quality.

sound quality is the area which focuses on how the vehicle's interior sound is perceived by the customer. Sound can communicate functional states and give an indication of the vehicle's health. It can also be used to strengthen positive aspects of the driving experience.

Historically, manufacturers have focused on lowering noise levels within the cabin. If the noise was sufficiently low, it may have been perceived as a sign of good build quality. Today however, the emphasis is not only on the removal of noise, but it is also focused designing sound to create a positive emotional response. Taken together, the emphasis of NVH engineering has shifted from being objectively led to focusing on subjectivity (Repik 2003).

Sound quality engineering can therefore be defined as the process of tailoring and enhancing a product's acoustic character, to support the customers' perception of overall product quality.

The emotional responses provoked by a car vary depending on the driving conditions and the type of car. Different functional states require different acoustic characteristics. When performing full load accelerations, customers expect the sound to reinforce the perception of the vehicle's dynamic character and its responsiveness to their control. At cruising, or at idle, they may expect the car to feel refined and luxurious. The degree of expectation of each of these qualities will depend on the type of vehicle. For a sports car, levels of luxuriousness can be compromised for improved perception of the dynamic character. On the other hand, in an executive saloon, the feel of refinement and

luxuriousness will take priority. The aim for the Jaguar brand is to achieve a sound with dual characteristics consisting of both a powerful and refined character (Dunne 2003).

An understanding of the value of sound quality has been gathered from automotive journalism, manufacturers' opinion, and consumer psychology.

It is rare to see car reviews of luxury saloons and sports cars which do not describe the sound of the car. As part of this research, car reviews written by journalists were monitored to see how many of them discussed the sound of a vehicle. This was done to gauge how important the sound of a car is from the media's perspective. This was conducted as the author believed that media can influence purchasing decisions. Out of a sample of 18 car reviews, a description of the vehicle's sound was not discussed in only 3 articles. It was also, however, deduced that the degree of importance of the sound depended on the car. For example, the sound of a sports car was considered more important than that of a hatchback.

Harley Davidson is a company whose products have a strong association with their sound. The aim of its product design process is to achieve the "Harley Sound" (Pierson 2003). They claim that customers are attracted to its motorcycles because of the perceptual feelings they experience whilst riding the bike, specifically how it looks and sounds. In this case, the "Harley Sound" differentiates its brand from others, so much so, the company believed it was worthy of a patent, which it applied for, but was not granted (New York Times 1995).

Within the automotive industry, Afeneh et al (2007) acknowledge that NVH attributes are not the main concern within the vehicle development programme. The attributes such as styling, safety and reliability are often the top priority. Usually NVH attributes are within the top 5, but this will depend on the type of vehicle. Hutchins et al (1992) from Lotus noted the effect of not matching the sound character to the vehicle. He implied that drivers' behaviour can be influenced by the sound, and they would avoid driving naturally to avoid the acoustic features they do not like.

Also from the automotive industry, Dunne (2003) illustrated the value of the product sound using the Kano Model. His interpretation of the value of sound quality is shown in figure 2. It is stated that the customer expects a basic level of NVH performance. These may be levels of noise and vibration that are unacceptable or if there are squeaks and rattles present. The achievement of this quality is illustrated by the "basic quality" line. Further improvement of these levels of noise and vibration does not induce proportionally increased levels of customer satisfaction. Beyond the achievement of the basic quality is the "spoken performance". This is referred to as the reduction of noise

and vibration. Decreasing these levels proportionally increases the level of customer satisfaction. However higher levels of customer satisfaction can be achieved with lower levels of noise isolation, if the sound has certain delighting qualities. That is if the sound is contributing to the customers' excitement during the use of their product. This type of quality is illustrated by the "excitement quality" line shown in figure 2. Dunne states that both the excitement quality and the spoken performance quality are achieved through the Sound Quality engineering process.

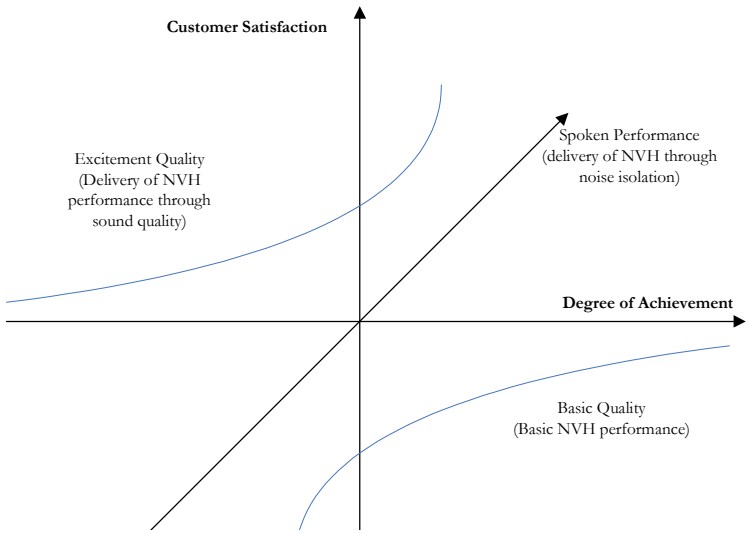

Figure 2: The Kano value of sound quality (Dunne 2003)

Manufacturers also try to draw the customers' attention through the media to the fact the sound has been engineered. When BMW launched the Z4, its adverts deliberately made reference to the engineering behind the sound of their vehicles. Freeman (2003) stated that BMW purposely engineered the sound depending on the type of the vehicle. It stated that if the vehicle was a luxury saloon the emphasis was on delivering noise and vibration isolation. On the other hand, if the vehicle was a sports vehicle, such as the Z4, then it would purposely increase the interior sound to enhance the enjoyment of the drive.

Researchers such as Schifferstein (2005) (2006) have shown how the importance of a product's sound is related to the products function. For products such as cars, vacuum cleaner and televisions, where the sound is a result of the products' function, the sound was considered highly important by users.

The sound of a vehicle is therefore a valued characteristic. Manufacturers direct investment towards research into sound quality methodologies and the development of

state-of-the-art simulation tools. The intention in this work is to support engineers who are largely accustomed to dealing with objective measures. In the same way as an artist's concept designs are turned into engineering drawings, these subjective impressions need to be used to set engineering targets.

2.2. Sound Quality Engineering

SQE is the process of delivering the vehicle's defined acoustic character. It is a discipline that needs to be embedded within the practice of Noise, Vibration and Harshness engineering (NVH), which as the terms imply, deals with those aspects of the vehicle. Typically the main concerns for NVH engineers are powertrain NVH sound quality, road NVH and wind noise. Other considerations include "operational sound quality" which concerns aspects such as seat belt and door closing sounds, seat motors and wind screen wipers. Also squeaks and rattles, which are considered error states, are also included in this. In addition, pass-by noise legislation needs to be accounted for.

This research has focused on the processes related to the subjective evaluation of powertrain sound quality as perceived from the driver's seat. This location is the so-called "auditory space" (Goldstein 1999) experienced by the customers. The sound heard here is a combination of the engine, road noise, the exhaust, the intakes and wind noise. Examples of the components that can be altered to manipulate the sound are shown in table 1 (Harrison 2004).

Source	Structure Borne	Airborne
Engine	Engine mounts	Firewall / floor
	Engine vibrations	
	Firewall / floor absorption material	Absorption material under bonnet
	Engine components	
Exhaust	Exhaust mounts	Absorption material under bonnet
	Floor absorption material	
	Coupling between catalyst and rest of exhaust	Door seals
Intake	Filter box and snorkel mounts	Firewall / Floor
	Door seals	Absorption material under bonnet
Road	Tyres	
	Suspension bushes	Firewall / floor
	Subframe bushes	Door seals
	Firewall / floor absorption material	
Aerodynamic Noise	Shape of the vehicle	
	Door seals	

Table 1: Components which can be altered to manipulate sound (Harrison 2004)

Other approaches to modify the sound of the car can include Active Noise Control (ANC). By emitting sound from the vehicle's speakers it is possible to cancel out or enhance particular sound characteristics. Traditionally J&LR considered ANC not appropriate for their product. They felt that customers of the Jaguar brand would consider this a deception of the product quality. However recently they have started using mechanical components that actively enhance the acoustic properties of the vehicle. These may have no impact on the performance of the vehicle. For example, on the latest XK, an intake feedback system is fitted, otherwise known as a bark tube. This device helps tune the sound of the car from the driver's perspective. It is a tube which runs from the intake manifold to behind the dashboard. Air pressure from the tube vibrates a small diaphragm piping engine sound into the cabin.

2.3.NVH Target Setting

Before an NVH engineer can begin engineering the sound of the vehicle, targets have to be established. This means deciding how the new vehicle will sound. Setting these targets is conducted at the beginning of the vehicle development programme, in advance of any detailed design effort. The decision of how the vehicle will sound is based on a mix of

engineering requirements and the opinions of customers, engineers and key decision makers within the organisation. (Afaneh 2007). However poor decisions at this stage can be expensive to amend once a prototype has been manufactured. To avoid these, processes to achieve the target sound have been developed. A visual tool used for the Jaguar brand, is a 2D map shown in figure 3 (Dunne 2003). Here the horizontal axis displays a measure of the subjective achievement of the "powerful" character of vehicles' sound. The vertical axis displays a measure of the subjective achievement of "refined" character. Powerfulness and refinement are the two verbal descriptors which represent the Jaguar brand during subjective sound appraisals.

The 2D map was constructed using results from a paired comparison evaluation in which 24 vehicles were assessed. The merit scores (see submission 1) were used as a measure for the position of the vehicles on the map.

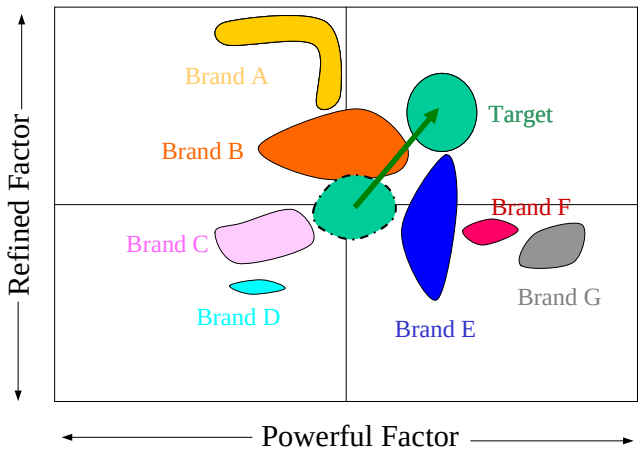

Figure 3: Target setting processes at J&LR

Dunne (2003) found it interesting that individual cars from each brand form clusters. However it is important to note the positions of the vehicles are relative not absolute. That means that if the group of brands used in the evaluation were different, it would result in different locations for each of the brands.

By displaying the results of subjective appraisals using the 2D map, decision makers can visualise how their own existing and prototype vehicles compare against current vehicles and competitors. It then allows them to identify target areas they wish to occupy with future vehicle developments. Their choice of target area is based on knowledge of how achievable the target is and what engineering effort is needed. Decisions are usually based

on understanding the compromises between other vehicle attributes such as handling and performance.

The descriptors powerfulness and refinement were identified using principal component analysis (PCA). Other sound quality practitioners use different descriptors, but with a similar meaning. Bisping (1995) for example identified the descriptors "power" and "pleasantness". Terazawa et al (2000) identified the descriptors "raciness" and "pleasantness". Both Bisping and Terazawa also used PCA to identify their descriptors. Both claim that if the process is conducted correctly, the descriptors that will be identified will have similar meaning to the ones above. One descriptor will usually represent the dynamic and performance of the vehicle, whereas the other will be related to the comfort and luxury aspects of the vehicle. For the purpose of the research project, the descriptors "powerfulness" and "refinement" were used throughout the project, as these are descriptors used for the Jaguar brand.

The 2D map is a useful visual aid, which can involve non-technical decision makers in the target setting process. It is a way for management to communicate to each other on a common platform. NVH engineers, however, require objectivity; they need quantifiable targets related to engineering components. The 2D map is the final outcome that needs to be achieved. It can not, however, provide the engineering detail required to develop a new sound. For this a target sound generation process is used.

2.4. Target Sound Generation

Target sound generation is the process whereby the engineers will develop a concept sound to see if it meets the target area defined from the 2D map. There appears to be no standard process for generating target sounds, as each manufacturer appears to have their own in-house approach, which is closely guarded. Jay-Cerrato (2007) provided insight into this process. Her approach, although generic, has been adapted according to the author's understanding of J&LR's processes (see figure 4). The aim of this process is to develop a model, based on objective measures, of how an assessor subjectively perceives a sound.

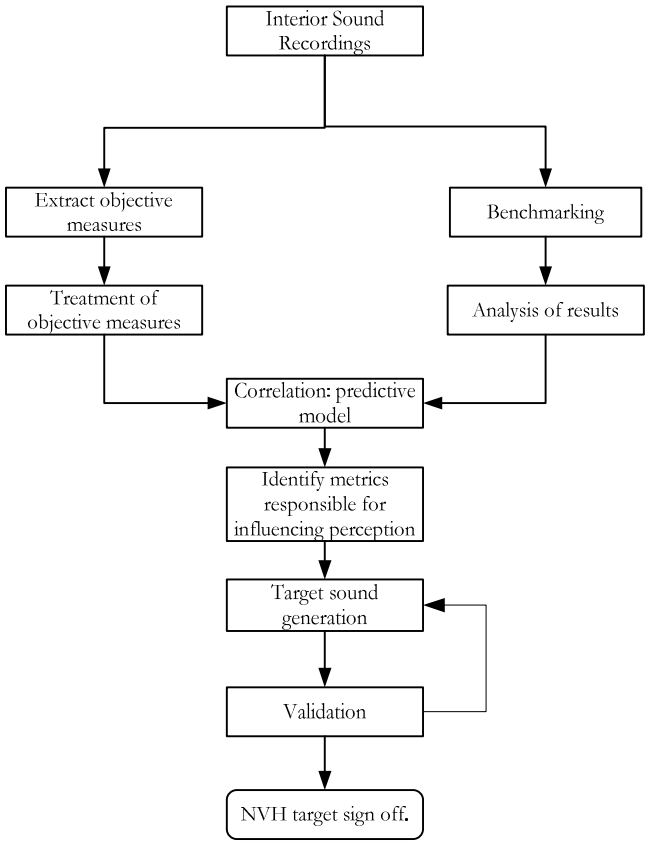

Figure 4: Setting NVH targets on subjective data

The first step in the target generation process is to select the vehicles that the manufacturer considers to be competition. This is followed by recording the interior sound of each of these cars. To ensure a like-for-like comparison between the vehicles, the recordings are made over a predetermined driving range. In the case for the Jaguar brand, the recording is made whilst the vehicle is accelerating in 2nd Gear Wide Open Throttle (2GWOT). Following this there are two paths that should be followed. The first is the extraction of objective measures from the sound. The objective measures are often referred to as psychoacoustic metrics, such as loudness, sharpness, and roughness (Zwicker 1999). The metrics are then treated so that they can be represented by a single number such as the average value.

The second path in the target generation process is the subjective assessment of the recorded sounds. This stage is referred to as "benchmarking". The opinions of brand mangers, NVH engineers and customers are some that may be taken into consideration.

The subjective impressions of the assessors are then also treated so that an overall impression of the sound can be represented by a single number. In the case where a rating scale may be used, the mean is often used. Attempts are then made to correlate the

representations of the subjective impressions with the many objective representations. It is likely that not all metrics will correlate. The ones that do, represent the acoustic features that influence the assessors' perception. This means that the engineers now have a model of how subjective impressions change with objective measures. This can then be used to understand if a target sound generated will perform well when subjectively assessed.

The target sound is created by applying acoustic filters to the recording of the interior sound. The filters represent engineering changes that can be applied to an existing vehicle referred to as a "donor". The purpose of this is to create a target sound, which may perform "better" than the donor when subjectively assessed against the competition. This second stage of subjective evaluations is referred to as validation. Here the target sound is compared against the competition and / or the existing vehicles, to assess if an improvement has occurred. If this occurs, the concept sound is used as a target which NVH engineers aim to achieve through the design of the vehicle.

Recently there has been much debate on the value of developing a model to simulate assessors' perception. The value of a metric will change according to the changes in the stimuli it was derived from. For example, if loudness was derived for a 2GWOT, then the value of loudness will be different at the beginning of the drive cycle compared to the end. This information is then lost when the metric is represented by a single number. The use of a model should not replace the task of listening to a sound and making a judgement based on experience, desired positioning of the brand and insight into the many engineering compromises that are often required. Nevertheless the need for the subjective evaluations is still significant.

The benchmarking and validation stages are referred to as jury evaluations. Here assessors listen to automotive sounds and make their preferences known. The methodologies employed within the assessment need to be carefully considered. As humans are essentially used as a measurement system, then various sources of variability need to be carefully dealt with. The integration of the human with the experimental interfaces needs to be facilitated, to ensure their attention resources are focused on the task. Jury evaluations were the focal point of this research project.

2.5.Jury Evaluations

A jury evaluation is the phrase used to describe the process of capturing the subjective opinion. For the Jaguar brand, assessors are asked for their opinion on which sounds are more "powerful" and which ones are more "refined" (Dunne 2003). There is a variety of data capture methods used to record their opinions. Otto (1999) discusses 5 types: paired comparison, rank order, rating scale, semantic differential and magnitude estimation. Although Otto's guidelines are widely accepted as the standard in the automotive industry, other researchers such as Fry (2006) and Schulte-Fortkamp (2006) used interviews to extract richer content from the assessors. Each of these methods is discussed in the next section. The selection of analysis methods used will vary depending on the data capture method. This topic is discussed in detail in submission 7 and is summarised in chapter 5.

2.5.1. Paired Comparison

With the paired comparison method, stimuli are presented to the assessor in pairs. They listen to each sound in turn, which is then followed by the by assessor making a choice between the sounds according to their own perception, and the question asked. The results can be analysed using the Bradley Terry method. Implementation of this method is shown in submission 1, and a detailed description of the paired comparison method can be found in David (1988).

Otto (1999) describes the paired comparison method as the "easiest" of the data capture methods for non-expert assessors. They have to simply choose one of the two sounds they have heard. The disadvantage of the method, however, is that as each sound must be compared against each other; the duration of the evaluation can be lengthy, especially if repeatability measures are applied. An evaluation that is too long can result in the assessor becoming tired and less attentive to the task.

To measure assessor performance, two measures can be calculated: consistency and repeatability. Consistency is measured by comparing groups of three sounds. For example, if sound B is chosen over sound A, and sound C is chosen over sound B, then it might be expected that sound C would be chosen over sound A. If this does not occur, a "circular triad" is said to have occurred, meaning the assessor has been inconsistent. For repeatability each pair of sounds needs to be presented to the assessors twice. If the same sound is not chosen on both occasions, unrepeatability is said to have occurred. It is possible to omit those assessors, who are too unrepeatable from the analysis.

2.5.2. Rank Order

The rank order method is used to place sounds in order of preference. The decisions that assessors have to make are slightly more complex than the paired comparison method, as they have to rank a number of sounds. The method is suitable when a quick analysis of a number of cars is needed, or to compare the effect of changing a component. However it is not possible to measure how different two sounds are.

2.5.3. Rating Scales

For rating scales, the assessors are asked to rate the sounds they hear on a numerical scale, say from 0 to 100. Otto suggests that this approach is unsuitable for non-expert assessors, as they tend not use the extremities of the scales in fear that the next sound they hear is either better or worse than the one placed at the extremity. This is caused by the fact that the assessor may not have a good understanding of the range of sounds available. This also leads to the different assessors using different ranges on the scales. NVH engineers at J&LR use a numerical scale ranging from 0 to 10 during the on-road evaluations.

2.5.4. Semantic Differential

This method is similar to the rating scale, but to avoid the limitations present with the use of numbers, bipolar adjectives are used at the extremities of the scale, for example "refined" and "coarse". Otto proposes that this makes the method more user friendly for the non-expert assessors, and promotes the use of the extremities of the scale.

2.5.5. Magnitude Estimation

This is also a scaling method, but sounds are rated relative to reference sound. The sounds placed on a scale are done so they are proportional to the reference sound. The use of the method requires the assessors to be trained on the use of the scale. The scale is used when correlating subjective opinions with objective measures.

2.5.6. Interviews

The interview technique involves asking the participant about their subjective opinion of the sound they hear. Coolican (1999) listed a number of advantages and disadvantages of interviews. For example the interview can provide a deeper insight into the customer preferences. It allows the interviewer to explore a deeper meaning of the assessors' comments, and the assessment is considered realistic. The limitation on the other hand is that the information collected can be unsystematic, and therefore different discussions will originate from different assessors. This leads to difficulties with the data analysis, and Coolican (1999) states that interviews are generally unreliable. This approach will be explored further, later on in this report.

2.6. Evaluation Environments

On-road tests and listening rooms are the two main environments in which jury evaluations can be conducted.

For an on-road test, existing and prototype cars belonging to the manufacturer and their competitors are driven by the assessors in any manner they choose. For prototype vehicles, engineering changes can be applied by changing mechanical components, and then assessing the vehicle again to see how it performs. The advantages of using on-road tests are that the assessor experiences all the stimuli and the sound is assessed in the correct context. They can also assess any of the driving conditions. The assessors may perform a variety of manoeuvres, as they drive the vehicle as they would drive their own.

On the other hand back-to-back comparisons of the cars are difficult. Assessors therefore may find it difficult to distinguish between the sounds. Experimental control is not easily applied, and therefore a like-for-like assessment between the cars and between the assessors is difficult to implement. This means that subjective opinion captured is dependent on evaluation conditions. If the evaluation is executed in a non-rigorous manner, then it is difficult to analyse the subjective impressions captured, which can be formed on different driving conditions by each assessor. This means that there are limited opportunities for making confident decisions.

Listening rooms are laboratories designed for the assessment of sound. They are purposely built so that the transmission of both external and internal noise sources is reduced. They are comfortable and neutral environments and offer the opportunity to conduct repeatable evaluation conditions for all assessors. A photograph of the WMG listening room is shown in figure 5. Sounds can be either replayed using the headphones

or the loudspeakers. The assessors follow the instructions on the screen at the front of the room, and up to six participants can take part in the evaluation at the same time.

Figure 5: WMG listening room

Within the listening room, vehicles can be easily assessed back-to-back at the press of a button. The sound of cars that do not yet exist can be simulated and therefore evaluated before prototypes are built. This can be done by manipulating the recordings of existing cars. These modifications represent potential engineering changes that can be made.

Experimental conditions are repeatable for each assessor, therefore rigorous analysis techniques can be applied which complement the OEM's decision making. The limitation however is that sound is assessed in isolation from the other stimuli, resulting in more attention being paid to the sound stimuli than would occur in a real car. Genell et al (2007) discusses the concept of *"cognitive capacity hypothesis"*. This is based on the assumption that a human has limited cognitive resources that are shared amongst the stimuli. Depending on the context of the situation they are likely to shift their cognitive resources to cater for the situation. Therefore if a task is dominated, say by assessing sound alone, all the attention resources will focus on the acoustic stimuli the ears receive, than say the information they see.

Another limitation of using the listening room is that fixed operating conditions are often assessed. For target setting for the Jaguar brand, fixed driving conditions of the sound of a vehicle accelerating in 2GWOT is often used as stimuli. However it is dangerous to assume that issues resolved at this full load condition result in equivalent resolution at part load, or indeed cruise or idle (Dunne 2007). This can therefore lead to errors being

made with regards to engineering changes needed to achieve overall vehicle sound quality.

2.7. Interactive NVH Simulation

It is therefore evident that on-road tests can only be used as initial benchmarking exercises, and then used again to check when the first prototype is produced. On the other hand, the listening room can be used to test the cars that do not exist. A recently developed method is interactive simulation. This enables sound to be presented to assessors whilst they drive through a virtual environment. This can be experienced whilst either sitting in a fixed base vehicle or at a desk.

The fixed base full vehicle simulator is a stationary vehicle positioned in front of a screen onto which a virtual environment (VE) is projected. An assessor sits in the car, and "drives" through the VE. Whilst driving, the assessor listens to the interior sound of the car through headphones. Vibration is felt through the seat, the floor panel and the steering wheel, as these are the contact points between the assessor and the simulator. An actual vehicle has been used for the WMG simulator. It is a Jaguar XJ with reinforced windows and body panels which minimise the transmission of external noises.

The author of this document was responsible for the design and commissioning of the interactive simulation facility. The facility was built next to WMG's existing listening room facility. This had a dedicated waiting room that is now shared with the simulator facility.

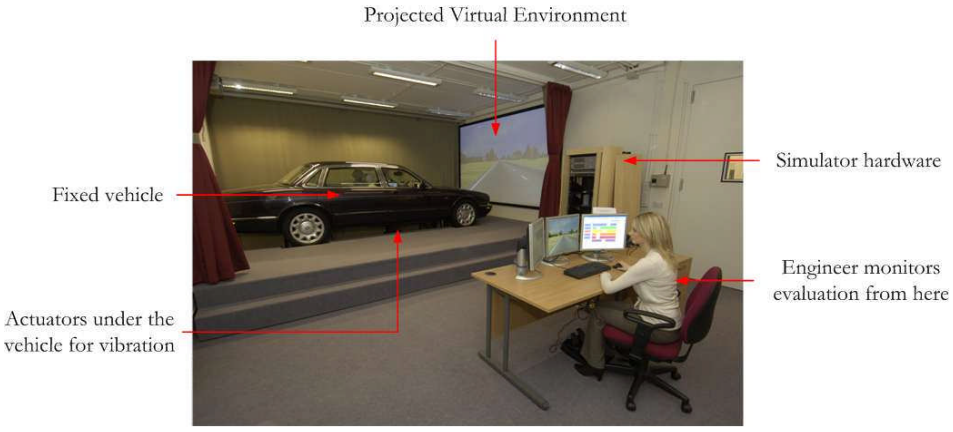

Figure 6: Exterior of WMG's interactive NVH simulator

The projection screen is aligned with the driver's position and close to the vehicle so they can not see the stationary floor in the place in between. There are two layers of black out

curtains that enclose the vehicle, so the driver can not be distracted with any events outside, nor can exterior light deteriorate the quality of the projected image. The lighting in the room is adjustable. The floor surrounding the vehicle is split into three platforms, all loaded onto wheels, so that it is possible to service the underneath of the vehicle if required, as this is where the power and signal cables enter the car.

