CHALMERS

Design methodology of 3D sounds for in-vehicle auditory information systems

Master's thesis in the Master's programme in Sound and Vibration

VIJENDRA BHAT & YIDAN LAI

Department of Civil and Environmental Engineering
Division of Applied Acoustics
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2013
Master's thesis 2013:4

MASTER'S THESIS IN THE MASTER'S PROGRAMME IN SOUND AND VIBRATION

Design methodology of 3D sounds for in-vehicle auditory information systems

VIJENDRA BHAT & YIDAN LAI

Department of Civil and Environmental Engineering
Division of Applied Acoustics
CHALMERS UNIVERSITY OF TECHNOLOGY

Gothenburg, Sweden 2013

Design methodology of 3D sounds for in-vehicle auditory information systems
VIJENDRA BHAT & YIDAN LAI

Master's thesis 2013:4

Department of Civil and Environmental Engineering
Division of Applied Acoustics
Chalmers University of Technology
SE-412 96 Gothenburg
Sweden
Telephone: +46 (0)31-772 1000

Cover:
Picture taken at A2Zound Studio, Semcon Tech Center, Göteborg

Chalmers Reproservice
Gothenburg, Sweden 2013

Design methodology of 3D sounds for in-vehicle auditory information systems
Master's Thesis in the Master's programme in Sound and Vibration
VIJENDRA BHAT & YIDAN LAI
Department of Civil and Environmental Engineering
Division of Applied Acoustics
Chalmers University of Technology

Abstract

Auditory signals are proven to be effective as a means of attracting the driver's attention and conveying relevant information about a given traffic situation. Two most common ways of conveying non-verbal auditory information are : by the use of earcons; and auditory icons. Owing to the fact that both earcons and auditory icons are unable to present the information about the direction of an impending danger, auditory signals can be aptly complemented by the use of directional sounds or otherwise called 3D sounds.

The objective of 3D sounds is to draw the driver's attention to a potential threat by localizing the sound in the respective direction. There have been studies that have addressed this idea of using 3D in-vehicle warning sounds but they have seldom focused on the development of it's design methods. Therefore, this study aims at investigating the concept of using 3D sounds as a means of enhancing the in-vehicle early warning systems, by developing and implementing a design and evaluation methodology for this purpose.

An iterative sound design process was proposed as the methodology. The 3D sounds were designed using a combination of an audio mixing software, a binaural 3D plug-in and different MIDI synthesizers. The early phase of sound design consisted of making earcon and auditory icon sound sketches based on the two chosen driving scenarios. Next, these sketches were chosen for evaluations on two jury groups in total, who's feedbacks were used to improve and perform modifications until satisfactory sketches were designed. These improved sketches were tested and then investigated by performing an individual listening test.

The investigation primarily involved studying the participants' emotional responses to the sounds; using scale ratings to test the appropriateness of sounds for the two chosen scenarios; and addressing the problem of 'front-back confusion' concerning the localization of 3D sounds. Finally, the final sound designs were ranked in terms of preference for each of the two chosen scenarios and the efficiency of the sound sketching method was also tested.

Keywords: Earcons, Auditory Icons, 3D Sounds, Sound Design, Psychoacoustics, Intelligent Transport Systems

Preface

This project was funded and initiated by A2Zound, a Semcon group company. This project is a part of the research project EFESOS (Environment Friendly Efficient Enjoyable and Safety Optimized System) with Semcon, VTI, Victoria Institute, HiQ, Chalmers University of Technology, Luleå University of Technology and Linköping University as the partners.

This report is the result of our master's thesis work carried out at Semcon AB, Theres Svenssons Gata 15. The thesis also fulfils the partial requirement for the master's degree in "Sound and Vibration" at Chalmers University of Technology, Göteborg, Sweden.

Acknowledgements

First and foremost, we would like to extend a word of gratitude to Jonas Klein from A2Zound, for giving us this immense opportunity to start the thesis at Semcon and for his constant words of motivation during the work. Secondly, we are greatly indebted to our supervisor Peter Mohlin, from A2Zound, who's constant guidance and technical expertise have been invaluable in the successful completion of this thesis work. We are very grateful for his help and suggestions in designing auditory icons for the chosen driving scenarios especially for the second evaluation and also the development of the MATLAB GUI for the individual listening test. Thirdly, we owe many thanks to our supervisor from Chalmers, Penny Bergman, for her contribution to this work. Her precious psychoacoustics expertise helped us improve the evaluation methods for testing our designed sounds.

We would also like to say thank-you to the people from the Department of Sound and Vibration in Chalmers, who have supported and helped us a lot since the commencement of the