Inside the vehicle there is a touch-screen monitor located on the passenger seat, where the evaluation interface is displayed. Using this interface the assessor can choose which vehicle to drive, and record their subjective opinions. The interior of the vehicle is shown in figure 7.

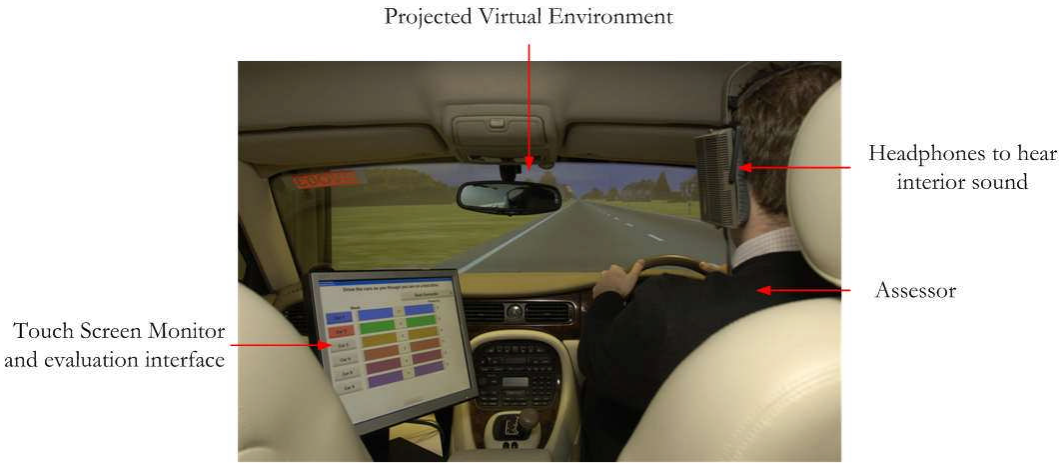

Figure 7: Interior of WMG's interactive NVH simulator

The desktop simulator is similar to the vehicle simulator. The vehicle is controlled by the assessor via a computer game steering wheel and foot pedals, whilst sat a desk. The subjective opinions are captured on a monitor using a mouse.

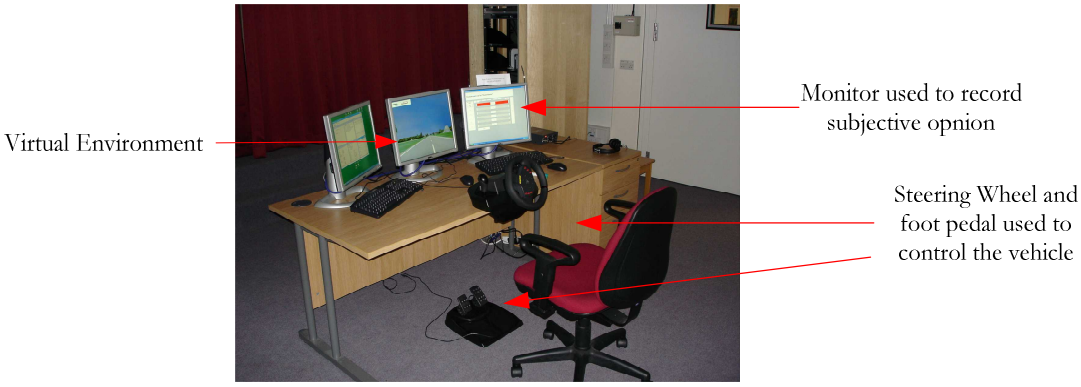

Figure 8: Desktop simulator

The computer hardware for the simulators is in a sound proof cabinet located in front of the desktop simulator. The amplifiers and the graphic equalisers can also be found here, amongst other items. Overall the facility is neutral, with the exception of the Jaguar branding on the vehicle.

With both simulators, accurate sound is generated in real time, in response to the driver's behaviour; this means that the assessor can drive the simulator as they would a real car. The assessor responds to visual, auditory and tactile cues, through the use of the throttle, pedals and the touch screen. These responses are converted into inputs to the simulator's performance model, which turns them into physical representations, which are converted into new audio, visual and tactile cues. These are then experienced by the assessor and the process continues (see figure 9).

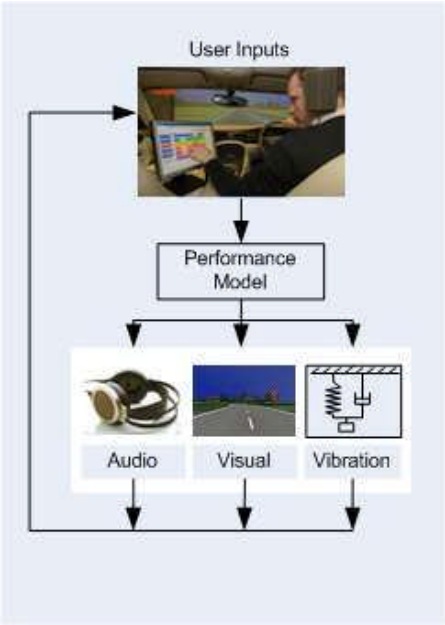

Figure 9: Control of the interactive NVH simulator

The intended purpose of the simulator is to offer an alternative to the listening room. The ability to drive is made possible through the use of the virtual environment and the real-time audio generation. The ability to drive the concept cars is what makes the evaluation more closely related to on–road driving. The assessor is allowed to explore all driving conditions, as they are able to do in a real car. This means that they are no longer focusing on a single drive cycle such as 2GWOT.

This ability to drive is reliant on how the sound is generated. This is through a hybrid of actual recordings and CAE data.

To generate the sound of a vehicle in the simulator, its sound is recorded at speed increments of 10mph on a test track. This creates a number of very short audio files. Depending on the operating conditions of the simulator, the corresponding audio file is played. For the operating conditions where the data would be missing, say between 10 mph and 20 mph the sound heard is a blend between the two existing audio files.

The second feature of the simulator software is its ability to "break down" the sound recorded. This way the individual contributions that make up the sound can be heard and manipulated individually. This process is known as decomposition. This means that each of the mechanical components shown in Table 1 has acoustic properties whose effects can be heard independently and as a whole. Further details of the decomposition process can be found in Allman–Ward (2003) and Crewe (2003). The decomposition process is illustrated in figure 10

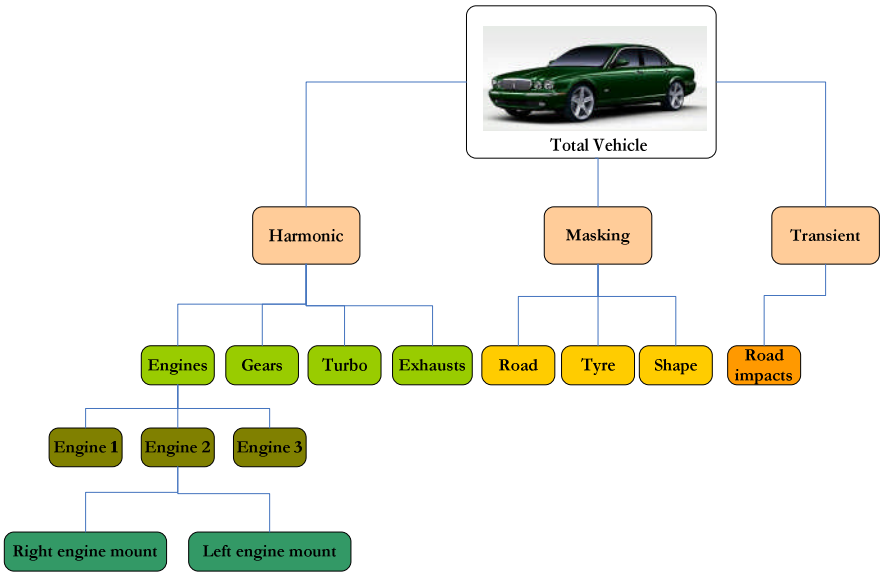

Figure 10: Decomposition process

2.7.1. Other simulators

At the time of writing this report, there appeared to be three simulators in the market place. The one discussed above was developed by Sound & Vibration Technology Ltd, and is now distributed by Bruel and Kjaer (B&K). This is the simulator that has been used as part of this research project.

Head Acoustics also supplies a similar set-up. They have various versions of their simulator. One version is very similar to the simulator developed by SVT, using a

stationary full vehicle simulator and a desktop simulator. However, other versions have recently been developed so that the simulator software can be fitted in real cars; these appear to be developments resulting from the OBELICS project (Objective Evaluation of Interior Car Sound). A European funded project involving two OEMs, three engineering consulting companies including Head Acoustics and LMS, and two German universities. The functionality remains the same as above. It is possible to generate the sound of target and existing vehicles in real time. Details of the Head Acoustic simulator can be found in Genuit (2001 and 2007) and Schulte-Fortkamp (2007).

Another Simulator was discussed by Goetchius (2001) from Daimler Chrysler. At the time it was said to have problems recreating high fidelity vibration, and there may be reason to believe that it eventually developed into the system sold by Head acoustics.

LMS has also developed a simulator (Janessons 2006), but the available literature is limited as it does not give a full description. The papers that do describe it suggest that it is being used as an engineering tool, and not for subjective evaluations. However this statement remains unfounded until further details can be found.

There are a number of other simulators, which appear to be more of machines developed by OEMs, for generic in house testing. An example of this was from Mazda (Fukuhara 2002). In addition there are a number of other sound and vibration rigs, which play-back the stimuli, and do not allow interaction; an example of this is Ford's vehicle vibration simulator (VVS), use of this is described by Pielemeier (2001) and Amman (2005).

2.7.2. Validation of Interactive NVH simulators.

The two companies actively promoting the use of simulators are B&K and Head acoustics. Validation tests have been conducted on both these simulators, that is checking that subjective impressions formed on-road, can be replicated in the simulator. Allman-Ward (2004) showed that the rank order between subjective results from their simulator was similar to those of the on-road tests. Genuit (2007) conducted a similar test using the Head Acoustics simulator, and found significant correlation between simulator and on road evaluations

2.7.3. Target setting using interactive NVH simulation.

Whilst it is unlikely that the interactive NVH simulation will replace on-road testing altogether, they are likely to offer alternative opportunities for the assessment of NVH attributes.

In a presentation to an EIS conference in 2005, Roger Williams, a developer of the SVT / B&K simulator demonstrated when concept cars can be assessed, compared to when real cars are available this is illustrated in figure 11.

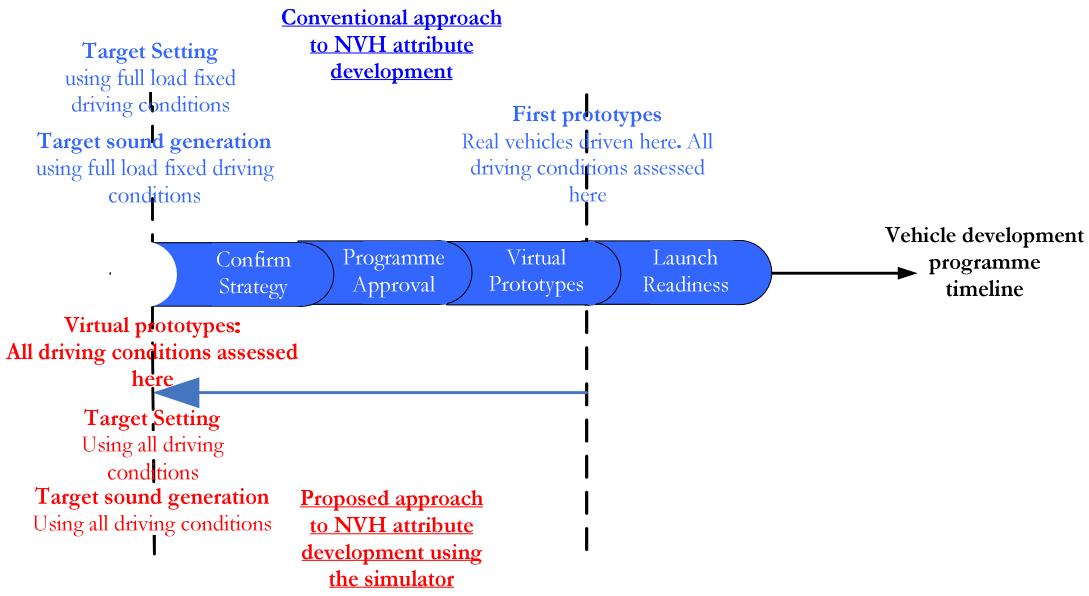

Figure 11: Simulation period within the vehicle development programme

With the use of the simulator it is possible to experience the vehicle before the first prototype is built. Therefore the right products can be designed 'right first time' by considering part load assessments as well as the traditional 2GWOT manoeuvres.

Williams et al (2006) also showed how the simulator could benefit the target setting and concept sound generation processes, which does not need the identification of psychoacoustic metrics. The proposed method is a four step approach and is illustrated in figure 12.

1. Step 1: Use the simulator to capture customer opinion through jury evaluations. This is equivalent to the benchmarking stages.
2. Step 2: NVH engineers and the key decision makers use this insight to design a concept sound, using a database of simulator component models.
3. Step 3: The ideal concept sound is cascaded down to the component level and negotiations can take place with the engineering design team.
4. Step 4: The concept sound is validated using another jury evaluation; this is the equivalent to the Validation stage shown in figure 4.

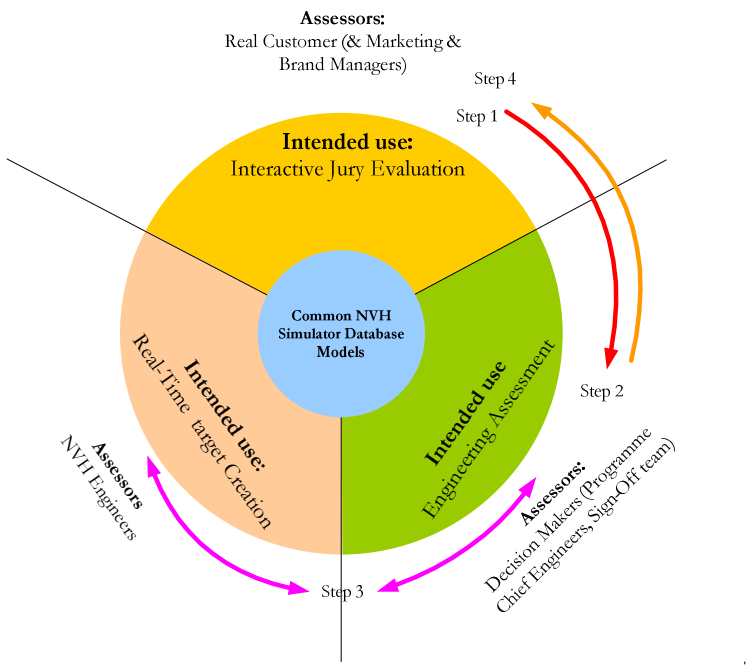

Figure 12: Potential concept generation process

2.7.4. Methodologies for interactive simulations.

It was evident through the discussion in section 2.7.3 that the jury evaluation is a key element of the NVH target setting process. Prior to this research project the interactive NVH simulator had been used in a vehicle development programme (Williams 2005) using provisional methodological approaches. The methods implemented were, however , proposed in an ad-hoc manner with little evidence justifying their application. The original approaches were initially discussed by Allman-Ward (2004), and are reviewed in submission 3.

It was apparent that two levels of interactivity existed; these are referred to as "fully interactive" and "non-interactive". For the fully interactive mode the assessors can use

the simulator as a real car. As it responded to inputs such as throttle use and steering wheel, the assessors' subjective opinions were captured on an interactive scaled interface (IS) (see figure 13). This interface facilitated a back-to-back comparison with the vehicles. For example the assessors could be driving car A at 70 mph, and then using the interface, they could instantly switch to car B, and all they notice is a change in sound and vibration. The driving conditions remain constant. The interface is shown in figure 13

Each car is represented by a button, which when pressed permits the listener to hear the sound of that car. Each vehicle can be rated by using the sliding scale. The reshuffle button rearranges how the sounds appear on the screen into rank order. The assessors are allowed to choose between the use of adjectives or numbers to rate each of the vehicles between 0 and 100.

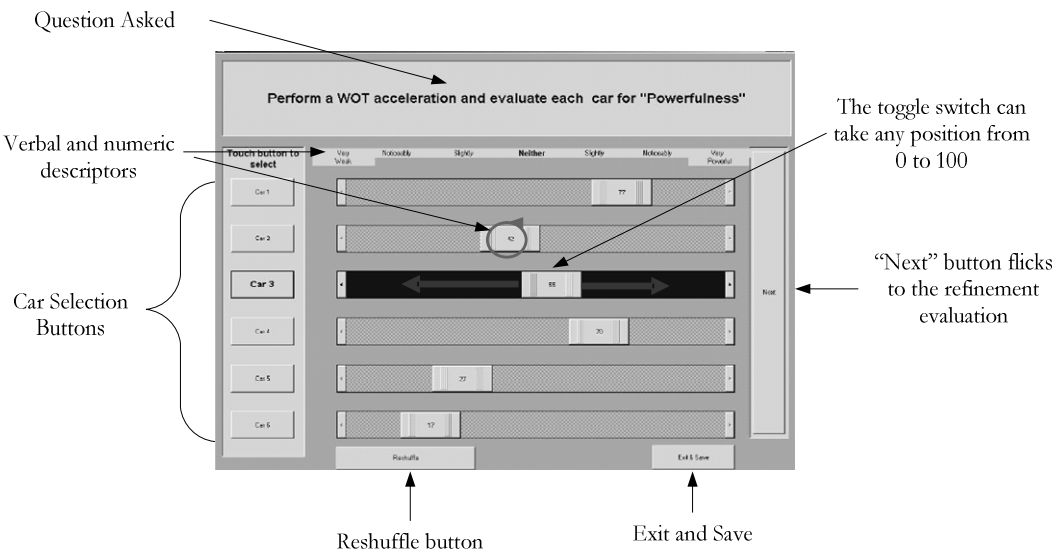

Figure 13: Interactive scaling method

For the non-interactive mode the assessors could not interact with the simulator. The assessor listened to vehicles performing 2GWOTs. They could not control the steering, or the speed of the cars they listened to. Their subjective opinions were captured using the paired comparison method (PC). With this method vehicles' sounds were presented to the assessor in pairs. The assessor was asked to choose the most "powerful" one, and then the most "refined". Figure 14 illustrates the methodologies employed for each level of interactivity.

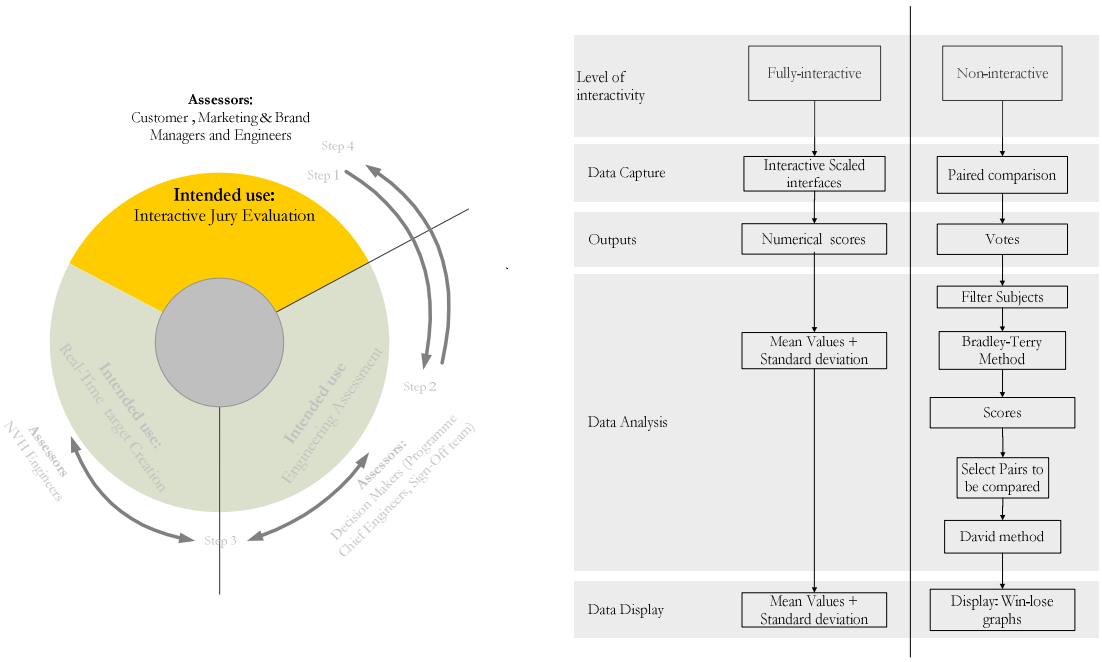

Figure 14: Original methodology

The non-interactive approach is similar to that employed in conventional evaluation methods employed within listening rooms. These take into consideration the guidelines proposed by Otto (1999). These are well established and are often referenced by many Sound Quality practitioners. These guidelines are considered by many to be best practice for Sound Quality evaluations; however they can fall short when applied to new technologies such as the simulator. For example they do not give insight into how to optimally use interactivity, or which data capture interface is most appropriate. They do state that the assessors should be made comfortable, however how does one do this when using a virtual environment? And what additional factors need to be taken into consideration such as interactivity? In the simulator the assessor not only makes a response through the data capture interface, but also through the use of the vehicle. This, as will be shown, can have implications on how the subjective impressions formed by the assessors are interpreted.

Towards the end of the EngD research programme, Head Acoustics proposed equivalent methodological approaches for evaluations in their simulator, which are also very similar to the ones illustrated above. Their process is shown in figure 15.

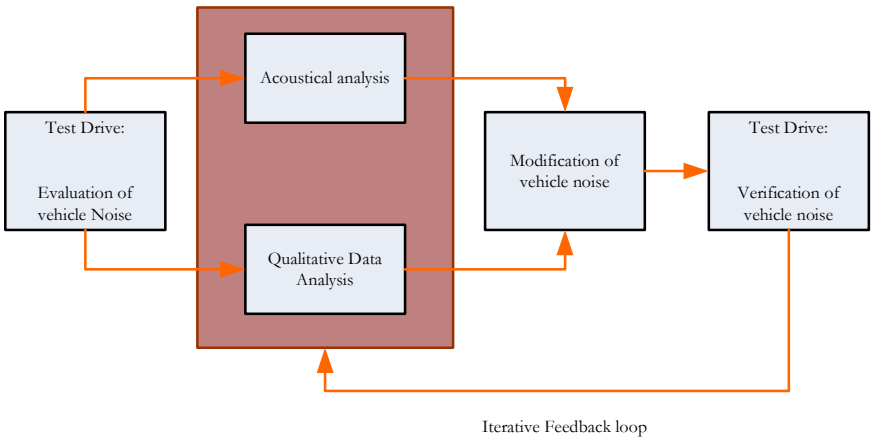

Figure 15: Target setting processes proposed by Schulte-Fortkamp (2007)

Their means of data capture was to conduct an interview with the assessors whilst they are driving. The verbal data is then analysed to interpret which acoustic features the assessors were referring to. However there are some fundamental issues within this approach which means it is difficult to use it alongside the NVH target setting processes discussed so far. It is difficult to relate the comments, from the interview, to key acoustical features. Fry et al (2006) used this method to identify the reasons why assessors choose sounds according to how powerful and how refined they are. Issues were however caused in the analysis of the results. Customised methods had to be developed (by the author of this document) to convert lexicon data into ranked data which was compared to objective psychoacoustic metrics (Fry et al 2006). It was an analysis method, whose use could be justified, although it lacked the ability to apply a measure of statistical confidence within the results

In addition, each assessor is likely to verbalise their feelings using different approaches. Some assessors may be less vocal than others. The form of the data itself can be troublesome. Processing transcripts can take a long time, and capturing key themes in the transcripts can be challenging and equally time consuming. Furthermore there is limited scope to provide a measure of how much more one sound is preferred over another. Hence their use is limited within the NVH target setting processes.

However interviews do have a place in sound quality assessment. If these are conducted prior to establishing the NVH target setting processes, it is possible to identify a list of key words that are used by assessors to evaluate sounds. Then using the principal component analysis, the dimensions of the assessment can be defined, as was conducted

by Dunne (2003), when he identified the descriptors "powerfulness" and "refinement" for the Jaguar brand.

Given this, guidelines for sound quality evaluations using interactive simulation are needed to ensure experiments are set and conducted in a scientific manner, and which exploit the interactivity functionality. In addition these guidelines have to complement existing target setting processes at manufacturers such as J&LR.

2.8. Summary of chapter

- The sound of a vehicle can influence the drivers' perception of the brand and the vehicle. Therefore a vehicle's sound is an important characteristic that needs to be considered during its design.
- It is important to establish a target sound before the detailed design phases of the vehicle development programme.
- NVH engineers therefore need to understand what influences the customer's perception of sound.
- To achieve this, evaluations are conducted, wherein assessors compare sounds and their subjective impressions are captured.
- Traditionally evaluations have been conducted in listening rooms and with on-road tests. Both of which have limitations.
- The simulator offers an alternative approach to the conventional evaluation environment, by providing the potential for experimental control and the ability to evaluate existing and concept cars in real time.
- This new tool requires a formal methodology for its use within existing target setting processes.

The next chapter discusses how people make decisions and how the research areas pursued in this project were identified.

3. HOW PEOPLE MAKE DECISIONS

3.1. Introduction

Within a jury evaluation people are asked to make decisions based on their subjective impressions of the sounds they hear. The task that they are asked to complete must therefore facilitate their decision making processes without confusing or distracting them from the objectives of the evaluation. So just as previous chapters looked at how OEMs make decisions it is necessary to understand how assessors might make decisions. Both of these topics helped focus the research areas that needed to be tackled to formulate a methodology for the simulator. The purpose of this chapter is to show the rationale behind the origins of the specific research questions.

3.2. Decision Making

The main difficulty of the sound quality engineering process is that it requires humans to inform engineers of their subjective opinions. This means the engineering processes need to be able to deal with subjective data, which cannot be treated and interpreted in the same manner as engineering data. Instead the methodologies used to extract the information from an assessor need to be sufficiently robust so that decision makers can confidently use the results to decide how best to develop a vehicle.

During the jury evaluation the assessors have to make decisions on which vehicle they believe is more closely associated to the criteria used in the assessment e.g. how "powerful" each sound is; which can be an intuitive procedure. As humans we make decisions every day, and we regularly compare, contrast and select certain products over others. The decision making processes include processing physical signals from the environment surrounding the human. These are perceived and then compared to stored memories, before a decision is made and followed by an action (Wickens 1999).

Wickens (1999) described how humans make decisions. This is shown in figure 16. It is a multiple component process, which starts with a person perceiving their environment. The output of their perception is a decision resulting in a response. This response influences the environment, which is again perceived by the person, and therefore there is a continuous process.

Signals from the environment are first processed by biological receptors. In the case of sound, the ear first detects acoustic signals, which are then perceived. For this they compare the sound against those which may be stored in their memory.

The memory component is made up of two storages; the "long-term" and "working" memory. The long-term memory recalls past events, whereas the working memory focuses on the events that have just occurred. The memory and the perception together with the cognition enable the human to make a decision. Both Västfjäll and Kleiner, (2002) and Blauert (1994), researchers of sound quality, agree that decisions are based on cognition. Once a decision has been chosen, a response is executed.

Attention resources are continually influencing each of the stages of the stimuli processing, decision-making and the response selection, but these are dependent on the context of the situation as was discussed in section 2.6.

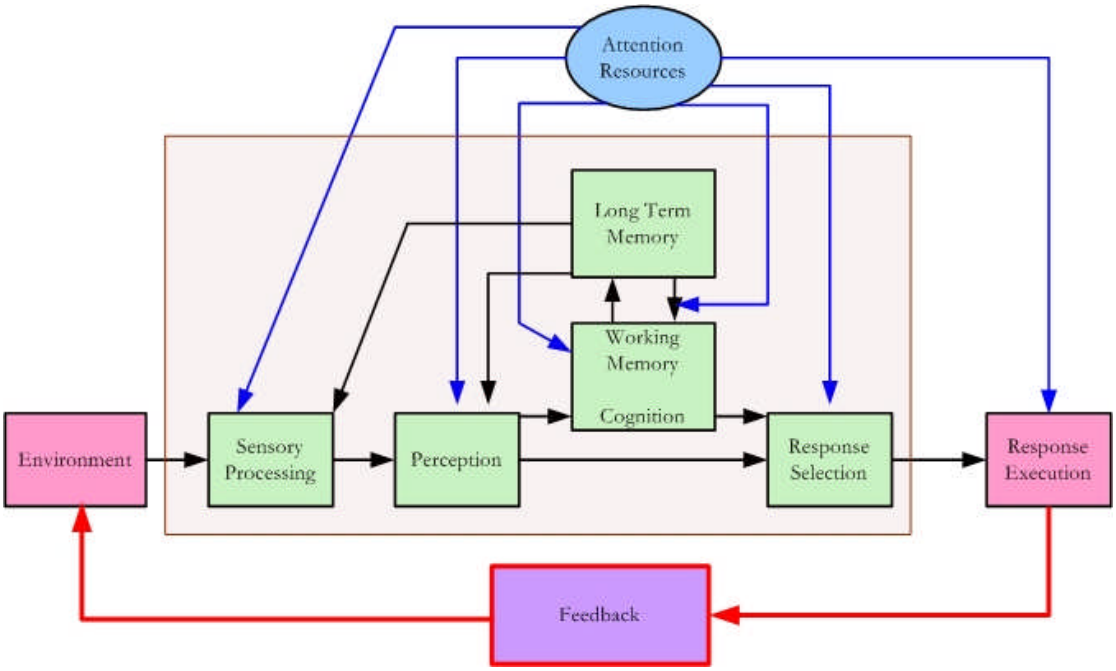

Figure 16: Decision making model proposed by Wickens (1999)

3.3. Decision making in jury evaluations

Zeitler et al (2006) proposed a decision making model related to the activities of a jury evaluation conducted in a listening room. This is shown in figure 17. It shares some of the features of the model proposed by Wickens (1999) (figure 16), shown by the similar colouring.

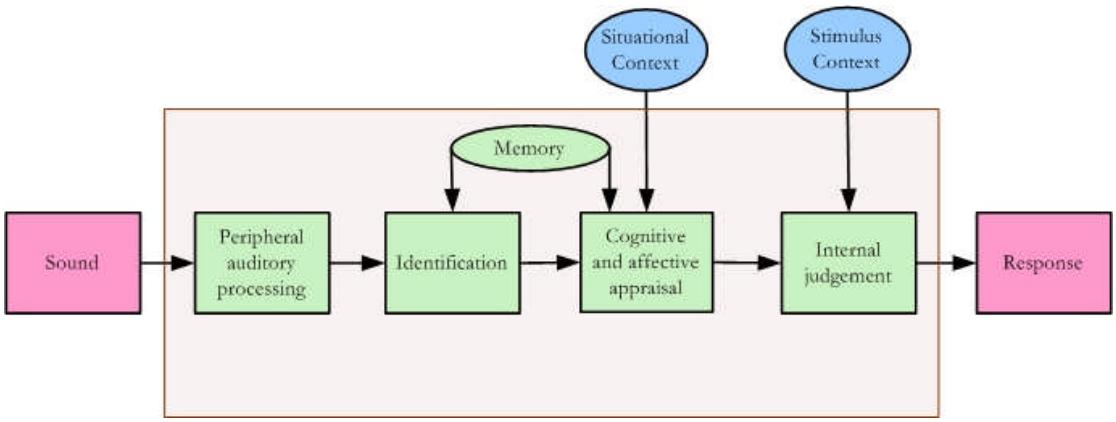

Figure 17: Decision making model proposed by Zeitler (2006)

Wickens (1999) used the generic term "environment" to describe the source of the stimuli. In listening room evaluations, the environment is referred to as "sound" as this is assessed in isolation. Therefore, given the principles of cognitive capacity hypothesis, much of the attention is focused onto one stimulus. As with the Wickens model, the stimulus is captured by receptors and then processed using perception, memory and cognition until a decision has been made on how to react. The decision in this case is which sound is, for example more "powerful", and by how much.

3.4. Research areas

Decision making within a simulator is likely to be represented by the model suggested by Wickens. The interactivity aspect of the simulator means there is a feedback loop between the simulator and the assessor, as the sound is generated in real-time in response to the assessor's behaviour. The assessors' choices are influenced by what they hear. What they hear is dependent on how they drive. Therefore to know which vehicle components are influencing subjective opinion, it would be necessary to capture how the assessors drive. Knowing this would enable the subjective impressions collected to be linked to driving conditions and therefore the form of the stimuli evaluated. The sound of a vehicle will be dependent on the driving conditions. Therefore knowing the driving conditions assessed will enable engineers to link subjective impressions to the mechanical components responsible of influencing the assessors' perception.

Central to both the models illustrated above is the assessor. How they perceive the sounds they hear can influence the outcome of the evaluation. Their attention resources are distributed according to task, the stimuli and the environment, and therefore the context of the evaluation needs to be considered.

From an evaluation perspective it was not known how to set up evaluations in the simulator, similarly it was not know how to choose between data capture methods and their corresponding data analysis techniques.

Furthermore evaluations using the interactive simulation had to take into account the NVH target setting processes, and be supported by principles of experimental psychology.

In conclusion the following five research areas needed to be better understood before a new methodology could be proposed:

1. **The assessor:** How they perceive and engage in the simulator based evaluations
2. **The context:** How to set up evaluations in the simulator, this includes:
    a. The task the assessors are asked to perform. (PC or IS)
    b. The stimuli they hear
    c. The type of simulator, desktop or full vehicle
3. **Interactivity**: What is the influence of interactivity?
4. **Data capture**
    a. From the data capture interface
    b. Of driver behaviour
5. **Analysis of**
    a. Data from the data capture interface
    b. Driver behaviour

In conjunction with the industrial partners it was decided that the following specific research questions were of a high priority. These are also illustrated in figure 18 .

1. How do assessors engage and perceive the simulator and the task?
2. Understand how to set up experiments using the simulator:
    a. How to design experiments.
    b. How to select data capture methods.
    c. How to analyse the data collected.
3. Understand interactivity:
    a. How to capture how people make decisions.
    b. How it could be used to enhance the target setting processes.
    c. What is the influence of interactivity?

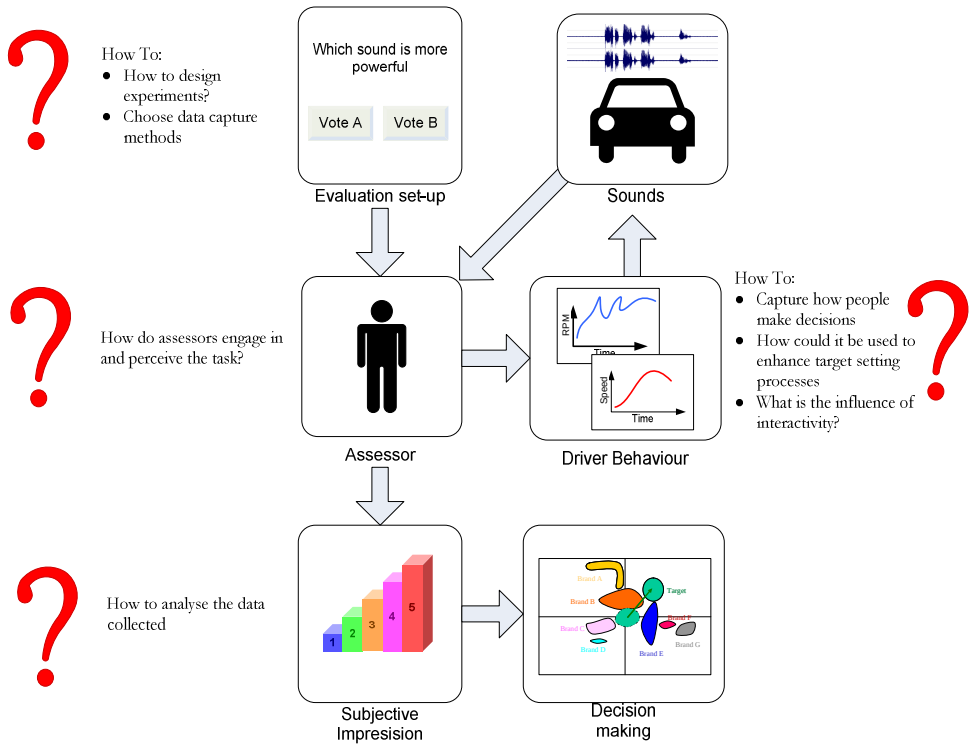

Figure 18: Research questions

This chapter has illustrated how people make decisions and the research questions that have been addressed in the project.

The next chapter discusses research resulting from question 1. Chapter 6 discusses research from question 2, whereas chapter 7 discusses the findings from question 3. The 4th research question needs to be applied to both research question 2 and 3, and therefore it is also discussed in chapters 6 and 7.

4. ASSESSORS' PERCEPTION OF SOUND QUALITY EVALUATIONS USING INTERACTIVE SIMULATION.

4.1. Introduction.

Assessors are the integral part of the sound quality evaluation. They are in essence a detection and measurement system. Integrating their needs and expectations into the evaluation procedure should result in them making confident assessments of the sounds. Hence the methodological approaches suggested must take into consideration how the assessors perceive the experience and engage in the task of evaluating sounds using interactive simulation.

The full vehicle simulator is a real vehicle positioned in front of a screen onto which a virtual environment is projected. This should not be an intimidating set-up, as humans we are regularly exposed to vehicles and computer generated graphics on a daily basis. It is, however, the context of the experiment that may be alien to the assessors. They are asked to listen to vehicles' sound and rate them whilst driving through a virtual environment. This is not an everyday occurrence for most people; it was therefore necessary to understand how they react to this new situation.

From conducting sound quality evaluations in the listening room, there was a general awareness to ensure the assessors were first comfortable in the evaluation environment; and engaged in the task of evaluating sounds. It was also important to grasp how willing they would be to return for future evaluations. Assessors are a rare commodity; having a database of assessors willing to return can reduce the time taken to set-up evaluations. Recruiting and organising assessors is one of the more time consuming aspects of sound quality evaluations. In addition, by keeping a database of assessors it is also possible to monitor how subjective opinions can change over time.

4.2. Assessor engagement.

The investigation of the perception of the evaluation experience was researched as part of this research programme and it is discussed thoroughly within submission 4. An informal and guided interview was conducted after the first ever evaluations in the simulator with non-expert assessors. This was done in the fully interactive evaluation mode and in conjunction with the Interactive Scaled interface. Overall the assessors found the experience was pleasing and enjoyable. However, when prompted for improvements or causes for concern, a number of issues were captured. These fell into

three categories: the briefing; the equipment; and the comfort and involvement. Each of these is now discussed in detail, together with a willingness to return for the evaluation.

4.2.1. Briefing:

The briefing is the medium through which instructions are communicated to the assessor. It was found that the content and delivery method of the briefing could influence the assessor's behaviour during the evaluation.

Through the research, it was established that the most efficient way to deliver the briefing to the assessors was verbally, as opposed to through an automated briefing as is done in traditional testing environments. This means that the tester has to make an effort to ensure consistency between the deliveries of the briefing, so that all assessors are exposed to the same experimental conditions. An outcome from this research project was to set-out clearly the briefing method. This was included in submission 4. It is a step-by-step guide of how the experimenter should deliver the briefing for both the full vehicle and desktop simulators

4.2.2. Equipment set-up: sensitivity adjustments

The exploratory studies, by the author, also identified difficulties the assessor had when interacting with the data capture interface whilst driving. Assessors reported going off road, or having to make corrective manoeuvres following interaction with the data capture interface.

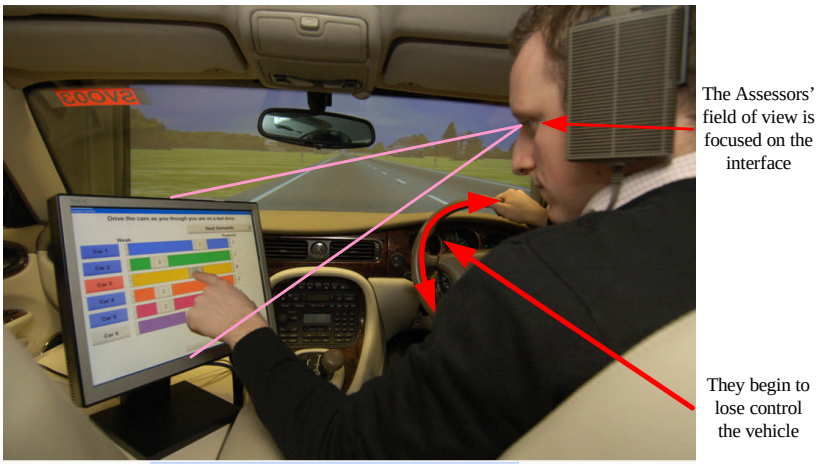

Figure 19: How assessors can be distracted

The impact of this is two fold. First, if the assessor is off road, then they are not listening to the sound in the correct context. Also if they are concentrating on their steering then they are not concentrating on their task, which is to evaluate cars. Conducting the tests on straight roads reduced the occurrence of this issue. The drawback of this is that it is potentially reducing the level of interactivity, and therefore moving the simulator experience further away from a real appraisal. However this was seen as a necessary compromise. Some provisional adjustments had to be made to the steering, but the real alternative was to upgrade the steering wheel to incorporate a force feedback system. At the time, it was not financially viable; however this is now being pursued in a collaborative research project initiated from the recommendations of this EngD project.

Usability issues were also identified. The sensitivity and responsiveness of the touch-screen, steering wheel and foot pedals were found to causing the assessors to focus upon these issues, this was possibly detracting their attention on the task of the evaluation. These issues were raised and prioritised with the developers of the simulator and contributed to their product development programme (Williams 2005).

4.2.3. Comfort and Involvement.

In the initial study, 2 out of the 10 assessors who took part reported suffering from nausea. A review of the literature revealed that the nausea could have been caused by simulator sickness. Research by Mourant and Thattacherry (2000) provides insight into this phenomenon.

Mourant's research was of interest as he also used a fixed base simulator. He found that the VE was responsible for inducing symptoms of simulator sickness, and that the gender of the participant influenced the magnitude of the symptoms. Mourant found that the symptoms were less prevalent with those participants who drove in simulated city environments than those who drove in rural ones. This was accredited to the speed of the vehicle. He also found that women were more susceptible than men to the symptoms. Min (2004) also investigated the causes of simulator sickness, the findings agreeing with those of Mourant when considering the gender of the assessor.

Lo (2001) however, introduced another variable referred to as *"scene oscillations"*. Scene oscillations" is the phrase used to describe perceived motion induced by the simulation, such as pitch, yaw and roll. Lo found that participants were more susceptible to simulator sickness when exposed to scene oscillations compared to those assessors where scene oscillations were less present.

Following this new insight a second exploratory study was conducted with a larger sample of 21 assessors. The purpose of this was to explore the impact of simulator sickness and its effects on the assessors. To investigate this, assessors were asked how comfortable their simulator experience was. Assessors were not asked directly if they suffered from simulator sickness to avoid misleading their perception of the evaluation. In addition to comfort, the assessors were asked how involved they felt in the evaluation. The purpose of using a simulator over the listening room is that the assessor should feel engaged in the evaluation, and in control.

The questionnaire responses indicate that most assessors felt involved in the task, suggesting that the experimental set-up was appropriate for evaluations.

Comfort levels were also acceptable. However as with the first exploratory study a similar percentage of assessors felt the symptoms of simulator sickness. It was also found that assessors' involvement in the evaluation was related to their comfort. Figure 20 shows the relationship between comfort and involvement.

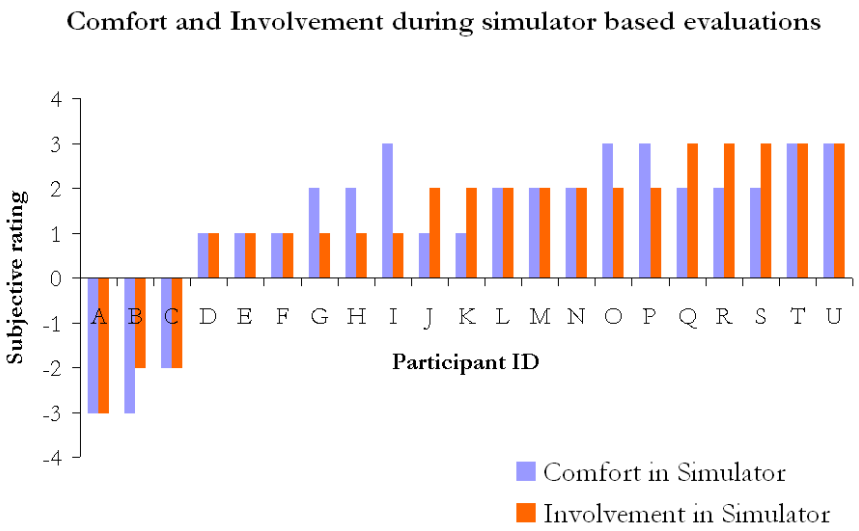

Figure 20: Comfort and involvement in simulator-based evaluations

The negative responses towards comfort were related to symptoms of simulator sickness and rising temperature within the simulator. It was found that approximately 10% of assessors suffered from simulator sickness. A recommendation was made to inform the assessors of the symptoms of simulator sickness. This precautionary warning was fed into the briefing. The assessors were asked to stop the evaluation prematurely and leave the vehicle. They were also asked not to drive a real car until they felt better.

4.2.4. Perception: Willingness to return for future evaluations

In a separate study, assessors were asked to select which evaluation environment they were more willing to visit for an evaluation if the evaluation were conducted interactively. Assessors experienced evaluations within the full vehicle simulator and at the desk, both in the fully and non-interactive mode. The results are shown in figure 21.

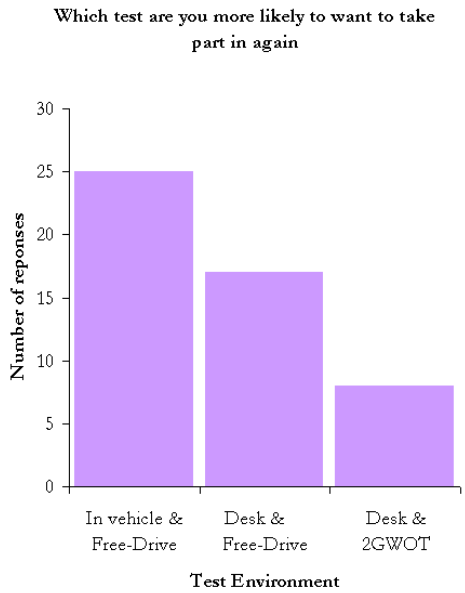

Figure 21: Preference of evaluation

Fully interactive in-vehicle evaluations were the most favoured. The second best was the fully interactive mode at a desk, and the least favourite was the non-interactive mode at a desk.

4.3. Summary

This chapter has reported on the research conducted on the first evaluations using non-expert assessors in the interactive NVH simulator. Prior to this research, only key decision makers from OEMs and engineers had experienced it. The exploratory studies reported here focused on how to engage the assessors with the interactive simulation. Minor functionality issues were identified and rectified such as steering wheel sensitivity. A standard briefing method was also prepared. This should be delivered in conjunction with a demonstration.

It was also identified that simulator sickness affects approximately 10% of the assessors recruited. This means it is necessary to notify the assessors of the symptoms of simulator sickness prior to the evaluations during the delivery of the instructions. The assessors

should be asked to end the evaluation prematurely if they feel any symptoms of nausea and sickness. As a precaution the VE was modified so that only straight roads were used. This would decrease the occurrence of scene oscillations that potentially cause the symptoms Also assessors are now asked to not drive immediately after the evaluation, until the symptoms have gone away.

5. EXPERIMENTAL PROCEDURES FOR SOUND QUALITY EVALUATIONS IN THE INTERACTIVE NVH SIMULATOR

5.1. Introduction

The research presented within this chapter focuses on understanding how to set up an evaluation. This includes the implementation of principles for the design of evaluations and the selection of data capture methods. This stage of the research encompassed increasing levels of complexity; therefore many of the details have been omitted from this Innovation Report. The reader is urged to refer to submissions 3, 4 and 6.

5.2. Experimental design

The initial methodological approaches adopted by the developers of the interactive simulation were benchmarked and used as a starting point for the guidelines proposed within this research programme. Through this approach it was evident that a rigorous application of experimental procedures had not been adopted. Stone and Sidel (2004) provide a list of effects and errors that should be avoided in setting up sensory perception evaluations. These had not been taken into account in the original methodology and could influence the results. The effects are:

1. Halo effect
2. First sample error
3. Leniency error

The *"Halo effect"* occurs if the order in which an experiment is conducted could affect its outcome. With the simulator's original methodology the assessors were always asked to first assess powerfulness followed by the assessment of refinement. This meant that if the assessor had tired during the first assessment, the second assessment would have not received an equal amount of attention.

To avoid this, the repeated measures design (Tilley 2004) has been recommended by the author of this document. This ensures that the order of the questions asked to the assessors is alternated for each person that takes part.

Similarly the order the sounds appear on the IS interface needed to be randomised each time the evaluations takes place. If the order the sounds are displayed on the screen is

fixed, then it is possible that order effects will bias the results. Again from the literature, this was referred to as the *"first sample error"*.

Following these recommendations the simulator's software has been updated. Now it has the functionality to easily randomise the order of the sounds, and alternate the order of the questions asked to the assessors. These recommendations should provide confidence in the results obtained, as the evaluation is free from experimental biases, and it is also easier to set up the evaluation as this functionality is now automated, meaning that rigorous evaluations can be set up now in minutes, as opposed to hours.

The "*Leniency"* error is as a result of non-standardised delivery of the instructions to the assessor. Recommendations for the briefing were discussed in chapter 4 and in submission 4. Included in these is a standardised briefing conducted in conjunction with a demonstration of how to use the simulator.

5.3. Data capture methods for interactive simulation

Traditionally sound quality evaluation processes have employed the techniques discussed in section 2.5. Similar methods were initially adopted for the use of the simulator. The paired comparison method and the interactive scaling were used with the non- interactive and fully interactive method respectively. The decisions to use each of these methods had been implemented in an ad-hoc manner. There was little understanding of the capabilities of the IS method and how the data collected via its use should be analysed. There was also limited understanding regarding how to choose between data capture interfaces.

5.4. Classification of the IS method

To understand the nature of the IS method, a review of methods used in sensory perception of food and in sound quality evaluations was conducted. Furthermore an understanding of measurement theory was used as well.

Through this review it was evident that the IS interface was an amalgamation of many different approaches. From the sensory perception disciplines, it was evident that the scales were defined as "category" scales with opposite bipolar adjectives. Relative to the methods used in the automotive industry, the IS method is a hybrid of three approaches. Rank order, rating scales and semantic differential. The presence of the "reshuffle" button on the IS interface, means that assessors can rank the order the sounds appear on the screen. The presence of adjectives along the scale indicates that it is a semantic differential, although the numbers displayed on the toggle bar demonstrate it is a rating

scale as well. Through a review of measurement theory and comparing this to practices of sensory perception, it has been assumed that data collected from this interface would be of the interval level. Levels of data are discussed in the next section.

The next stage in this research was to define the statistical approaches that should be used with the IS method. In particular it was necessary to understand whether parametric on non-parametric statistics should be used.

5.5. Data analysis methods

Upon a review of how the data capture interfaces were being used in conjunction with the simulator, it was evident there was a need to introduce analysis methods which would complement the decision making processes. Allman-Ward (2004) analysed results from the IS method using Exploratory Data Analysis (EDA). Whilst EDA remains essential for any form of data, they are limited in the degree of information that can be communicated. This is because there is no indication if the sounds are significantly different from each other or if the observed differences are in fact due to random effects. As part of this research programme, a framework to select the statistical analysis methods was needed. This is an outcome from this research and plans are in place to incorporate the framework into a new version of the simulator software.

Through the use of statistics it would be possible for key decisions to gain a deeper understanding of the customers' perception. There are two forms of statistics that can be applied; these are referred to as parametric and non-parametric. Parametric statistics assume the data originates from an interval level scale and conforms to a normal distribution. This means that descriptive statistics such as means and standard distributions are suitable descriptors of the data. Non-parametric, are usually referred to as distribution-free statistics, because they do not assume the data conforms to the normal distribution. In addition the assumption is made that the data originates from an ordinal-level scale.

5.5.1. Level of data

Within the field of subjective evaluations there has been an on-going debate. Sensory practitioners, psychologists and statisticians often argue about which type of statistics should be used to analyse subjective data (Bower 1995). The argument focuses on defining the "level of data" that is collected in subjective evaluations. There are four levels of data, nominal, ordinal, interval and ratio. All four levels are discussed in detail in

submission 6. Only the ordinal and interval levels of data have been discussed here, as these are the focal point of the dispute. For example, an assessor may give the products A, B, C, D the scores 10, 65, 85, 30 respectively. If the level is considered "ordinal" then the scores are ignored and the products are assigned the order 4th, 2nd, 1st and 3rd respectively. If the level is considered interval then the scaled difference between the sounds is considered. By considering it as ordinal data there are those who argue that some information is lost (Stone 2004) and the scale between the sounds is ignored, reducing the output from the rating scale to the same as that from a rank order test.

Gardener (1975), Resurreccion (1998) and Stone (2004) discuss the presence of a "grey area" in the interpretation of the level of data. Gardener (1975), Bower (1995) and Stone (2004) argue that parametric statistics are sufficiently "powerful" to withstand a violation of the level of data, and therefore it is possible to use parametric statistics. Although it is Gardener (1975) that states that the use of the type of statistics is dependent on the risk that one is willing to accept. If they are cautious, then they are likely to use non-parametric statistics. If on the other hand they are willing to accept potential violations of the level of data, then parametric statistics can be used. Such principles seem to have been put into practice, although it has not been understood if this occurred whilst considering the level of risk one is willing to accept. Villanueva (2000) and Stone (2004) who are concerned with the sensory perception of food recommend the use of parametric statistics. In addition the British standard for sensory perception (BS ISO 4121:2003) considers the use of scales, similar to those of the IS method, to lead to an interval level of data. On the other hand texts related to medical applications consider subjective data to be ordinal in nature. Jakobsson (2004), who discusses best practice in nursing, states that non-parametric statistics should be used, and Tilley (2004), a text book for experimental psychology, approaches the issue conservatively, and states that as it falls into this grey area it is safer to treat the data as ordinal.

It was concluded that determining the level of data was not within the remit of this project. An investigation into such an area would have not have not led to contributions in understanding how to utilise interactive simulation for sound quality evaluations. Therefore given this project falls within the remit of sensory perception and the similarity between the IS method layout and the graphical scales illustrated in the British standard for sensory perception, the assumption has been made that the data collected using this method is at the interval level.

The issue of normality however can be potentially of greater significance to the interpretation of the results. Knowing if the mean is an appropriate descriptor of the

results can influence the decisions made whilst target setting. For these reasons, methods to test for normality were investigated.

5.5.2. Normality

Parametric and non-parametric methods make assumptions regarding the shape of the distributions, and therefore it is necessary to identify this before selecting a method. This means that a method to determine if the data is normally distributed is also necessary. To do this, Bower (1995) stated that one should observe the data to see if it "looks" normal. This approach seems to be reliant on the skills of the observer and can lead to biased results. This therefore requires a common way for testing normality. For this Roberts and Russo (1999) recommend the use of the Kolmogorov-Sminrov test. Once the normality of the data has been confirmed then it is possible to select the statistical tests which complement management's decision making. If the data is not normally distributed then it is likely that the observer may have to check to see which descriptive statistic is more relevant, the mean or the mode.

5.5.3. Statistical Analysis Framework

Decision makers in industry wish to understand how their own existing or concept vehicles are performing against the competition. Their immediate concerns are not related to how competitor cars fare against each other. This therefore means a number of pairwise comparisons are needed between each of the competitor vehicles and their own vehicle. This concept was initially introduced by Fry (2006); he made use of the so-called "win-lose graphs" which show the difference between two vehicles and if that difference is significant.

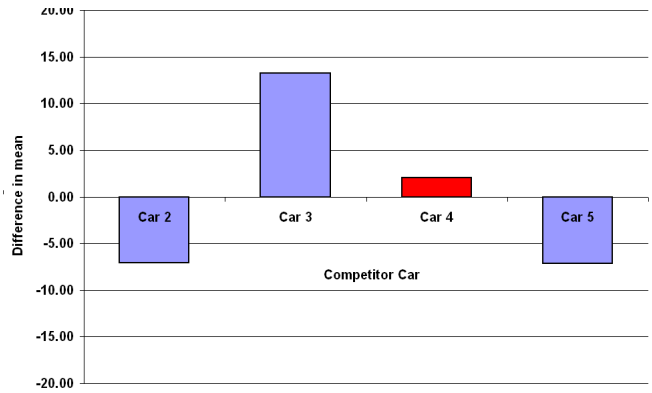

Figure 22: Win - Lose graphs

For the paired comparison method, the critical difference between sounds can be calculated with following formula (David 1988).

$$\Delta_{95} = INT\left\{1.64\left(\frac{NT}{2}\right)^{\frac{1}{2}} + \frac{1}{2}\right\}$$

Whre N is the number of sounds and T is the number of comparisons made. If there have been 10 assessors, and each assessor evaluates each pair twice, then 20 comparisons have been made. INT indicates the result has to be rounded off to the nearest integer.

This approach is favoured by J&LR, as it illustrates the scale of the difference between vehicles and if that difference is significant. However equivalent statistical approaches were needed for the IS method. These were identified through this research programme, such led to the following statistical analysis framework.

The first stage in the data analysis is therefore to use the test for normality as suggested above. This will indicate which stream of statistics should be use. If the data collected conforms to a normal distribution, then parametric statistics should be used. For this, the author of this document identified that analysis of variance (ANOVA) method can be used. This approach is recommended as it is suitable for analysis of 3 or more variables. In the case for automotive sound quality evaluations usually 6 sounds are assessed in evaluations for the Jaguar brand. In addition the use of the ANOVA method has been observed in other sensory perception practices such (Piggot 1984) (Shen 1997) Kuti(2004) Villanueva(2000) Lheritier (2006) Stone(2004) and Resurrecion (1998). The non parametric equivalent of the ANOVA is the Friedman test.

The ANOVA method works by identifying if the sounds used in the evaluation, have led to significant differences in the subjective impressions formed by the assessors. The limitation however is that it cannot indicate which sounds are causing the significant difference, and therefore further analysis is required. If the ANOVA yields a significant result, then further comparisons need to take place. There are two approaches for this. The Tukey HSD test calculates a critical difference, (between the means of the scores for each sound) that must be exceeded for the comparison to be considered significant. This value applies to all of the possible pairs of sounds. This approach can however lead to high probability of the type I errors (this means the findings could be wrong). To prevent this from happening, the paired t-test could be used to compare if two vehicles are

significantly different from each other, as long as the number of comparisons made between sounds is limited.

The literature is however vague in defining the maximum number of comparisons that can be conducted. Instead the use of Bonferroni adjustments is suggested. This makes the paired t-test more stringent. For example if significance is requested at the 0.05 level but there are 5 comparisons needed, then 0.05 is divided by the number of comparisons. Hence each comparison must be significant at the 0.01 level.

The non- parametric equivalent of the paired t-test is the Wilcoxon test. This ignores the rating associated with each sound, and only considers its ranking.

However parametric statistics tend to be preferred over the non-parametric equivalents as these are said to be more "powerful". By this it is meant they are more likely to correctly identify if a result is significantly different or not, that is the power to define its true significance (Bower 1995). For this reason if the data does not first conform to a normal distribution then the option to transform the data using z scores has been introduced. These make adjustments to the scores assigned by each assessor, ensuring that each set of scores by each assessor has the same mean and standard deviation. This could potentially transform the data so that it does conform to normal distribution. Following this, the data is checked for normality again, if it fails then non-parametric statistics should be employed, if it succeeds then the paired t-test is used on the transformed data. If the ANOVA does not highlight significant difference then it is possible that the sounds evaluated were not significantly different. Given this, the following process for data analysis has been proposed by the author (see figure 23).

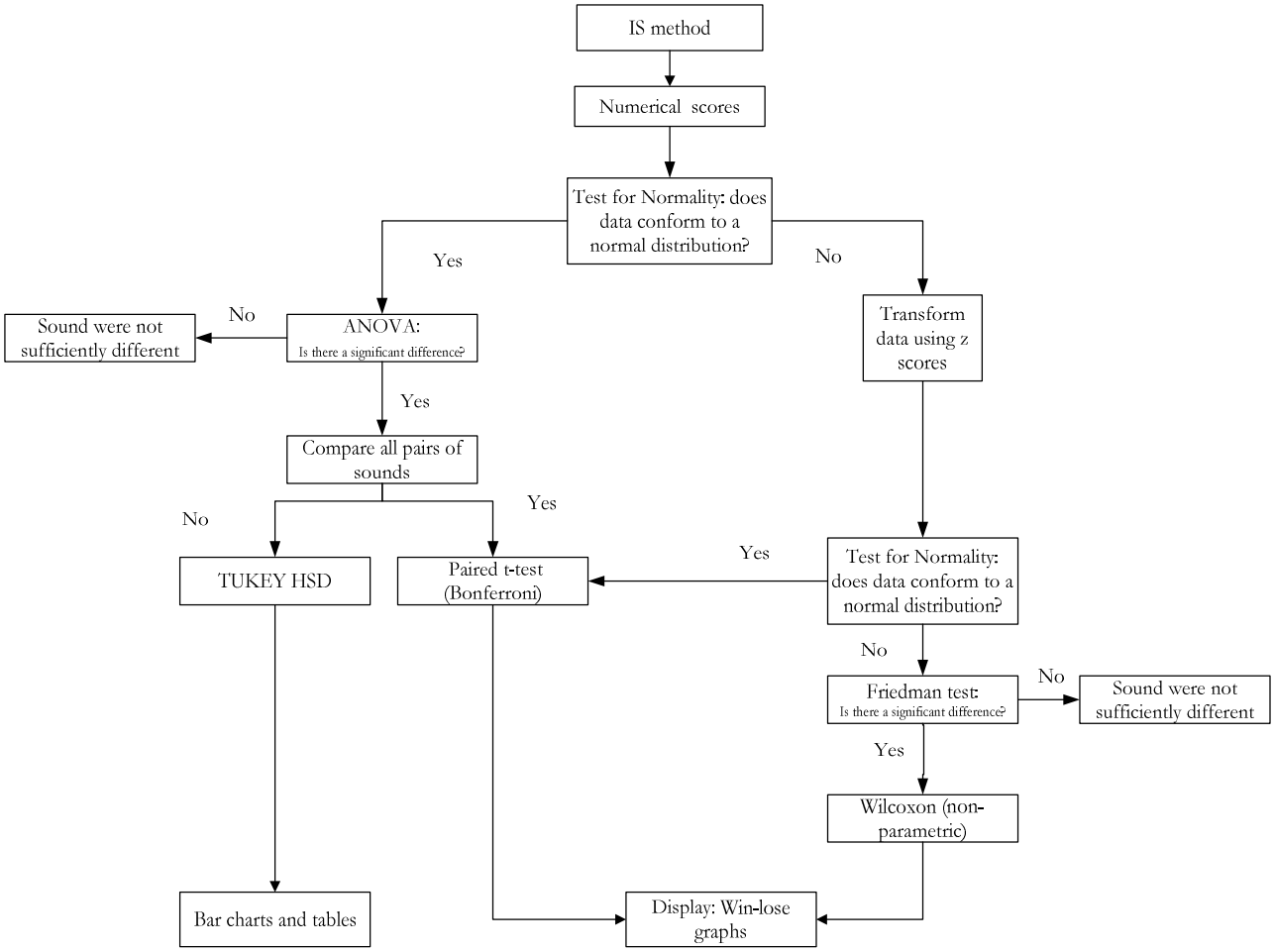

Figure 23: Data analysis methods for IS method

Detailed workings of the ANOVA method and the TUKEY HSD method were discussed in detail in the post module assignment for the applied statistical analysis module. In addition methodology was checked by the head of psychology at the University of Warwick and statisticians within the engineering department. This therefore led to the integration of these methods into updated versions of the simulator software.

Once the IS method and its corresponding analysis techniques were identified, it was necessary to understand if it was suitable for evaluations using interactive simulation, and when it and / or the paired comparison method should be used. To conduct this investigation, it was necessary to define assessment criteria. This is discussed in the next section.

5.6. Criteria for data capture methods

The criteria were formulated by considering the demands on the assessors, and the requirements of the NVH target setting process. In addition, ways of distinguishing between data capture methods had been shown by Bodden (1998) and Maunder (1998). Measurement theory concepts were also observed, these were initially considered by Rossi (2003). The criteria formulated can be segmented into two categories. The first is referred to as qualifiers, the second is differentiators. Qualifiers are those criteria that the function of the data capture interface must satisfy before it can be considered for use within the evaluation. Differentiators are those criteria which help distinguish which data capture interface is more suitable.

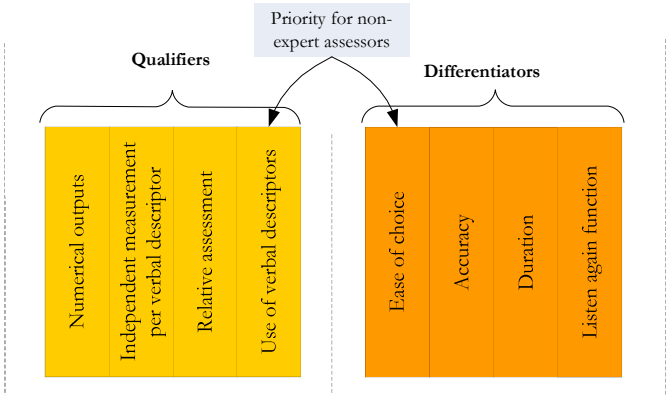

Table 2: First assessment of data capture interfaces using the criteria

Both the qualifiers and the differentiators are now discussed. Much of the justification for the criteria is discussed in detail within submission 6. The reader is urged to consult that document for the origins and rationale behind each of the criteria.

5.7. Qualifiers

5.7.1. Numerical outputs

Numerical outputs, as opposed to verbal data collected from interviews, are needed to provide a sense of measurement. Using numbers can provide a point of reference for improvement and it complements the manner in which decisions are made within a vehicle development programme. They can also be used to correlate subjective impressions formed with objective measures.

5.7.2. Independent measurement per verbal descriptor

The subjective impressions of powerfulness cannot be derived from the assessment of refinement and vice versa, and therefore an independent evaluation is needed for each attribute under consideration.

5.7.3. Relative assessment

Manufacturers are interested in how their existing and prototype vehicles compare against the competition. They are not interested in a comparison against all cars in existence. This is why the assessment has to be a relative and not absolute.

5.7.4. Verbal descriptors

Otto claimed that non-expert assessors find it easier to assess sounds with the use of adjectives. It was found that non-expert assessors may find the use of numbers more difficult, as they have no point of reference. Therefore this criterion has been included for evaluations using non-expert assessors such as customers. It is not as important with expert assessors such as NVH engineers, as these people have been trained to use a numerical scale.

5.8. Differentiators

Differentiators are those criteria that enable the person designing the evaluation to understand which data capture interface is more suitable for their assessment needs.
There are also 5 differentiators, these are:

5.8.1. interactivity: "Listen Again" Function

Enabling the assessors to listen to the vehicle sounds as many times as they wish, before making a decision, can lead to increased levels in performance. This was accredited to the fact that assessors can get familiar with the sounds before they make a final decision. Some interfaces will enable the assessor to listen to the sounds again more than others.

5.8.2. Accuracy

This refers to the ability to statistically distinguish between sounds using a data capture method and its corresponding analysis technique. NVH engineers need to know, with degree of confidence if a sound is significantly different from another, or if the apparent differences are due to random variability in the data. Some approaches may be more accurate than others.

5.8.3. Duration

The assessors' level of attention paid during an evaluation may be related to the duration of the experiment. The longer the evaluation is, the less attention is paid to the task. The duration of the evaluation also affects the number of participants that can be used in the time-frame available. Therefore data capture methods that take less time should be seen as more favourable.

5.8.4. Ease of choice (from the assessors' perspective)

It is best to make the task as easy as possible for the assessors. For example, it is easier for a non-expert assessor to choose their preferred sound, as opposed to choosing their preferred sound and then deciding by how much.

5.9. Using the criteria

Using the criteria it was possible to evaluate the suitability of the methods available using interactive simulation. These were the paired comparison in the free-play and fixed play modes and the interactive scaling method. The initial review identified that there was insufficient information to decide if the IS method led to a relative assessment of the vehicles' sounds. It was also evident that the paired comparison fixed play did not satisfy the "listen again" function whereas the IS method allow the assessors to listen as many times as they wish before completing the test. The paired comparison free-play is "better" than the equivalent fixed-play mode, however the assessor is not as free to "listen again" as they can with the IS method. This is because the assessor can listen to each pair of sounds as many times as they wish before making a choice between the two. Once this choice has been made the assessor can not go back to listening to the previous pairs of sounds. In addition it was not understood which method was more accurate and

which one took less time to complete. As for the difficulty of choice, the paired comparison approaches were considered easier than the IS method.

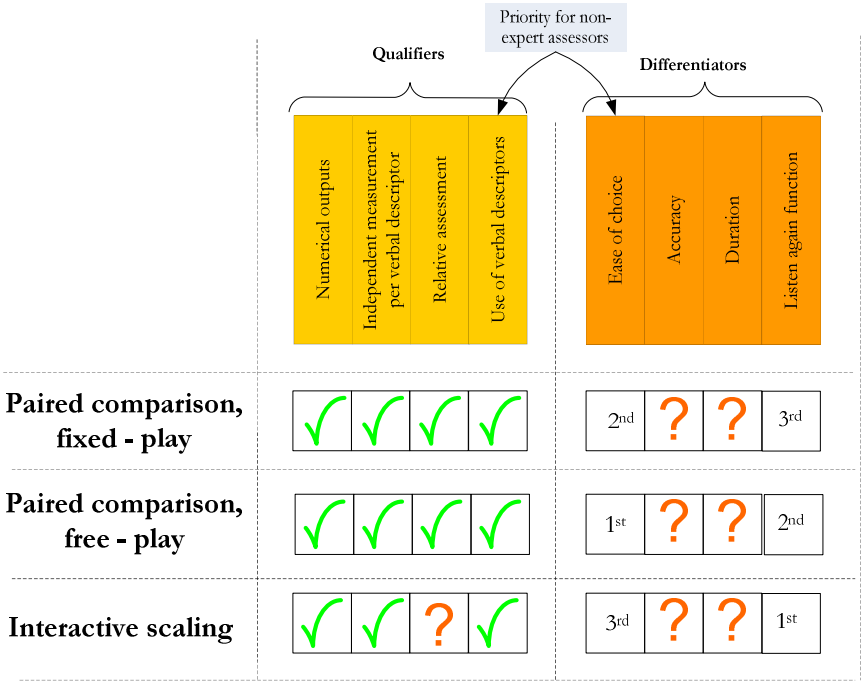

Table 3: First assessment of data capture interfaces using the criteria

To evaluate whether the IS method led to a relative assessment of the vehicle sounds, it was necessary to observe how it was being used by the assessors. For this two observations were made. The first was an observation of the use of the reshuffle button; the second was the range of scores given to a set of cars. The use of the reshuffle button would indicate if the assessors were trying to evaluate the cars against each other, the range would demonstrate the outcome of how they were using the scales. As most of the vehicles used in the evaluation are from competing brands, with similar engines, it would be expected that the impressions of these vehicles to be similar, and therefore the range would be very narrow. However if the assessment is a relative one, then it would be likely that the range would have been broad. Figure 24 illustrates this. Both interfaces show that the same order has been associated with the vehicles, but the scale used is different.

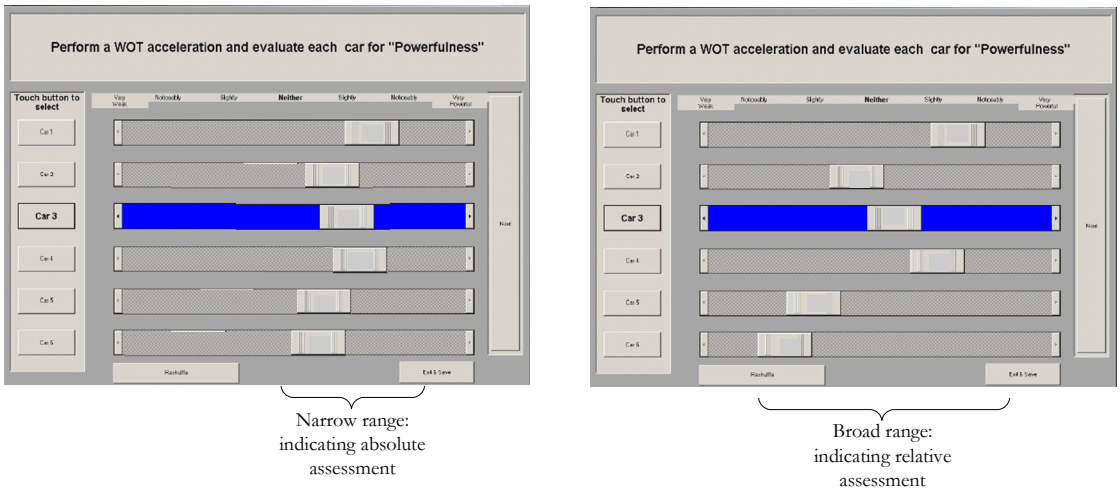

Figure 24: Absolute and relative assessments

5.10. Methodology

To capture the assessors use of the IS data capture method, 21 assessors were recruited for a fully interactive evaluation in the full vehicle simulator. They were asked to evaluate 6 vehicles according to how powerful and how refined they perceive each one to be using the IS method. The duration of the evaluation and use of the interface was monitored using data from the driver observation module (DOM), this will be discussed in the next chapter. In addition the scores given to each car were observed and the range was calculated for each assessor.

5.11. Results

5.11.1. Number of reshuffles

The observational study concluded that approximately 76% and 71% of assessors used the reshuffle button for the assessment of powerfulness and refinement respectively. The average range used was 46 for the assessment of powerfulness and refinement. These findings led to the assumption that the use of the IS method results in a relative assessment of the vehicles. This is shown in (diagrammatically) in the following figures 25 and 26.

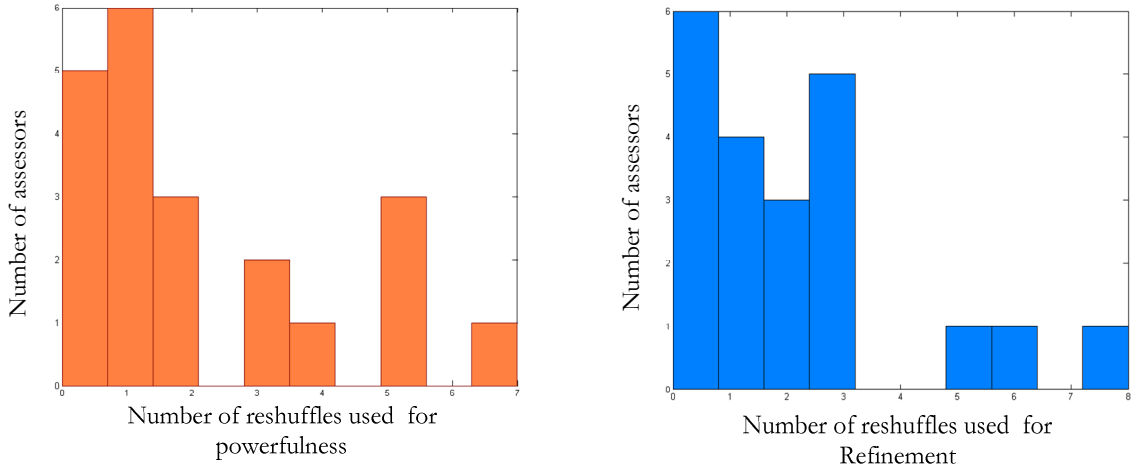

Figure 25: Number of times the reshuffle button is pressed

5.11.2. Range of scores

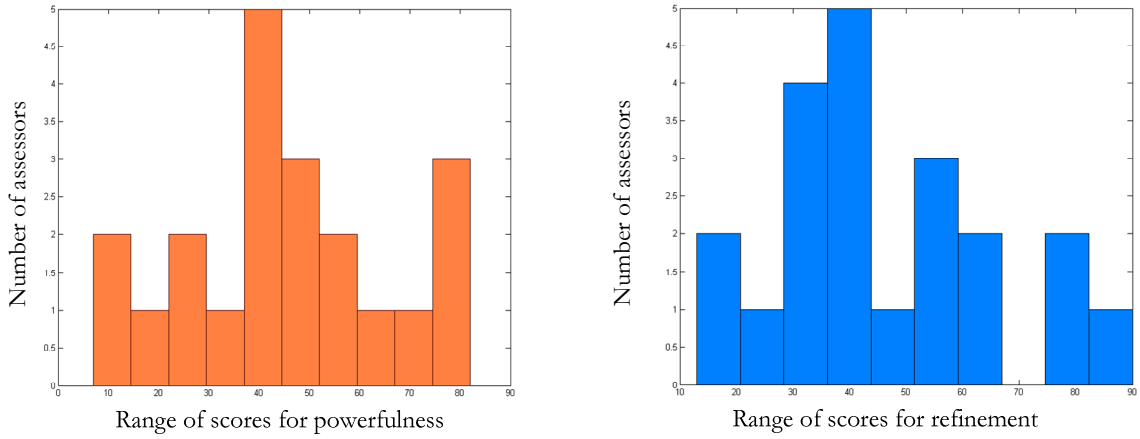

Figure 26: Scale range used

5.11.3. Duration

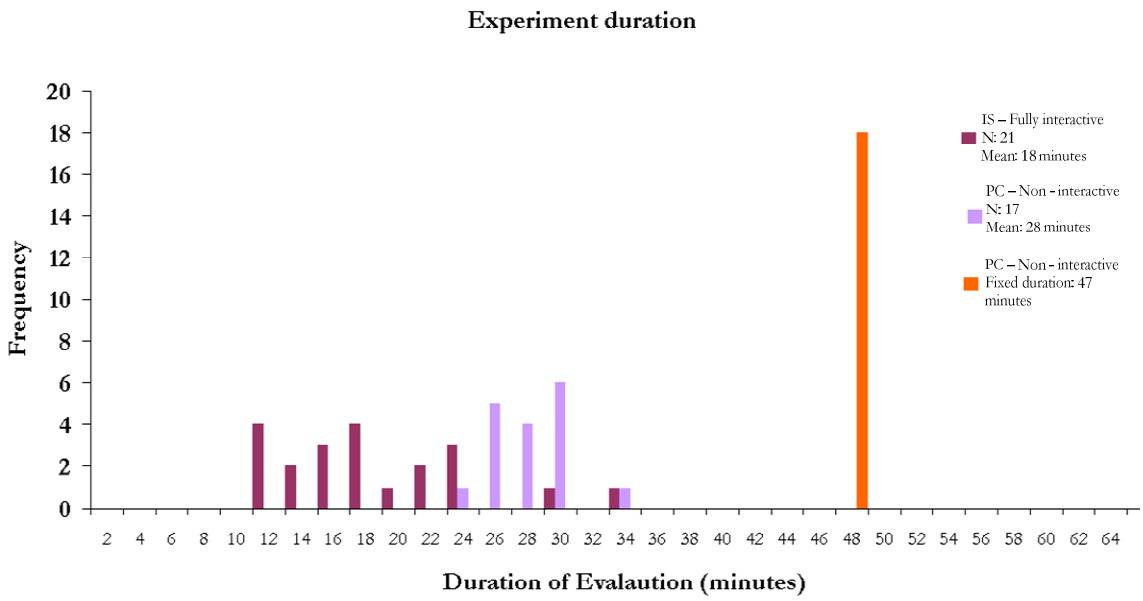

Figure 27: Experiment duration

The time taken to complete an evaluation was also recorded. The average time to complete the fully interactive assessment with the IS method was 18 minutes, whereas for the non-interactive level in conjunction with the paired comparison free play and fixed play modes, the duration was 28 and 46 minutes respectively. Furthermore, as will be shown in the next chapter, the duration of the stimuli in fully interactive evaluations is longer. This means that if the fully interactive level is used with either of the paired comparison modes, the duration of the evaluation would be even greater.

5.11.4. Accuracy

A separate experiment was conducted to compare the level of accuracy between the PC and IS method. The reader again is urged to read submission 6 for further details. The experiment focused on the evaluation of 6 sounds. The verbal descriptors "powerfulness" and "refinement" were used. The study concluded that the number of times paired comparison detected a significant difference between sounds and IS method did not was 2 for powerfulness and 4 for refinement. And the number of times IS method identified a difference and paired comparison did not was 1 for powerfulness and none for refinement. Further results can be found in appendix A. This gives an

indication that for the paired comparison can give higher levels of accuracy, as opposed to the IS method.

This stage of the research concluded that the paired comparison in free-play mode and the interactive scaling method both have an equal amount of advantages and disadvantages. Therefore the decision of which one to use becomes dependent on the priorities of the evaluation. If the sound quality practitioner deems the accuracy of the evaluation to be important, then the paired comparison method should be used. If on the other hand, time is limited then the IS method is more appropriate.

This philosophy is synonymous with NVH target setting processes. In the benchmarking exercises many vehicles may be compared, however the level of accuracy of the results is lower than that required for validation exercises, when fewer cars are evaluated and the results form the foundations for key engineering decisions. Hence for the benchmarking evaluations, the IS method should be used, whereas for the validation stages the paired comparison would be more appropriate.

5.12. Summary of Chapter

The purpose of this chapter was to demonstrate research into how evaluations in the simulator should be set up. The research led to the introduction of experimental design principles, statistical analysis techniques, and the identification of the most suitable data capture and analysis methods. For this to be possible, it was necessary to develop criteria which would help identify suitable data capture interfaces. This led to the approach of selecting data capture interfaces depending on the purpose of the evaluation.

The development of criteria for a data capture method was generated with the requirements of the target setting processes in mind. The criteria were divided into four qualifiers which any data capture method must adhere to before it can be considered for use. In addition four differentiators are also needed. These helped distinguish between the methods which satisfy all the qualifiers.

Three data capture interfaces were then evaluated using the criteria. The paired comparison in free-play and fixed-play mode and the Interactive Scaling method. The latter method formed part of the original methodology employed within the simulator. Using the criteria it became evident that there was a lack of understanding regarding the appropriateness of the IS method.

Through experimental work it was proven that the scales provided a relative assessment of the sounds, and therefore all qualifiers were satisfied.

Following this it was necessary to see how the IS method compared against the paired comparison method using the differentiators. Unknown aspects of the IS method were its capability to allow the statistical distinction between vehicles, and the duration of an evaluation.

To understand the accuracy of the IS method, a framework for selecting analysis techniques was proposed. Then through experimentation it was found that the paired comparison method was more adept for the distinction between sounds. However an evaluation with the IS method lasted less time. It was concluded that for benchmarking evaluations the IS method would be more appropriate, whereas for the validation of concept vehicles, the paired comparison method is more accurate and therefore suitable for key engineering decisions.

6. INFLUENCE AND EXPLOITATION OF INTERACTIVITY IN SOUND QUALITY EVALUATIONS

6.1. Interactivity

The term "interactive" in the context of computer or electronic systems, such as the interactive NVH simulator, encompasses the two way flow of information between a user and the system. This can lead to an input from the user resulting in a response in the system. The resulting responses in the system again influence the user, and so the process is an on-going cycle. Most situations a human is presented with in everyday life are interactive. As we make decisions based on the environment that surrounds us, and we influence our environment through our actions. Hence having an interactive evaluation could mean a more intuitive experience for the assessor. This can therefore provide a deeper understanding of how the assessors evaluate the sound of a vehicle.

There are many forms of interactivity during the evaluation in the simulator. For example, the assessor continually assesses thier location within the virtual environment and adjusts the steering wheel accordingly. They can drive the vehicles however they wish and they can interact with the data capture interface.

The two main forms of interactivity from the evaluation perspective are the manner in which the assessors drive and how they use the data capture interface. Understanding these two processes provided insight into the decision making processes of the assessors and provided the necessary knowledge into how to most effectively use "interactivity".

Previous studies concerning interactivity with a data capture interface were conducted by Baker and Jennings (2004). Conventional paired comparison evaluations are referred to as fixed-play. This means the assessor can only listen to each sound in the pair once, before making a decision. Baker and Jennings9 (2004) studied the concept of free-play. With this mode, the assessor was allowed to listen to each sound as many times as they wished before making a choice. Therefore they were allowed to further interact with the data capture interface.

Despite the fact that the assessors were allowed to listen to each sound numerous times the duration of the evaluation was reduced. This was because repeating each pair of sounds was no longer necessary. The drawback however, is that there is no fixed time for the duration of the evaluation. This makes planning an experiment slightly more complex. A time buffer is required between the end of one evaluation and the scheduled start of the next one. In addition, Baker measured differences between levels of

consistency, and noticed that when using the free-play mode, the assessors achieved higher levels of consistency. Furthermore, this mode of interactivity can lead to the assessor being more confident as they can assess each vehicle as many times as they wish before making a decision. Hence interactivity in this context can be a benefit to the evaluation process.

Dunne et al (2006) presented the perspective from the manufacturers' point of view. In a presentation at an EIS workshop he stated that interactivity was the most desirable feature of the simulator, as it allowed for all driving conditions to be evaluated.

To understand the influence and potential ways of exploiting interactivity other fields of research were consulted in particular the ability to interact with the field of view. Cain (2005) discussed different representations in customer appraisal of hand held products. Koppius et al (2004) discussed that deficient product representations could affect the price a customer was willing to pay for a product. Sales of flowers at auctions were investigated. Flowers displayed on a monitor fetched a lower price than those actually displayed in front of an audience. Krapichler et al(1998) discussed the benefits of using virtual reality in training medical students. The ability to view internal organs at any angle was seen as a benefit. This broad body of research showed that interactivity can be a benefit to the evaluation of products and imagery. The scope of the literature used however had to be broad because there was no previous work that is specific to the interaction between assessor and sound. A topic that needed to be explored was how vehicles were represented in sound quality evaluations; this is discussed in the next chapter.

6.2. Stimuli.

The stimuli used in the evaluation are important, as the assessors make their decisions according to what they hear; and what they hear is related to how they drive.

The stimuli used in conventional jury evaluations, in the pre-prototype stages of the vehicle development program are fixed driving conditions. For the Jaguar brand 2GWOTs are used. Other manufacturers however will use different driving conditions as there is no particular standard. For example Hutchins (1992) at Lotus engineering Ltd reported using 3GWOTs. However, once the first prototype is built on-road evaluations can be conducted. Here the driving conditions assessed vary slightly. Full load conditions such as 2GWOTs would be assessed, but other aspects of the car will be also considered. For example Amman (1999) from Ford research laboratories discusses the assessment of

"performance feel" by allowing the assessors to drive their vehicles however they see fit. Zeller (2005) from BMW stated that passenger comfort is dependent on the *"driving noise at constant speed"* whereas for the *"dynamic situations"* there is a need to hear more engine noise. For their evaluations they assessed vehicle at a constant speed of 100 km/h, in top gear, and then between 100 km/h to 140 km/h at full throttle.

Otto (1999), who as mentioned earlier, provided the guidelines for listening room evaluations, stated that driving conditions used in the evaluations must reflect how a customer would drive a real car. However he also stated that the evaluation should be conducted by testing the product to an extreme such as 2GWOT, as it is easier to pinpoint the differences between the sounds at these operating conditions. Hutchins (1992) offers an alternative argument to this and states that drivers would not use full load conditions, to avoid the harshness aspects of the sound. Afeneh (2007) offers a managerial perspective to the NVH target setting process. He stated that many driving conditions need to be assessed, and it is necessary to identify the driving conditions most used by the customers, and these driving conditions should then be used in the jury evaluations.

The use of an interactive NVH simulator in the pre-prototype stages of the vehicle development programme provides the opportunity for the assessors to choose which driving conditions to consider in their assessment. For example, it is believed that the refinement attribute may be related to the reduction of internal noise whereas powerfulness is a dynamic character. The interpretation of the attribute however, was not previous allowed by the assessor, as they had no freedom to choose which driving conditions they assessed during the listening room evaluations. Within the simulator, the assessor can choose the driving conditions they believe represent the attribute. On the downside, an extra source of variability is introduced. Each assessor will drive differently and this in turn will influence their subjective impressions. This occurs as the assessors make decisions based on the sounds they hear, and the sounds they hear are based on how they drive. The problem with this is that the sound heard at different driving conditions is dependent on different components. For example, whilst accelerating at full throttle, the predominant sound heard originates from the engine, whereas at cruising conditions, the sound heard is aerodynamic and tyre noise, depending on the speed.

The engineering decisions are based on the results from subjective opinions. Using fixed driving conditions such as 2GWOTs means the engineers have a good understanding of the stimuli's composition. With fully interactive evaluations the engineer could not know how the assessors drove. Therefore they could not link the subjective impression with

the stimuli. Hence they could not make informed decisions with regards to creating a target sound.

It is therefore evident that interactivity could be beneficial to the evaluation process, but it was not known how it influenced the outcome from the evaluations, and how it can be best exploited. For either of these cases it was necessary to observe driver behaviour during their evaluation. This provided a greater degree of insight into assessors' decision making strategies, which then revealed how to best use the interactive aspect of the simulator. However before this could occur it was necessary to develop the simulator so that it made a record of the assessors' behaviour during the evaluations and corresponding data analysis techniques for this type of data. This stage of the research programme is discussed in the next section.

6.3. Capturing Driver Behaviour

A first attempt to understand how assessors make decisions was made during the exploratory studies discussed above. For this Coolican (1999) was consulted to understand how to conduct observational studies. The most appropriate method for this scenario was event based observation. For this method it was necessary to watch the juror behaviour from the simulator's control desktop and make notes when changes in state occur. A state was referred to as the one of the following:

- Starting the vehicle
- Selecting a new car
- Change in speed: accelerating or decelerating or cruising.
- Rating cars
- Pressing the reshuffle button
- Revving the car whilst idle.

The implementation of the method provided some insight into how the assessors made decisions. A typical observational log for one evaluation is shown in figure 28.

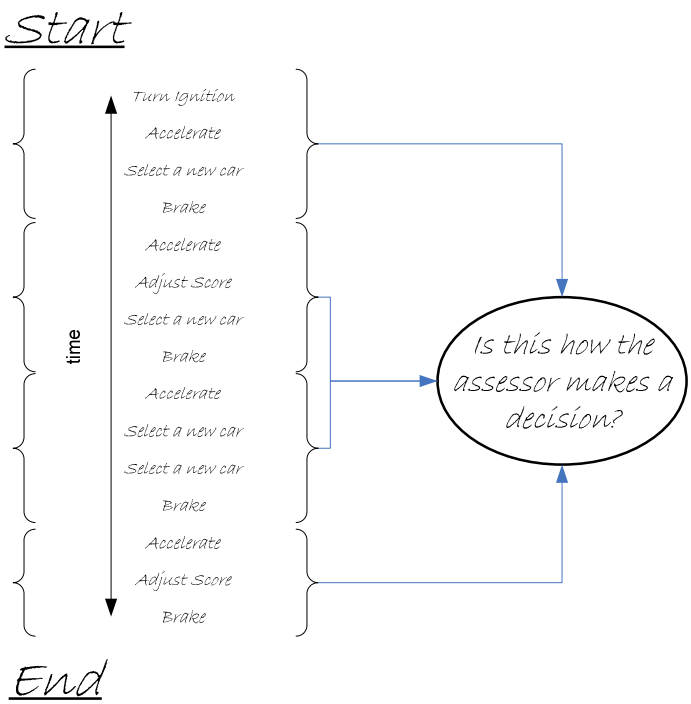

Figure 28: Typical observation log from one evaluation

The method demonstrated how the assessors were making decisions. It appeared that the assessors performed a driving manoeuvre, and then interacted with the data capture interface, either by selecting or rating a car. They then performed the manoeuvre again until they were satisfied with the scores given to each of the cars. With this knowledge it is possible to hypothesise how the actual decision making model may look for evaluations in the simulator. This is illustrated in figure 29.

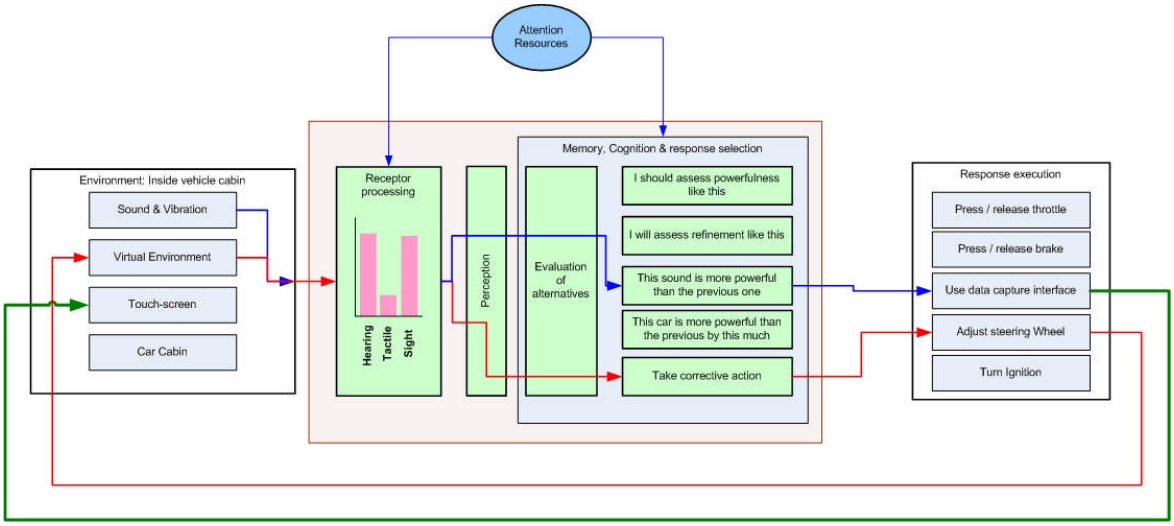

Figure 29: Theoretical decision making model of simulator based evaluations (suggested by the author)

The model above makes the distinction between the interactivity feedbacks. Showing that at any one time a number of interactions can occur. Also shown are how the levels of attention change for each of the main stimuli. The attention resources are influenced by the actions of the assessors at any one time. For example, given the principles of the cognitive capacity hypothesis (Genell 2007), whilst the assessors are listening to the sound their attention resources are more focused on the information received through their hearing, and also they may be focusing on keeping the car on the road. When this changes to the interaction with the data capture interface, then the attention resources are more focused on information being received through the sight and touch. In the example above the assessor has just evaluated the sound of a vehicle and compared it to that of one which is in his memory. This means that most attention is paid to the auditory and visual stimuli. The assessor receives new information from the data capture interface upon interaction with it. This means more attention is paid to the visual and tactile stimuli. The process continues until the evaluation is complete.

6.3.1. Driver Observation Module

The "manual" method discussed above, although informative, was not sufficiently robust to clearly identify the actions of the assessor. It relied on the observer, and this meant limited degree of detail could be collected. For example, it was not possible to observe if the "acceleration" was at full or part load, similarly it was impossible to accurately determine what speeds the assessors were accelerating from and to. The lack of robustness in the method would eventually lead to misinterpretations and biases in future observational studies. It was therefore recommended to the developers of the simulator to create an additional software module based on a specification generated by this research (see submission 5).

The simulated vehicles represented by the simulator are governed by simulated physical parameters which are defined within its performance model. This provides the performance characteristic for the vehicles used in the evaluation. If for example the assessor presses the accelerator, the performance model calculates the vehicles response in road speed and engine speed. The response is experienced by the assessor through the perceived motion, the audio heard and vibrations felt. The use of the data capture interface is also communicated to the performance model, which calculates the responses, i.e. which car to play.

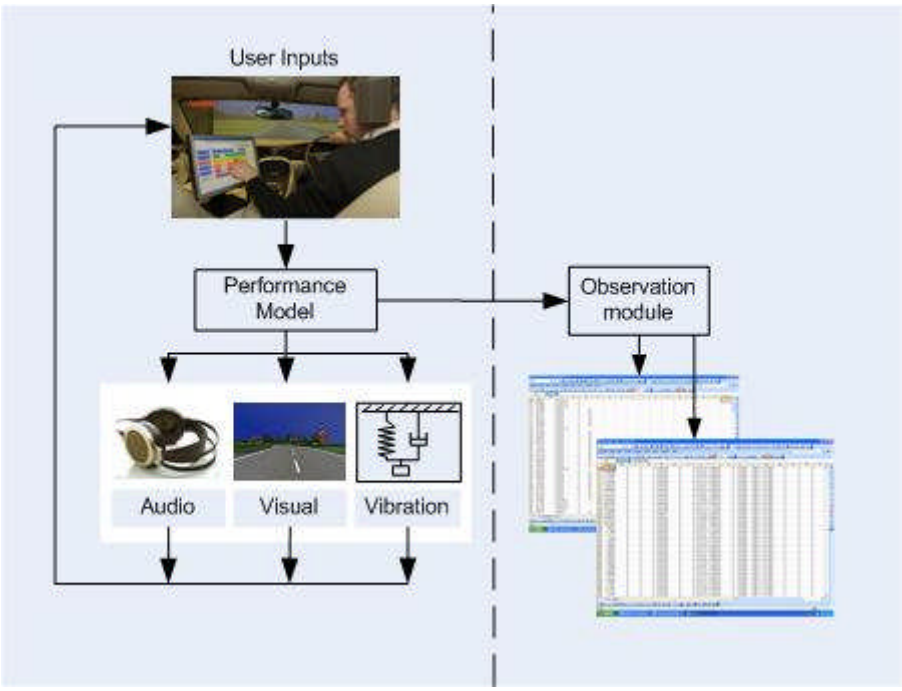

Figure 30: Driver Observation Module

The driver observation module (DOM) works by extracting data from the simulator's performance model. DOM works by keeping a time stamped record of all the information that enters and leaves the performance model. The recorded outputs are in the form of two spreadsheets. One of the sheets has a recording of all the interactions between the assessor and the touch-screen interface. The other sheet has a recording of the state of the vehicle. This includes information related to 20 parameters, including, speed, RPM, steering, and throttle position.

The benefits offered by DOM are that observational studies are now automated and can lead to more accurate observations being made. It also allows for greater detail on driver behaviour to be collected. However to analyse the observational new data, custom approaches had to be developed.

6.3.2. Visualisation Methods

Of the existing simulators in the market place, there was limited information demonstrating that they are being used to collect observational data. Schulte-Fortkamp (2007) uses the simulator alongside an interview. Here the assessors are asked to talk about their perceptions of the sound as they drive the simulator. A part of the analysis technique was to superimpose the recording on the interview onto the recording of the

sound of the vehicle during the evaluation. The advantages and limitation of interview techniques were discussed in section 2.5.6.

Other fields of research ware explored to understand how data from driving observation studies could be analysed. Most of these studies were based on driver safety, examples of these studies can be found by Hancock (2003) and Dorn (1992). Their analysis methods were based on the use of exploratory statistics such as observations of the mean speed and reaction times. Such methods, although they can provide an overview of the state of the vehicle over the duration of the evaluation, did not give insight into the assessment making strategies employed by the assessor. This prompted an investigation into the most suitable data interpretation techniques. This phase of the research was discussed in detail in submission 5.

There were two stages to developing the data analysis techniques. The first was the creation of a number of displays that provided an insight into the nature of the data collected. The second concerned the use of this data to understand how assessors made decisions.

For the first stage, interfaces were developed which illustrated how a recorded parameter changed over time. Included in these displays were the assessors' interactions with data capture interfaces and the cars selected during the course of the evaluation

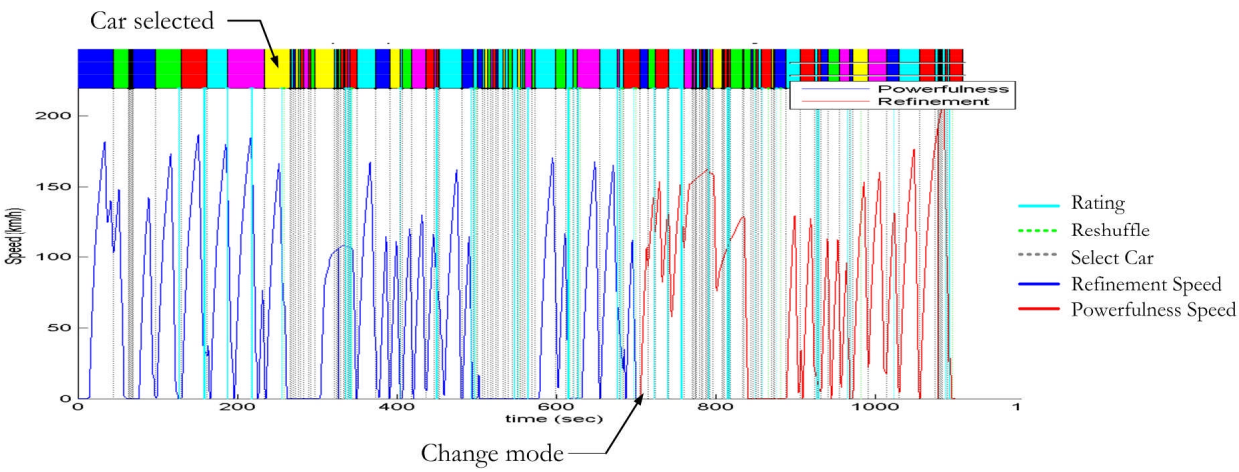

Figure 31: Speed time profile with interactions. Powerfulness and refinement.

Figure 31, shown above illustrates the speed time profile for one assessor. Other parameters that can be plotted are:

- Acceleration
- Engine speed
- Steer angle
- % brake pedal compression
- % throttle pedal compression

These forms of interfaces are useful in providing detail regarding how the assessors behaved during the evaluation. From an engineering perspective however, these interfaces contain too much detail. The NVH engineer is interested in approximately how the assessors evaluate sound, so they can understand the relationship between the sounds and the scores. For this a series of displays were created which were more effective at communicating the behaviour of the assessor during the evaluation. Each of these is now discussed.

6.3.2.1. Engine speed vs. Road speed.

The RPM versus speed shows which gears and engine speeds are being used the most. An example of this is shown below:

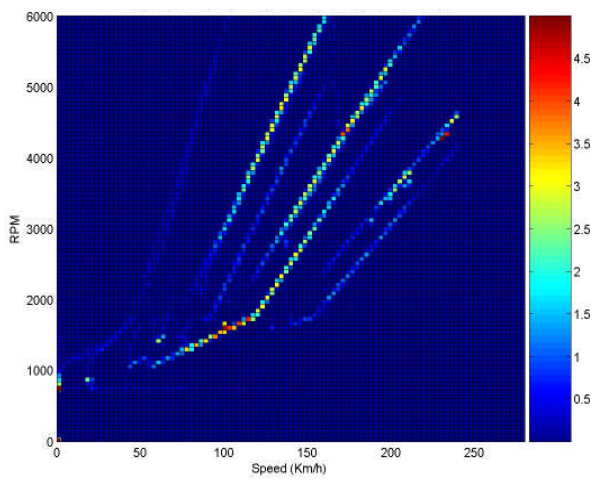

Figure 32: Engine speed vs. road speed

The amount of time spent at each RPM/Speed increment is illustrated by the colour. Bright colours relate to high usage. Figure 32 above shows how one assessor used the cars. The display method can be used to demonstrate how either a group of assessors or an individual evaluated the cars.

6.3.2.2. Acceleration vs. Time

This interface uses the same principle as above. The difference is that each acceleration and speed increment is shown instead. As before, the graphs can also be used for an individual or for a group of assessors.

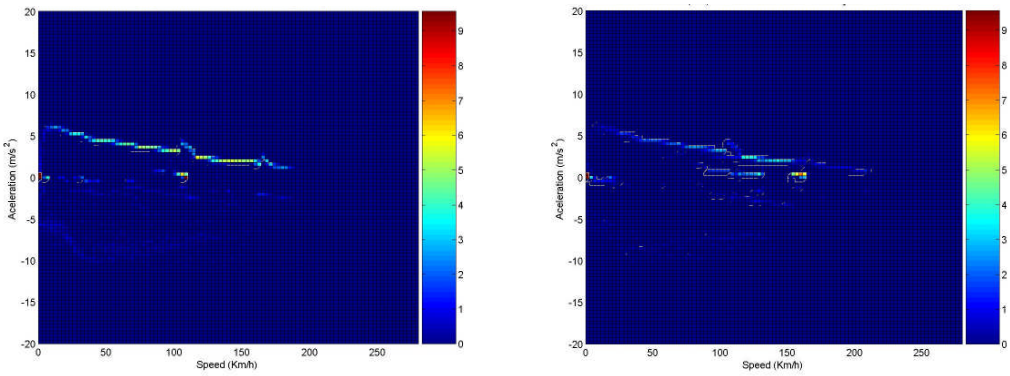

Figure 33: Acceleration vs. Speed

6.3.2.3. Overlapping

Although these 3D frequency distribution graphs do provide insight into the assessment strategies and the behaviour adopted, it was still not possible to identify how the assessors drove to make an assessment. That is, it was not possible to visualise the driving conditions used to make a judgement. This was because each of these interfaces was devoid of information from the time domain. These limitations gave rise to the development of a new data visualisation methods which shows the driving conditions evaluated. This method is referred to as "Overlapping".

The initial idea was formulated following observations of assessor behaviour using the speed time profile graphs (see figure 31 for an example). It was evident from these displays that following the assessors' interactions with the data capture interface a driving manoeuvre would be performed. This gave rise to two assumptions:

1. The interactions with the data capture interface symbolise judgements or decisions made by the assessor.
2. If the assessor had made a conscious effort to assess the vehicles in a consistent manner, then each of the driving manoeuvres, between interactions with the data capture interface, should bear some similarity to each other.

The first assumption is related to the decision making models discussed in chapter 3. The interaction with the data capture interface symbolises the "response" and this is an action that can be observed and recorded.

The second assumption is related to how involved the assessor is in the evaluation. It is necessary to know if they are comparing like-for-like, or if their behaviour is random.

The overlapping method works by providing visual information which shows the "drive – cycle" assessed by the participant. There are four steps to the overlapping method referred to as:

1. Data separation
2. Cutting
3. Overlapping
4. Display

Each of these steps is now described:

6.3.2.4. Data Separation

The recorded data is divided depending on the assessment of the sound relative to the verbal descriptor. In this case the data was separated according to "powerfulness" or "refinement". Each verbal descriptor needs to be considered in isolation as different driving conditions may have been used by the assessor to evaluate the vehicles. In the graphs shown in figure 31, the assessment of powerfulness and refinement is shown by the blue and red coloured lines respectively.

6.3.2.5. Cutting

The cutting process involves dividing the speed-time profile within the time domain, using the assessors' interactions with the touch-screen as markers. This divides the speed time profile into a number of different "events", each representing the stimuli considered to make a decision. The events for the assessment for each attribute are saved into a separate file, which can be used later on in the process.

6.3.2.6. Overlapping

The overlapping process is the most complex of the stages. It involves shifting each of the individual events on top of each other to maximise the degree of overlap between them. The complexity of this step will be discussed in the next section

6.3.2.7. Display

For the display step, the overlapped events are shown on a 3D frequency distribution chart. These show the areas of high intensity which should represent the driving

conditions, used by the assessor, to make an assessment. The overall process is summarised in figure 34.

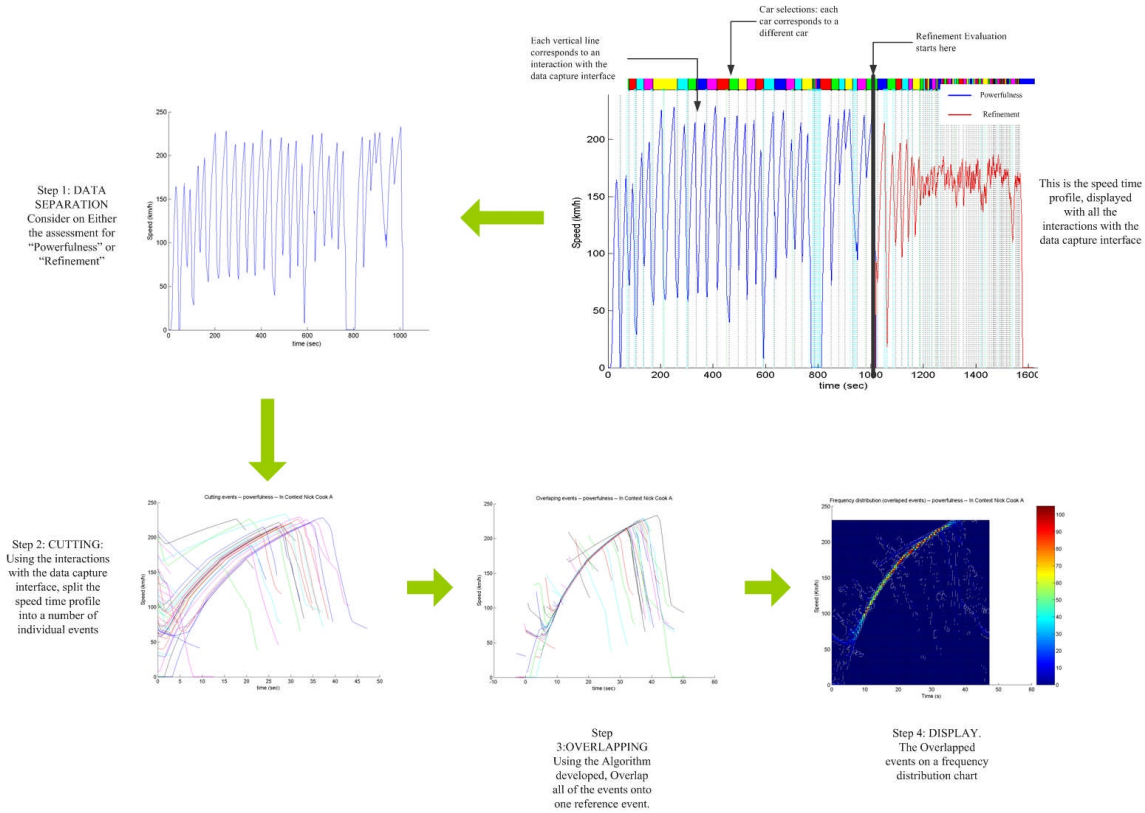

Figure 34: Overlapping process

6.3.2.8. Overlapping approach

A review of possible methods was conducted to find a suitable pattern recognition algorithm which could identify patterns in the assessors' driver behaviour. One method investigated was referred to as cross-correlation. An automated program was developed to consider the events in pairs. One pair was fixed in time, whereas the other was shifted along the time domain until the maximum point of correlation between the two events was found.

The limitation with the method was that it seeks to find the best fit for the whole of the events, rather than looking for specific areas within the events which illustrate how the assessors are making decisions. This was therefore the reason for developing another method.

The new method works by making comparisons between each possible pair of events. The first event is fixed in time; the second is shifted along the time domain. The second

event is shifted along the time domain in increments equal to that of the DOM's sampling frequency (16 Hz).

The difference between the two events is measured within the time domain occupied by both events; this is shown in figure 35. The sections of either event which are not within the same time domain are ignored. The increments are 0.064s long. If the difference between the two sections of events is less than 3km/h then a match is said to have occurred. For every phase shift, the number of matches between the two events is calculated and summed.

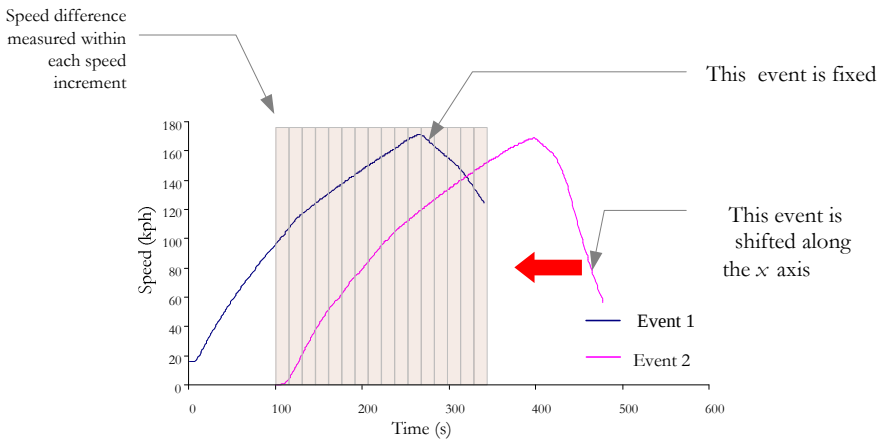

Figure 35: Phase shift with two events

For each event, the total of the number of matches made with every other event is calculated. The event with the highest total is referred to as the reference event, onto which all other events are shifted to.

The overlapping algorithm has been written as a MATLAB script, which enables users to feed raw data collected from the simulator's performance model into a program which automatically generates the overlapping diagrams.

Below are two interesting examples of the application of the overlapping approach. Figure 36, shows the behaviour of one assessor performing WOT and deceleration manoeuvres. Such was conducted by a J&LR employee. It represented a test usually conducted on road, referred to as "tip in".

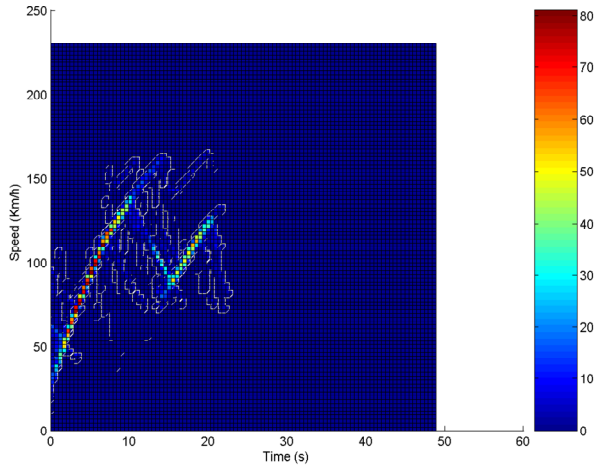

Figure 36: Example of overlapping: Tip-in

Figure 37 shows a different strategy, this is an attempt at cruising with some of the elements of the tip in manoeuvre.

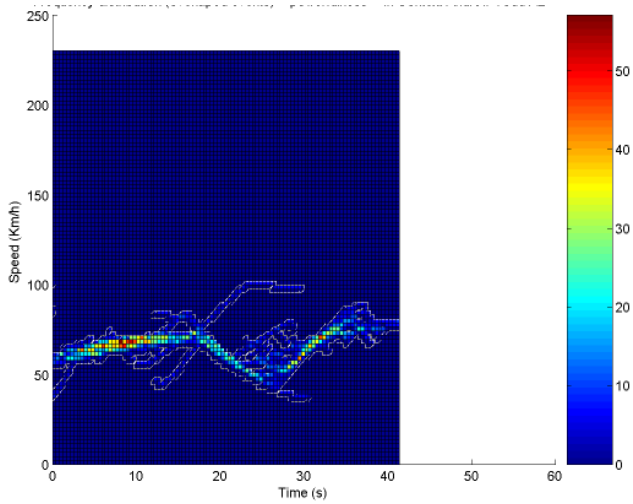

Figure 38: Example of overlapping: cruise

6.3.3. Interpreting the results

To interpret the results, the user then has to select driving manoeuvre that best represents, in their opinion, the drive cycle displayed on the overlapping display. They can do this by viewing the figures generated in the overlapping process (see figure 34). It is then possible to use the speed and time data from the event selected to identify the equivalent pedal positions and RPM trace during this event. This process was also automated so that the data from the selected speed time profile would be used to scan through all the speed time data collected, and using the same approach implemented for the overlapping method and identify the times when similar events occurred. This

returns multiple timings within the data collected. At each of these times the pedal positions and RPM values are captured. In many of the occasions each drive cycle identified will be very similar. From here a drive cycle representative of the majority of the cycles identified should be selected. This is then a visual and numerical representation of how the assessor drove to make an assessment of the vehicle. This process is illustrated in appendix B.

With this data it is now possible to create a fixed driving condition, using the simulator's non-interactive, mode to experience the driving conditions, and the stimuli used by the assessor to make their decision. The use of this functionality will be discussed further on in this chapter.

6.4. A new tool box:

The strengths of the new visualisation methods are their ability to capture un-predictable human behaviour, and produce a visual summary of decision making strategies. They provide scope for new areas of research into Sound Quality and allowed for a greater understanding of the influences of interactivity in subjective perception. Development has already begun to ensure this new tool box is incorporated into updated versions of the software for the simulator.

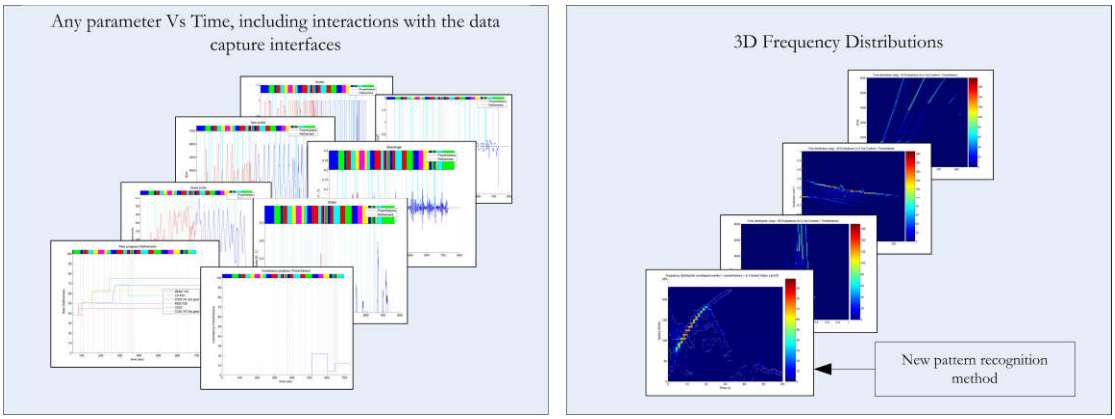

Figure 39: Visualisation tools for driver behaviour

6.5. Influence of interactivity on Sound Quality evaluations.

This new capability allowed for a greater understanding of how interactivity within evaluations, can influence the output. To achieve this, three approaches were conducted in a new experimental study. The first was a comparison of the outputs of interactive and non-interactive evaluations. The second was a questionnaire, asking the assessors to describe if they had assessed the two attributes, powerfulness and refinement, with a different assessment strategy. The third approach was the use of DOM and the visualisation and analysis tools described in the previous section.

Given that two modes of interactivity needed to be experienced by a large group of assessors, the time taken to complete the study became a concern. For this reason the IS method was used as opposed to the paired comparison method. As previously mentioned, the verbal descriptors powerfulness and refinement, are used in sound quality assessments for the Jaguar brand. This study was conducted on behalf of J&LR who were interested in understanding how using an interactive NVH simulator would influence the outcome of the evaluation compared to evaluations conducted using 2GWOTs.

Bisping (1998) conducted a similar study using an earlier version of the interactive simulation which was built into real cars. A dummy head located on the passenger seat captured the sound of the car during the test drive. Then using synthesis techniques and active noise cancellation, the sound heard by the driver was manipulated. Comparisons were made between two driving conditions; both were 2GWOTs but with different rates of change in sound pressure level (SPL). However the assessor had control over both drive cycles as on-road evaluations are "fully-interactive".

His results indicated that different rates of SPL response of the same car yield different subjective impressions. Thus confirming that the use of the product can dictate how it is perceived by the user. Bisping however, does state that further research was needed to explore the effects of acoustic feedback using a variety of sounds and driving conditions.

The purpose of this study however, is to evaluate the effect of the level of interactivity on the assessors' perception of sounds. Therefore the vehicles assessed had the same performance model. In addition, in this study the assessors are not being told how to use the vehicle. In Bisping's study the driver had to perform 2GWOTs. Here however, 2GWOTs were experienced with the non-interactive level, and the assessor could choose how to drive with the fully interactive level. Furthermore Bisping had only used 7

assessors. Within the study conducted as part of this research programme, 36 assessors from various backgrounds took part.

The impact of this study was twofold. First it demonstrated the use of the observational tool box. Second it demonstrated how the attributes powerfulness and refinement would have been assessed if the participant is given the freedom to choose how to interpret the attribute.

The study showed agreement between evaluations using 2GWOTs and those conducted in the fully interactive mode, when powerfulness was assessed. The contrary was true for refinement. Further exploration of the questionnaire responses showed that for powerfulness the assessors evaluated the attribute by accelerating at full throttle, whereas for refinement there was more emphasis on constant speed assessment. Then using the observational data it was possible to understand how each attribute had been assessed.

An initial overview of the Overlapping graphs found that 68% of assessors performed acceleration manoeuvres for the assessment of powerfulness, 9% used a constant speed assessment. For Refinement 53% performed acceleration manoeuvre, 19% drove at constant speed. No pattern could be observed with the remainder of the assessors, however through exploration of the RPM vs. Speed graphs it became apparent that there had been some activity at idle. For the assessment of both verbal descriptors the same 4 assessors (approximately 11%) performed an assessment at idle. No pattern could be found for 12% and 25% of assessors for the assessment of powerfulness and refinement respectively.

Representative drive cycles were identified using the method described in section 6.3. For the assessment of refinement at constant speed the mean speed was 58 mph which occured between 2300 and 2100 RPM. The average maximum throttle position was 44% and the minimum was 35%, and average duration of the decision making event was 4 seconds. For the assessment of refinement during acceleration, the average maximum speed was 90 mph; the average minimum was 16 mph. The average maximum engine speed was 5900 RPM and the average minimum was 1070 RPM. The average maximum throttle was 100% and the minimum was 31%. On average 72 % of the decision making events were spent at full throttle, and the most frequent number of gears changes was 2.

For powerfulness there were fewer people conducting constant speed manoeuvres, and the range of driving strategies was much broader for this verbal descriptor. The average maximum speed was 47 mph; the average minimum speed was 38mph. Most of the assessors performed acceleration manoeuvres which were similar to those for refinement. For the full results see Appendix B.

Figure 40 shows the mean scores for the assessment of refinement and the two main assessment strategies. The modifications simulated for Target O had been made by considering the acoustic properties of car A at full load. As it can be seen, at these driving conditions, there has been an improvement. It has been perceived as being more refined than the other cars in this class using these assessment conditions. The Lexus however is substantially outperforming all the vehicles for the assessment of constant speed. This is possible, as Jaguar has not used constant speed assessment within jury evaluation process. Therefore the acoustic characteristics of these driving conditions have not been considered during the target sound generation phase. This is an example of how improvements made considering the full load do not necessarily result in equivalent improvements at other driving conditions.

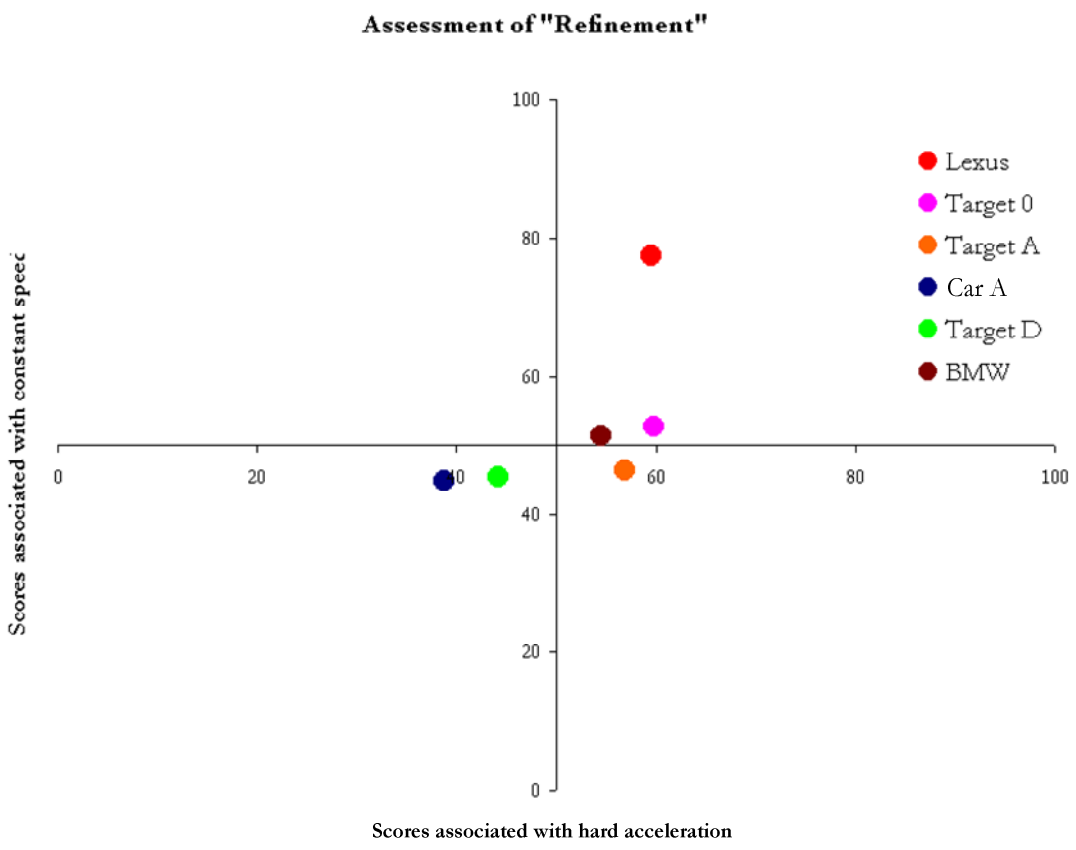

Figure 40: Assessment of refinement

Using the insight captured above, it was possible to remove those assessors who drove at constant speed for refinement, from the comparison between the levels of interactivity. Conducting the comparison again, showed a better agreement between the two levels of

interactivity. Such confirms the hypothesis that the stimuli heard by the assessor can influence their subjective opinions formed.

The study also demonstrated drawbacks in J&LR's existing target setting processes. As the driving manoeuvres other than those represented in the 2GWOT stimuli are assessed. In particular for refinement, the assessor associated this verbal descriptor with sound heard whilst driving at constant speed.

This finding demonstrated the importance of capturing driver behaviour. Without the observational tools provided, the user of the simulator could not associate subjective impression with the stimuli heard. This would inevitably result in uninformed decisions whilst identifying the key acoustic features that influence perception. If these are wrongly assumed then the engineering decisions, regarding mechanical component design and selection would be misdirected. Eventually this could lead to necessary hardware changes after the physical prototypes have been built, thus defeating the purpose of the simulator and sound quality evaluations.

6.6. New data for decision making

Using the visualisation tools developed, it has been possible to link the driving strategies with subjective perceptions. For the J&LR case study this was done for the refinement attribute where two different driving strategies seemed to be used. For powerfulness the majority of the assessors applied the same assessment strategy, and therefore the data could not be split into groups.

Using the groups specified through the observational studies, it would be possible to display the subjective impressions formed on win-lose graphs, but relate these to the driving manoeuvres assessed. Such is shown in figure 41 and 42.

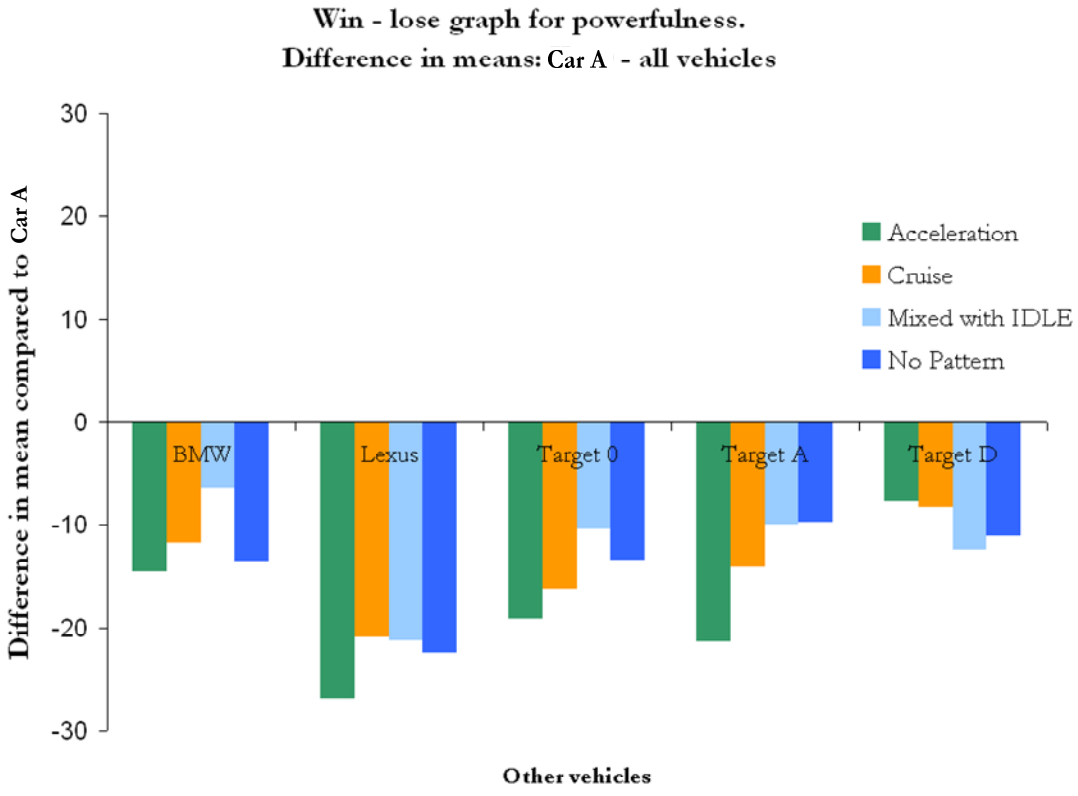

Figure 41: Win - lose graph for powerfulness, segmenting driving style

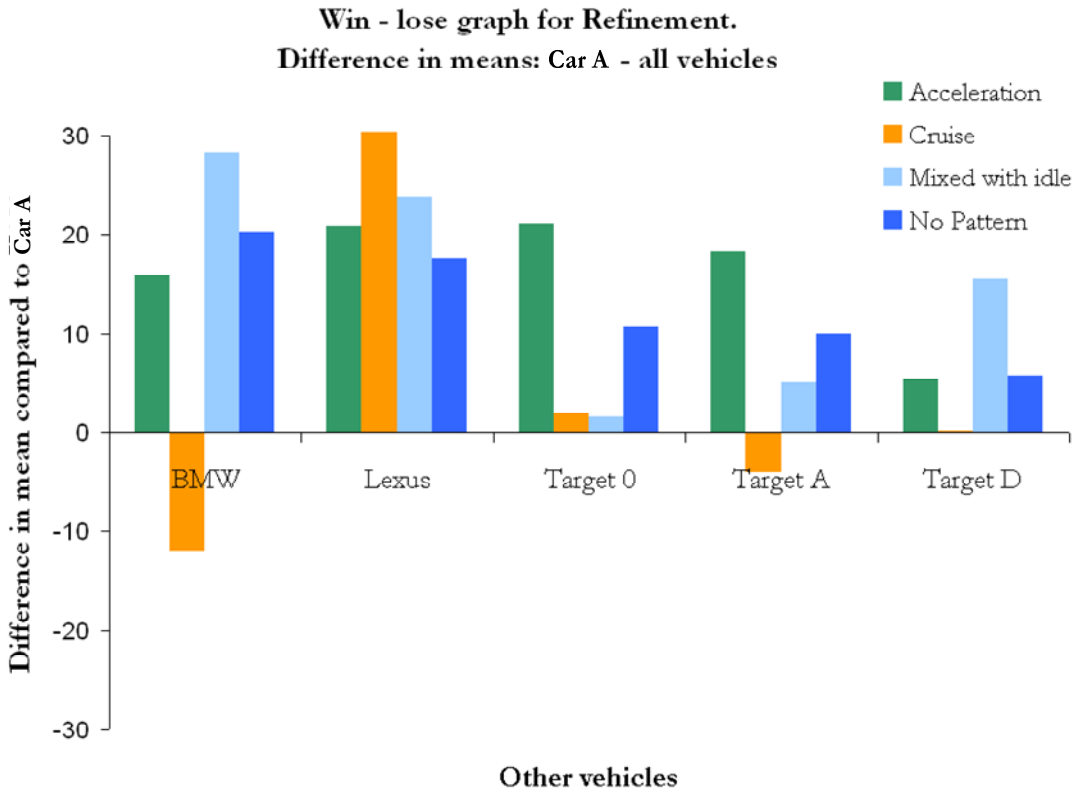

Figure 42: Win - lose graph for refinement, segmenting driving style

The author did not feel that for this demonstration the use of the statistical methods should be applied, given the small numbers of participants in some of the groups.

However the approaches suggested go beyond the interpretation of numerical data. The NVH engineer can now visualise the driving parameters most used in the evaluation, and using this data it is possible to recreate, in the simulator the driving conditions assessed. In this way the engineer can hear what the assessor heard to make their assessment, and can conduct further acoustical analysis on the drive cycle selected.

Using these tools it is now possible to observe how many different types of assessor evaluate the sound of a vehicle using interactive simulator. For example, it is possible to capture how customers, key decision makers in the company and other NVH engineers make an evaluation. Then using the driver data collected it would be possible to re-create the driving conditions in a non-interactive evaluation. The people then involved in the decision making process can play-back the stimuli assessed by all assessors and use this form of information when formulating their decisions, as they gain a further understanding of how each other evaluated the sound, and their subjective impressions were formed.

The benefits of using the simulator in its fully interactive mode meant that assessors could choose which conditions they wanted to assess. However, until the developments to the data capture methods made as part of this research project, it was not possible to relate assessment strategy to subjective opinion.

This new approach means that decision makers can visualise the degree of achievement of the sound quality at each of the driving conditions. They can use the simulator to set vehicle targets for different driving conditions other than WOT manoeuvres. This was not possible before the introduction of the simulator and before the use of the visualisation approaches which ware developed during this research.

In the aftermath of this study,J&LR began further work to develop a database of vehicles that could be driven interactively in the simulator. Their original database of vehicle sounds was made up from recordings of 2GWOTs of their own vehicles and their competitors. Their ambitions were however to drive all of their vehicles in the fully interactive mode. However to establish a new database of sounds, which would be as extensive as the original one, would take time and incur high costs. Instead they developed the vehicle models from their existing 2GWOT sounds, and developed a method to convert these from non interactive stimuli to fully interactive ones.

This signifies their ambitions to conduct fully interactive evaluations in the simulator (Dunne 2007).

6.7. Summary of Chapter

- The interactive assessment of concept cars before the prototyping stages of the vehicle development programme was not possible before the introduction of new simulation tools.
- Key decision makers in industry believe that interactivity is an advantage offered over traditional evaluation tools
- DOM was incorporated into the simulator software, following recommendations made from this research project.
- Visualisation tools were developed to summarise the data collected through DOM.
- The new visualisation tools were used to understand the influence of interactivity on the outcomes of the evaluation of powerfulness and refinement.
    - The study showed that when given a choice assessors will choose to evaluate driving conditions other than 2GWOT.
    - The study also showed that assessors associate the evaluation of refinement to driving at constant speed and acceleration.
- To analyse the data collected, subjective impressions formed must be grouped according to how people drove.
- Before this research programme, NVH engineers could observe the assessors subjective impressions. Now however they are able to observe how assessors drove and hear what they assessed to formulate their subjective impressions. This provides a fuller understanding of the key acoustical features that influence perception.

7. DISCUSSION & RECOMMENDATIONS

7.1. Discussion

The purpose of this chapter is to present guidelines on how to use the simulator for sound quality evaluations. These guidelines are underpinned by the findings from the research discussed in the previous chapters. This chapter will enable the reader to understand how the findings from the different research areas previously discussed merge, and how the new methodological approaches and tools should be applied. The reader should soon appreciate that there are no set rules to setting up and running evaluations in a simulator. Instead potential users are provided with best practice guidance and suggestions to optimise the use of the simulator.

The purpose of developing the simulator was to provide NVH engineers with tools that would enable them to listen to and build existing and concept vehicles in real-time, in more time efficient and effective manner. It was also intended for the simulator to be used as an evaluation tool; this is why the original methodology proposed by the developers was formulated. However the developers of the simulator were, and are, very proficient in developing the simulator technology. Their own methodology was based on many years of experience, and did prove to be a good starting point. However there was a lack of evidence underpinning their methodology and this could have led to limitations in its use. It also fell short in terms of recommending how to cater for the assessors, and how to best capture and analyse data reflecting the assessors' subjective impressions. In addition it did not cater for the interactive aspect of the simulator. This is the main difference between the original methodology and the one proposed here.

The following research questions were raised with the aid of the industrial partners J&LR and SVT.

1. How do assessors engage and perceive the simulator and the task?
2. Understand how to set up experiments using the simulator:
    a. How to design experiments
    b. How to select data capture methods
    c. How to analyse the data collected
3. Understand interactivity:
    a. How to capture how people make decisions
    b. How it could be used to enhance the target setting processes
    c. What is the influence of interactivity?

The answer to each of these research questions is now presented.

7.1.1. How do assessors engage and perceive the simulator and the task?

During the course of the research project it was possible to gain an appreciation for how the assessors engage and perceive the simulator task. Overall the assessors enjoyed their experiences, often looking forward to the evaluation. In fact the majority of assessors indicate that they are more willing to return for interactive evaluations using the simulator than non-interactive ones.

Through exploratory evaluations conducted at the outset of this research programme it was found that a verbal briefing conducted in conjunction with a demonstration can best communicate to the assessor how to the simulator functions.

An issue identified through the use of the full vehicle simulator was simulator sickness. This affects approximately 10% of assessors who take part in the evaluations. It is caused by a mismatch in perception. Sight and hearing give the sensation of movement, the body however is stationary and conflicts with these messages. Experiencing these symptoms led to a decreased level of involvement.

7.1.2. How to design experiments

The initial phase of the research programme was to take into account where the evaluations were conducted and how they were set up. Prior to this research programme the simulator was located within a NVH testing facility. At the time this was a workshop environment unsuitable for meeting and greeting the assessors. There were uncontrollable sources of vibration and noise from the rest of the workshop. This meant that it was not suitable for the evaluations and the development of methodologies. A first step towards achieving the guidelines was therefore the design, installation and commissioning of a new simulator facility.

The author was responsible for commissioning the simulator in its new location at the University of Warwick. The setting now is suitable for assessors; as it is not in a workshop. In fact the new facility has been used to generate much interest in the research through open days and visits from VIPs from around the world. Setting up a dedicated testing facility is recommended, as part of the simulator guidelines. This should be a clean, neutral and a comfortable environment. Particular attention should be paid to

isolating the evaluation area, both in terms of extraneous noise and vibration and outside distractions.

Benchmarking the original use of the simulator was the next stage of the research programme. For this, the original use of the simulator was compared against the literature. It was evident that aspects of setting up the evaluation had not been taken into consideration. This meant that errors such as the "halo effect", "first order errors" and "leniency errors" were present. It is standard practice within the discipline of experimental psychology to remove these sources before the evaluation. Suggestions, by the author of this document, have been made for the design of experiments so to avoid these errors corrupting the observations made.

This ensures that observations made are due to changes in the sounds, and sources of variability are minimised.

For the halo effect, the user of the simulator is recommended to alternate the order of the questions asked to the assessor. For the Jaguar brand the assessors were usually asked to first evaluate vehicles according to how powerful and then according to how refined each one was. Recommendations, by the author of this document, were made so that the order of the question could be alternated for each assessor.

To avoid the first order effects, it is recommended that the user randomises the order of the sounds on the data capture interface. Both randomising the order of the vehicles and alternating the question asked to the assessors have been made easier with recent upgrades of the simulator. Recommendations, by the author of this document, were made to the developers of the simulator to include functionality which enables the user to randomise the vehicles and alternate the questions at a click of a button, when setting up an evaluation. These improvements also provide benefits to the user of the simulator software. Prior to input from this research programme, it would take approximately an hour to modify the programming code responsible for the set-up of the evaluation. Now through the click of a button the evaluation can be set-up in the appropriate manner.

The leniency error originated as there was no standard way of delivery instructions to the assessors. Conventional evaluations in the listening room employed an automated briefing through a PowerPoint presentation. This ensured that all assessors were exposed to the same experimental conditions. For the simulator, this briefing had not been put into practice. The use of an automated briefing and a similar one delivered verbally in conjunction with a demonstration were tested. Assessors responded more favourably to the verbal briefing. This was accredited to the fact that the simulator environment is a real car; however the context of the evaluation is still alien to an assessor. The simulator

responds as a real car and the data capture screen is self explanatory. The assessors however tended to need reassurances regarding how to approach the evaluation. The user is now recommended to follow a standard briefing. Content for this is set out in submission 4.

During the brief it is also fitting to warn assessors of the symptoms of nausea and sickness. Upon setting up the simulator for the first time, the user is recommended to calibrate the new system by adjusting the sensitivities of hardware interfaces. This would include calibrating items such as the steering wheel, foot pedals; the data capture interface, and the alignment of the projection system. The user is also recommended to take this into account when recruiting assessors, as the assessors who suffer from the symptoms should be instructed to stop the evaluation prior to its completion rendering their evaluation void.

7.1.3. How to select data capture methods

To capture subjective impressions the original methodology employed the use of the Interactive Scaling interface (IS). This remains a core aspect of the new methodology. Evidence of the use of this method was found in previous studies. Criteria were proposed and used to evaluate its suitability for interactive evaluations. The criteria have been developed to provide guidance on the suitability of interfaces. They consist of "qualifiers" and "differentiators. The qualifiers are a list of conditions that must be satisfied for an interface to be suitable for subjective evaluations. The differentiators enable the developer of new interfaces to compare it with original interfaces.

A recommendation is that the user implements these criteria if a new interface is wished to be used. Comparisons between new data capture interfaces will often require experimentation and exploratory studies. This was conducted for the interactive scaling interface. It was compared against the paired comparison method. The IS method fared better than the PC method in terms of duration and interactivity. The paired comparison however provided greater levels of accuracy. By this it is meant that the manner in which the results are collected leads to greater potential for distinguishing if one sound is perceived significantly different to another. The paired comparison is also an easier evaluation of the assessors; they simply have to choose one sound in the pair.

The criteria however, are a guide which help the user decide which interface to use. There are not structured rules regarding their use as it would all depend on the context of the evaluation being conducted. If the duration of the evaluation is a priority, then the

author recommends the use of the IS method. If on the other hand, crucial engineering decisions are based on the results from the evaluations, then accuracy becomes the priority and the PC method is recommended.

7.1.4. How to analyse the data collected

Prior to this research programme results from the IS method were being analysed through the use of Exploratory Data Analysis (EDA). To enhance decision making the use of statistics was recommended by the author. This will enable the NVH engineer to decide if a particular sound is significantly preferred over another. Further understanding of the IS method was needed to understand its origins, the level of data attained from its use and which statistical methods to apply. This information was already defined for the paired comparison method.

The analysis methods make the distinction between the use of parametric and non-parametric statistics. The former being preferred over the latter. Parametric statistics are dependent upon two assumptions. The data must be at the interval level and it must also conform to a normal distribution. A review of methods was conducted and led to the assumption that data from the IS method would be interval. However no assumptions were made regarding normality. For this the author identified the Kolmogorov- Smirnov test. If the data does conform to a normal distribution then parametric statistics should be used. The first test suggested is the ANOVA method. If this yields a significant result then two options are available. If the user wishes to conduct a comparison between every possible pair of sounds, then the Tukey HSD method is recommended. However if the user is only interested in a specific comparison between sounds, then the paired t-test can be used. If numerous paired t-tests are required then Bonferroni adjustments must be implemented.

If the data does not conform to a normal distribution, the non-parametric equivalent to the ANOVA and the paired t-test are the Friedman test and Wilcoxon test respectively.

7.1.5. How to capture the way people make decisions

A key aspect of the target setting generation process is linking subjective impressions with the stimuli. This is done so that the key acoustic features that influence perception can be identified. Using the simulator's fully interactive mode the assessor can drive however they wish, meaning that each assessor may evaluate the vehicle's sound at different driving conditions. The assumption was made that assessors make decisions

based on what they hear, and what they hear is dependent on how they drive. Therefore for an NVH engineer to know what key acoustical features influenced the perception of the assessor, they would need to know how the assessor drove and how they perceived the vehicle. The former was not possible before this research programme.

The data capture capabilities of the simulator were further enhanced through the recommendations for the DOM. This records data from the simulators performance model. This includes how the car is performing and the usage of the data capture interface. The inclusion of the DOM into the simulator software was recommended through this research programme; it is recommended that it is used during subjective evaluations. DOM facilitates the use of the fully interactive evaluations by providing insight into the assessors' evaluation strategies.

However to identify the drive cycles most used, new visualisation methods were needed

Effort was focused on developing a method to identify how the assessor drove whilst making decisions. The Overlapping method was one of the tools.

The Overlapping method looks for patterns in the drivers' behaviour and identifies the drive cycle most used. This method uses the data from the data capture interface and how the vehicle was driven, to illustrate a driving condition most used by the assessors. It is a four stage process. The first separates the data collected according to the verbal descriptor assessed. The second cuts the speed time profile into a number of events; using the assessors interactions with the data capture interfaces as markers. The third overlaps each of the events using a custom approach developed within this research programme. The final stage displays the overlapped events on a 3D frequency distribution diagram.

To analyse the data, the user must use all four stages of the overlapping approach, although in reverse order. The frequency distribution diagram will show the intensity of usage of a particular driving cycle. It is recommended that the user works backwards looking at the events which were a result of the cutting stage. This would help them pick the driving condition which represents how the driver assessed the vehicle. This drive cycle can then be used to identify equivalent engine speeds and pedal positions which can be used to construct a fixed driving event file for this simulator, which can be replayed using the simulator's non-interactive level.

7.1.6. How it could be used to enhance the target setting processes

The approach described above provides the NVH engineer with a sound file which represents what the assessors listened to the most for their evaluation. This can be used to identify the key acoustic features that influenced subjective impressions. Using this approach it is possible to listen and experience how each assessor, whether it would be an engineer, key decision maker or customer evaluated the vehicle. This provides the NVH engineers an additional level of information which they can use to decide what acoustical features influenced perception.

In addition it would be then possible to group the assessors according to how they drove. In doing this, it is then possible to observe the scores given to the vehicles at the grouped driving conditions. The scores associated with each group, can then be analysed using the statistical approaches recommended. In doing so it is possible to observe driving conditions which do not achieve the desired sound quality character.

7.1.7. What is the influence of interactivity?

Using the methods proposed above it was possible to study the influence of interactivity. A comparison between non-interactive and fully interactive evaluation was conducted. This concluded that given when evaluating sounds using the interactive level, the assessors choose driving conditions other than those represented by the 2GWOT stimuli. In a case study for the Jaguar brand, it was found that assessors associated the refinement verbal descriptor to steady state driving conditions as well as wide open throttle manoeuvres. This meant that different subjective impressions were formed due to the driving conditions evaluated.

7.2. The Guidelines

The guidelines are based on an understanding of the experimental conditions necessary for subjective evaluations and they also take into account the assessors needs. Evidence was also found for data capture methods, and suitable data analysis methods were proposed. The enhanced data capture capabilities of the simulator were progressed. Now, in addition to the assessors' subjective impressions, their assessment strategy can also be captured. This can lead to enhanced understanding of which aspects of a car need further improvement to achieve an overall vehicle sound quality which meets or exceeds

the customers' expectations of the vehicle. More importantly this can be achieved before detailed designs of the vehicle are conducted, and before physical prototypes are built.

To understand the impact of this research to sound quality evaluations it is not only necessary to look at how the simulator was being used. It is also appropriate to look at how sound quality evaluations were conducted prior to this research programme. The use of listening rooms and the assessment of fixed operating conditions such as 2GWOTs were still the main approach for the evaluations. The data collected was simple. The stimuli upon which the subjective impression was formed were known, making the process of analysing the data straight forward. As for the simulator, this was a prototype tool with the potential to revolutionise how target sounds are selected. It introduced interactivity to the evaluation. The guidelines have provided new approaches to deal with this new variable, and optimise its use providing a new approach to evaluating vehicle sounds.

To demonstrate the changes which have taken place, the new methodology is discussed alongside the original one, initially introduced in chapter 2. Through this discussion reminders of how the recommendations produced are also given. Both the original methodology and the changes that have been made to it are displayed in figure 43. The methodologies proposed as part of this research programme are represented by the red flow chart. The original methodology, retained as part of the new one, is displayed in the blue flow chart. Supporting the methodology are representations of the tools developed during this research programme.

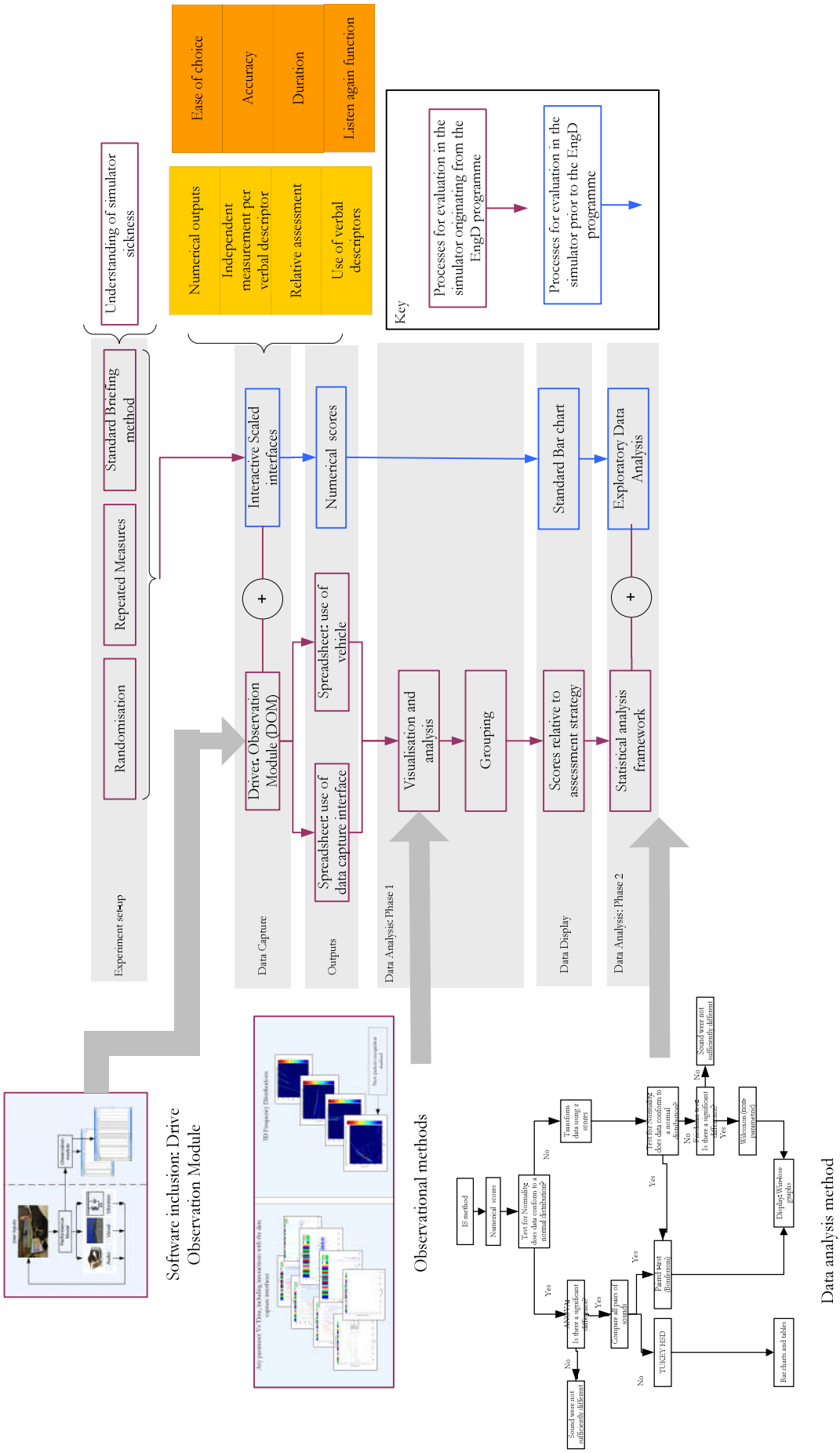

Figure 43: Guidelines for sound quality evaluations using interactive simulation

The following steps should be executed to use the simulator effectively:

1. Consider the location: The simulator should be isolated from extraneous sources of noise and vibration. The environment should be clean, comfortable and welcoming. A waiting area is appropriate; this ensures the evaluation environment is controlled and repeatable.

2. Evaluation design:
    a. to avoid the:
        i. Halo effect, alternate the order of each question posed to the assessor. This reduces a bias with regards to attention paid to the first question posed.
        ii. First order error, randomise the order the sounds appear on the data capture interface.
        iii. Leniency error, use a standard brief. This should be communicated verbally and in conjunction with a demonstration.

3. Consider the assessor.
    a. Simulator sickness affects approximately 10% of assessors. Warnings should be given to the assessor during the briefing before their evaluation. If the assessor exhibits the symptoms of nausea and sickness they should stop the evaluation and rest until they feel better.

4. Choose the data capture method:
    a. If the evaluation is a benchmarking exercise, then there are likely to be numerous cars. In this case the duration will be a concern, and the accuracy required may not be as high as that required for the validation stages, where fewer cars are assessed. For the former the author recommends the use of the interactive scaling method. For the latter the paired comparison is suitable

5. Use DOM with the fully interactive level:

6. Data analysis method:
    a. Step 1: use the visualisation methods, on each assessor, with the data collected from DOM.
    b. Step 2: group the assessors into categories dependent on how they drove.
    c. Step 3: To understand how each vehicle performed at each driving condition experienced, a statistical analysis for each category is necessary.
        1. First test the data in each category to see if it conforms to a normal distribution. If it does not, use z-scores to transform the data, and test for normality again with the Kolmogorov Smirnov goodness for fit test.
        2. If the data conforms to a normal distribution, the ANOVA method followed by the Paired t-test should be applied
        3. If the data does not conform then the Friedman test followed by the Wilcoxon test should be applied.

    d. Step 4: To understand the acoustic features that influenced perception:
        1. Use the visualisation methods to choose and create a drive cycle representative of each category.
        2. Create a fixed driving event for the simulators' non interactive level, and record the sound heard at these driving conditions
        3. Using the vehicles that were statistically significantly different, identify the key acoustic features that influence perception.

7. Use the engineering tools in the simulator to design a new sound given the new insight
8. Validate the results through another subjective evaluation.

8. FUTURE WORK

To date some recommendations from the research findings have been put into practice through upgrades to the simulator software. For the ones that have not, plans are in place to do so. However the contributions made so far could benefit from further research.

- Understanding the differences between the subjective impressions of key decision makers and customers

NVH engineers and key decision makers are eager to know how their own personal judgement of vehicle sound quality differs from that of customers. Often target sounds generated can be very similar. For these areas to be addressed there is a need to research into how different stimuli have to be before an assessor can tell the difference. This is expected to be different depending on the type of assessor. Expert NVH engineers may be more sensitive to small acoustical differences than customer. For this to occur there may be a need to investigate sounds using the Just Noticeable Differences method (JND). This is a method applied for the sensory perception of food. This investigation is now possible with the latest update of the simulator software; as it is possible to incrementally change the loudness of vehicles, and / or the contributions from the components which make up the overall sound

- Understand driver behaviour in the real world, and develop methodologies accordingly.

Developments in technology mean that the simulator software can be fitted to real cars. The observational tools and techniques can be further developed so they integrate into this new capability. It will be possible to capture how decisions are made in real cars. Capturing decision making in a real car, can then lead to guidance for further enhancement of methodologies in the simulator.

- Understand how to involve the customer in the design of target sounds.

The data capture methods investigated up to now are ideal for the integration of data from subjective evaluations into existing target setting approaches. In doing this however, the simulator's true potential has not been realised. There is the potential to develop methodologies which involve the user in the design of target sounds. Approaches such as user design allow the assessor to reconfigure the sound of a vehicle. In essence the assessor designs their own vehicle. Exploration of this method opens up a variety of challenges. It would be necessary to understand the limits of the method that is how many components the assessor is allowed alter. In addition it will be necessary to understand how to analyse data from the user design method, and how to design the layout of interfaces

The user design approach could also be used by car designers who at the present moment deal with the aesthetics of the vehicle. Designers could use the user design method to demonstrate how they think the car should sound as well as look.

On the other hand, as cars become mass-customised, there may be scope for an interactive NVH simulator in every showroom. Its use alongside the User Design method will provide the customer with the option of choosing how their car will sound.

9. CONCLUSIONS

This Innovation Report has reviewed the outcomes of a research project into the development of new and necessary methodological approaches for sound quality evaluations using interactive simulation. The purpose of the research programme was to develop tools and techniques, and to formulate user guidelines which ensure a rigorous and scientific approach to SQ evaluations within the simulator. The impact of the research project should contribute to improving the effectiveness and efficiency, of the NVH target setting process. This is done by maximising the use of the interactive aspect of the simulator.

The interactive NVH simulator used within this research programme was developed by Sound & Vibration Technology Ltd (SVT) to enhance sound quality engineering processes. Prior to the research project, the simulator's potential within SQEs was limited due to the manner in which it was being used. The initial methodological approaches and their implementation were not based on a rigorous research programme. This meant that although the simulator was being implemented within active vehicle development programmes, there was little confidence in the results. Similarly the ability to conduct the evaluations in an interactive environment was seen, by industry experts, as its most valuable asset. Yet methods to produce and treat data from interactive evaluations were not in existence.

The aim of this work was therefore to develop new practice, based on evidence, on how to set up, record and analyse subjective evaluations using interactive simulation.

1. Develop a set of "guidelines" for the design of experiments conducted using the simulator with, which optimise the new functionality.
    a. Propose a set of recommendations of how to set up and conduct sound quality evaluations in the simulator.
    b. Develop the tools and methods to capture how assessors make decisions when evaluating sound in the simulator.

The recommendations now encompass guidelines on how to set up the sounds on the data capture interface and which order to questions to the assessors. Issues relating to the assessors' health are also considered. Warnings of simulator sickness are given at the

beginning of each evaluation. With regards to the type of data capture interface, research was undertaken to identify criteria for the selection of data capture interfaces suitable for use with interactive NVH simulator. It was then used to justify the use of the method referred to as Interactive Scaling (IS). It was necessary to identify suitable statistical analysis techniques for this method. For this a framework was proposed which selects statistical tests according to the nature of the data collected.

Prior to the research project, observational studies were limited to the ability of the experimenter of the simulator to subjectively observe how the assessors may have driven. An outcome from the project was a Driver Observation Module (DOM), which recorded how the assessors drove the vehicles and used the data capture interface during an evaluation. Data collected as a result of the use of DOM needed the development of methods to visualise it. These help identify the driving manoeuvres the assessor used to formulate a decision regarding their opinion of the vehicle's sound. These new approaches enable NVH engineers to see and hear how assessors drove to make an assessment and hear the stimuli.

The new approach was used to demonstrate how assessors' evaluate the attributes powerfulness and refinement. Driving conditions other than 2GWOT (2GWOT) were being used to assess refinement. This meant the assessment strategy adopted by assessors within traditional evaluation conditions was not representative of how the assessor interpreted the attributes assessed.

The observational tools and the experiment design framework have already been partially implemented. Additional software updates will be incorporated in due course.
The research into driver behaviour during sound quality evaluations identified that WOT conditions alone do not match the requirements for the assessment of refinement. J&LR have therefore begun exploring part load conditions as well, evidence of this can be found at Dunne (2007)
The guidelines are set to be put into practice in a proposed collaborative study between the University Of Warwick and Windsor University in Canada, to explore the cultural differences between the two nations. Recommendations regarding methodologies have been provided to other OEMs who have recently adopted the simulator technology.
To date the research findings have been disseminated in the following way (for further details see appendix C):

- 1 International peer reviewed Journal (JSAE)
- 5 conference papers, including the SAE NVH biannual conference. For this conference the driver behaviour work was reviewed by some of the most established NVH researchers, including one of the author's of the original listening room evaluation guidelines.
- 1 Poster at the JSAE automotive exhibition. (Who and where)
- An additional conference paper regarding the selection of inputs for neural network was published and presented (Fry et al 2006)

The research as a whole has provided much needed insight into evaluations based in the interactive NVH simulator, and has a strong foundation upon which to pursue future research. This future research is set to be explored via a collaborative project between the University of Warwick, J&LR and Bruel & Kjaer.

The research conducted has led to new practice being implemented in the use of the simulator and has contributed to its development from a prototype to a solution for subjective evaluations. This is now being supplied to OEMs on a global scale through B&K.

10. APPENDIX A: ACCURACY TESTING FOR DATA CAPTURE METHODS

Tests for significance

Assessment of Powerfulness								
	data capure method		Paired Comparison	Interactive Scaling				
	Stat method		Critical difference	Paired T test	Paired T test	Paired T test	Wilcoxon's	TUKEY
	treatment of scores		ABS difference		Z Scores	Normalised: 0-100	Z Scores	ABS difference
	Car 1	Car 2	SIG	SIG	SIG	SIG	SIG	SIG
	Car 1	Car 3	SIG	SIG	SIG	SIG	SIG	
	Car 1	Car 5	SIG					
	Car 1	Car 6	SIG	SIG	SIG	SIG	SIG	
	Car 2	Car 3	SIG	SIG	SIG	SIG	SIG	
	Car 2	Car 5	SIG	SIG	SIG	SIG	SIG	SIG
	Car 2	Car 6	SIG	SIG	SIG	SIG	SIG	
	Car 3	Car 5		SIG	SIG	SIG	SIG	
	Car 3	Car 6	SIG					
	Car 5	Car 6	SIG	SIG	SIG	SIG	SIG	

Table 4: Significant differences for Powerfulness

Assessment of Refinement								
	data capure method		Paired Comparison	Interactive Scaling				
	Stat method		Critical difference	Paired T test	Paired T test	Paired T test	Wilcoxon's	TUKEY
	treatment of scores		ABS difference	Z Scores	Z Scores	Normalised: 0-100	Z Scores	ABS difference
	Car 1	Car 2	SIG	SIG	SIG	SIG	SIG	
	Car 1	Car 3						
	Car 1	Car 5						
	Car 1	Car 6	SIG					
	Car 2	Car 3	SIG	SIG	SIG	SIG	SIG	SIG
	Car 2	Car 5	SIG	SIG	SIG	SIG	SIG	SIG
	Car 2	Car 6	SIG	SIG			SIG	
	Car 3	Car 5	SIG					
	Car 3	Car 6	SIG					
	Car 5	Car 6	SIG					

Table 5: Significant differences for Refinement

11. Appendix B: Driver Behaviour Insights

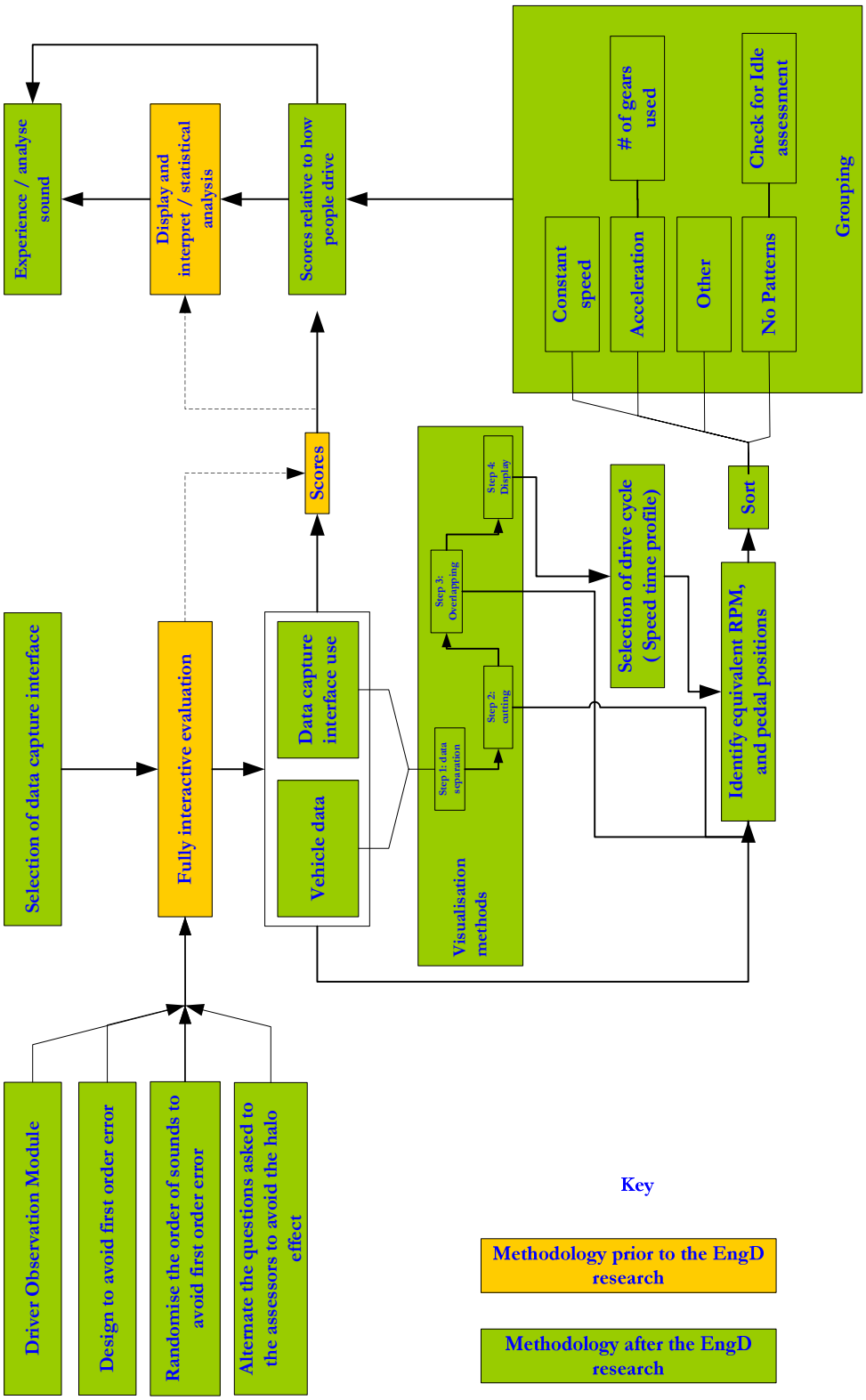

Figure 44: Process to capture drive cycles

Drive Cycle Type	Attribute	Number of people	Max speed (kph)	Min speed (kph)	Max RPM	Min RPM	% Time at full throttle	Max throttle (%)	Min throttle (%)	Average speed	% Time at cruise	Duration (s)
Constant speed 100% of	Refinement	7	94.21	92.86	2320.03	2109.39	0.00	43.84	34.59	93.59	100.00	2.95
	Powerfulness	3	68.43	67.75	1504.60	1471.58	0.00	34.15	27.47	68.02	100.00	4.99
Constant speed <100%	Refinement	1	75.15	61.88	1663.00	1414.50	0.00	42.86	43.66	66.55	44.02	13.01
	Powerfulness	0	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A

Drive Cycle Type	Attribute	Number of people	Max speed (kph)	Min speed (kph)	Max RPM	Min RPM	% Time at full throttle	Max throttle (%)	Min throttle (%)	Speed range (kph)	Number of apparent gears used	Duration (s)
WOT	Refinement	14	144.30	24.95	5913.39	1067.36	78.12	100.00	30.77	119.35	2.00	19.81
	Powerfulness	24	147.76	24.92	5907.64	1135.02	78.36	99.84	23.08	122.84	2.00	16.19
Part load	Refinement	3	114.05	17.61	4532.40	1095.65	0.00	89.08	0.00	96.44	2.00	16.25
	Powerfulness	0	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A

Drive Cycle Type	Attribute	Number of people
IDLE	Refinement	4
	Powerfulness	4

Figure 45: drive cycles identified

Overview of driving styles

Assessment type	POWER N = 34		REFINE (N = 36)	
	Number of assessors	Approximate percentage	Number of assessors	Approximate percentage
Accelerating	23	67%	16	44%
Cruising	3	9%	7	19%
IDLE	4	12%	4	11%
No Pattern	4	12%	9	25%

12. APPENDIX C: DISSEMINATION OF THE RESEARCH

Journals:

Jennings, P., Giudice, S., Fry, J., Williams, R., Allman-Ward, M. & Dunne, G. (2007) Developing Best Practice for Sound Quality Evaluations using an Interactive NVH simulator. *Transactions of the Japanese Society of Automotive Engineers*, 38, 31 - 36.

Conferences and presentations:

Fry, J., Jennings, P., Giudice, S., Williams, R., Allman-Ward, M. & Dunne, G. (2006) *Development and evaluations of "powerful" and "refined" neural networks for the efficient choice of automotive sound quality targets*. Proceedings of InterNoise. Honolulu, Hawaii.

Giudice, S., Jennings, P., Cain, R., Williams, R., Allman-Ward, M. & Dunne, G. (2008) *Optimising the value of Sound Quality Evaluations by observing assessors' driving strategies*. Proceedings of EuroNoise 08. Paris.

Giudice, S., Jennings, P., Cain, R., Williams, R., Allman-Ward, M. & Dunne, G. (2007) *Using and Interactive NVH Simulator to Understand Driver Behaviour*. SAE NVH. St. Charles, IL, USA.

Giudice, S., Jennings, P., Fry, J., Williams, R., Allman-Ward, M. & Dunne, G. (2006) *Development and evaluation of a new sound quality evaluation approach for interactive NVH simulators*. Proceedings of InterNoise. Honolulu, Hawaii.

Jennings, P., Giudice, S., Williams, R., Allman-Ward, M. & Dunne, G. (2005) *Developing Best Practise for Sound Evaluations using an interactive NVH simulator*. Proceedings of the Congress of the Japanese Society of Automotive Engineers. Yokohama.

Williams, R., Allman-Ward, M., Jennings, P., Batel, M., (2006). *Using an Interactive NVH Simulator to capture and understand customer opinions about vehicle sound and vibration*. Automobile Comfort Conference, Le Mans, SIA.

Posters:

JSAE Automotive Engineering Expo: Experiential Engineering

13. REFERENCES

Afaneh, A. H., Abdelhamid, M. K., Qatu, M., (2007) *Engineering Challenges with Vehicle Noise and Vibration is Product Development*. SAE 2007 Noise and Vibration. SAE. St. Charles, IL, USA, SAE: 2007-01-2434.

Allman-Ward, M., Dunne, G., Williams, R., Jennings, P., (2004) *The evaluation of vehicle sound quality using an NVH simulator*. Proceedings of InterNoise 2004, Prague.

Allman-Ward, M., Williams, R., Cockrill, M., Distler, H., Crewe, A., Heinz, T., (2003) *The Interactive NVH Simulator as a Practical Engineering Tool,* SAE 2003 Noise & Vibration Conference and Exhibition. Grand Traverse, MI, USA, SAE: 2003-01-1505.

Amman, S. and Blommer, M. (1999) *Psychoacoustic Considerations in Vehicle Ergonomic Design*. SAE International Congress & Exposition. Detroit, MI, USA, SAE: 1999-01-1269

Amman, S. A., Meier, R., Mouch, T., Gu, P. (2005) *A Survey of Sound and Vibration Interaction*. SAE 2005 Noise and Vibration Conference and Exhibition. SAE. Grand Traverse, MI, USA, SAE: 2005-01-2472.

Baker, S. Jennings, P., Dunne, G., Williams, R. (2004). *Improving the effectiveness of Paired Comparison Tests for Automotive Sound Quality*. 11th International congress on Sound and Vibration. St. Petersburg, Prague. p.3029 – 3036.

Bisping, R. (1995). *Emotional Effect of Car Interior Sounds: Pleasantness and Power and Their Relation to Acoustic Key Features*. SAE transactions: 951284, 104 (2), p. 2207-2213

Bisping, R. (1998). *Feedback of Sound as a Major Determinant of Sound Quality*. Proceedings of Euronoise 98. Munich. p.541-544.

Blauert, J, (1994), *Product-Sound Design and Assessment: An Enigmatic Issue from the point of view of Engineering?* Proceedings of Inter Noise 94, pp 857-862.

Bodden, M., Heinrichs, R., (1998). *Sound Quality Evaluation of Interior Vehicle Noise Using an Efficient Psychoacoustic Method*. Proceedings of Euronoise 98. Munich, p. 609 - 614.

Bower, J. A. (1995). Statistics for food science III: sensory evaluation data. Part A - sensory data types and significance testing., *Nutrition and Food Science* 6, 35-42.

BS ISO 4121:2003, (2003). *Sensory analysis - Guidelines for the use of quantitative response scale*. British standard.

Cain, R. (2005). *Involving Users in the Design Process: The Role of Product Representations in Co-Designing*. Department of Design and Technology. Loughborough University. Ph.D Thesis.

Coolican, H. (1999). Research Methods and statistics in Psychology. London, Hodder & Stoughton.

Crewe, A., Bogema, D., Williams, R., Ballam, M., and Allman-Ward, M., (2003), *Sound Decomposition - A Key to Improved Sound Simulation*. Noise & Vibration Conference and Exhibition, May 2003, Grand Traverse, MI, USA, SAE 2003-01-1423.

David, H. A. (1988). *The method for Paired Comparison*. London, Charles Griffin & Co. Ltd.

Dorn, L. and Barker, D. (2005). *The effects of driver training on simulated driving performance.* accident analysis and prevention 37, 63 - 69.

Dunne, G., (2003), *The Introduction of Sound Quality Engineering Process to Jaguar Cars*, University of Warwick, EngD thesis.

Dunne (2006). *Application of a NVH simulator in the product development process Presentation at EIS Perception of Product Sound and Vibration Workshop,* University of Warwick, 14 December 2006.

Dunne, G., Williams, R., Allman-Ward, M. (2007), *An Efficient Approach to Powertrain Sound Quality Decisions Making Based on Interactive Evaluations Using an NVH Simulator.* St. Charles, IL, USA, SAE:2007-01-2392.

Freeman, V, (2003), *BMW Z4 - BMW aims to become bigger noise*, Times Online, viewed 24 May 2004, < (http://driving.timesonline.co.uk/article/0,,12929-1073787,00.html)>

Fry, J., Jennings, P., Giudice, S., Allman-Ward, M., Dunne, G., (2006). *Development and evaluation of "powerful" and "refined" neural networks for the efficient choice of automotive sound quality target*s. Proceedings of Internoise 2006, Honolulu, Hawaii, USA. In06-516.

Fukuhara, C., T. Kamura, Suetomi, T., (2002). Subjective evaluation of engine acceleration sound with driving simulator. *JSAE Review* 23(4), 435-441.

Gardner, P. L. (1975). Scales and statistics. *Review of Educational Research* 45(1): 43 -57.

Genell, A. and D. Vastjall (2007). Vibrations can have both negative and positive effects on the perception of sound. *International Journal of Vehicle Noise and Vibration* 3(2), 172 - 184.

Genuit, K. and W. R. Bray (2001*). A virtual Car: Prediction of Sound and Vibration in an Interactive Simulation Environment.* SAE Noise & Vibration Conference & Exposition, Grand Traverse, MI, USA, SAE: 2001-01-1474

Genuit, K. and A. Fiebig (2007). The influence of combined environmental stimuli on the evaluation of acoustical comfort: case studies carried out in an interactive simulation environment. *International Journal of Vehicle Noise and Vibration* 3(2), 119 - 129.

Goetchuis, G., C. Ketelhut, et al. (2001 - 01 - 1590). Subjective evalaution of NVH CAE Model Predictions Using an Operator - in - the-Loop Driving Simulator. SAE Noise & Vibration Conference & Exposition. Traverse, MI, USA: 2001-01-1590

Goldstein, E. B., (1999), Sensation & Perception, United States of America. Brookes / Cole.

Hancock, P. A., Lesch, M., Simmons, L.,. (2003). The distraction effects of phone use during a crucial driving manoeuvre. *Accident Analysis and Prevention* 35, 501 - 514.

Harrison, M. (2004), Vehicle Refinement, Controlling Noise and Vibration in Road Vehicles, Butterworth Heinemann, London.

Hutchins, S. M., MacDonald A. M., (1992). *Noise, vibration, and harshness from the customer's point of view*. FISITA congress, London, FISITA.: 925181.

Jakobsson, U. (2004). Statistics presentation and analysis of ordinal data in nursing research. *Scandinavian Journal of Caring* Science 18, 437-440.

Janssens, K. (2006). A *sound design approach for vehicles and aircrafts,* Proceedings of Inter-Noise 2006. Honolulu, Hawaii, USA, INCE. In06-689.

Jay-Cerrato, G. (2007). Sound/Vibration Quality Engineering: Part 1 - Introduction and the SVQ Engineering Process. *Sound and Vibration Magazine.2007*, vol. 41, 4, pp. 16-25

Koppius, O. R., v. Heck,. E., Wolters, J. J., (2004). The importance of product representations online: empirical results and implications for electronic markets. *Decision Support Systems* 38, 161 - 169.

Krapichler, C., Haubner, M., Engelbrecht, R,. Englmeier, K., et al. (1998). VR interaction techniques for medical imaging applications. *Computer method and programs in Biomedicine* 56, 65 - 74.

Kuti, T., A.Hegyi, Kemeny S. (2004). Analysis of sensory data of different food product by ANOVA. *Chemometrics and intelligent laboratory Systems* 72, 253-257.

Lo, W. T. and So, R. H. Y. (2001). Cybersickness in the presence of scene rotational movements along different axes. *Applied Ergonomics* 32, 1-14.

Maunder, R. M. S. (1998). An interactive subjective assessment method for recorded sound. *ImechE conference transactions*, London, Mechanical Engineering Publications, London, ROYAUME-UNI (1998) (Monographie).

Min, B. C., Chung, S. C., Yoon-Ki, M., Kazuyoshi, S., (2004). Psychophysiological evaluation of simulator sickness evoked by a graphic simulator. *Applied Ergonomics* 35, 549-556.

Mourant, R. R. and T. R. Thattacherry (2000). *Simulator Sickness in a virtual environments driving simulator*. Human Factors and Ergonomics Society Annual Meeting Proceedings, Proceedings 1 - Virtual Environments, p. 534-537.

New York Times (1995) Milwaukee (Reuters), New York Times, sections 1, 2 p. 26.

Otto, N., Amman, S., Eaton, C, Lake, L., (1999). *Guidelines for Jury Evaluations of Automotive Sounds*. Noise & Vibration Conference & Exposition. Traverse City, MI, USA.

Parizet, E., N. Hamzaoui, Sabatie, G., (2005). Comparison of some listening test methods: a case study. *Acustica united with Acta Acustica* 91, 356-364.

Pense-Lheritier, A. M., Guilabert, C., Bueno, M., Sahnoun, M., Renner, M., (2006). Sensory evaluation of the touch of a great number of fabrics. *Food Quality and Preference* 17, 482-488.

Pielemeier, W. J., J. Greenberg, Otto, N., Jeybalan, V., Meier, R., (2001). *Some Factors in the Subjective Evaluation of Laboratory Simulated Ride*, SAE Noise & Vibration Conference & Exposition,. Grand Traverse, MI, USA, SAE: 2001-01-1569.

Pierson, R., and Bozmoski, A., (2003), 'Harley-Davidson's 100th Anniversary – The Sound of a Legend', Sound and Vibration, 2003, vol. 37, 3, pp. 14-17.

Piggot, J. R. (1984). Sensory Analysis of Food. London and New York, Elsevier Applied Science Publishers.

Repik, E. P. (2003). *Historical Perspective on Vehicle Interior Noise Development*. SAE 2003 Noise & Vibration Conference and Exhibition. Grand Traverse, MI, USA, SAE: 2003 - 01 - 1518.

Resurreccion, A. V. A. (1998). Consumer sensory testing for Product development, Apen Publishers. Springer; 1st edition.

Roberts M.J., Russo, R., (1999), A student's guide to ANALYSIS OF VARIANCE, Routledge, London.

Rossi, G., F. Crenna, Codda, M., (2003). Measurement of quantities depending upon perception by jury-test methods. *Measurement* 34, 57-66.

Schifferstein, H. N. J. (2006). The perceived importance of sensory modalities in product usage: A study of self-reports. *Acta Psychologica* 121(1), 41-64.

Schifferstein, H. N. J. and Cleiren, M. P. H. D. (2005). Capturing product experiences: a split-modality approach. *Acta Psychologica* 118(3), 293-318.

Schulte-Fortkamp, B., K. Genuit, Fiebig, A., (2007). A New Approach for Developing Vehicle Target Sounds. *Sound and Vibration,* 12 -14.

Shen, W., and Parsons, K. C., (1997). Validity and reliability of rating scales for seated pressure discomfort. *International Journal of Industrial Ergonomics* 20, 441 - 461.

Stone, H. and J. Sidel (2004). Sensory Evaluation Practices. Amsterdam, Elsevier Academic Press. 3rd edition.

Terazawa, N., Y. Kozawa, et al. (2000). Objective Evaluation of Exciting Engine Sound in Passenger Compartment during Acceleration. SAE 2000 World Congress, March 2000, Detroit, MI, USA, SAE 2000-01-0177

Tilley, A., (2003), An Introduction to research methodology and report writing in psychology, Brisbane, Pineapple Press.

Västfjäll,, D, and Kleiner, M, (2002), *Emotion in Product Sound Design*, Proceedings of Journees Design Sonare, Paris 2002.

Villanueva, N., Patenate, A. J., Da Silva, M., (2005). Performance of the hybrid hedonic scale as compared to the traditional hedonic, self adjusting and ranking scales, *Food Quality and Preference* 16, 691-703.

Wickens, C. D. and J. G. Hollands (1999). Engineering Psychology and Human Performance. New Jersey, Prentice Hall. 3rd edition.

Williams, R., Allman-Ward, M., Jennings, P., Batel, M., (2006). *Using an Interactive NVH Simulator to capture and understand customer opinions about vehicle sound and vibration*. Automobile Comfort Conference, Le Mans, SIA.

Williams, R., Henderson, F., Allman-Ward, M., Dunne, G., Cockrill, M., Adlard, S., (2005). *Using an Interactive NVH Simulator for target setting and concept evaluation in a new vehicle programme*. SAE 2005 Noise and Vibration Conference and Exhibition, Grand Traverse, MI, USA, SAE: 2005-01-2479

Zeller, P. and Zeitler, A., (2005). Psychoacoustic-Based Sound Design in Vehicle Engineering. JSAE Annual Congress. *Yokohama, JSAE*. 6-05, 17-20.

Zeitler, A., Fastl H., (2006). Methodological approaches to investigate the effects of meaning, expectations and context in listening experiments. Internoise 2006, Honolulu, Hawaii, USA. In06-123.

Zwicker, E. and Fastl, H. (1999). Psychoacoustics: Facts and Models (Springer Series in Information Sciences). 2nd updated ed.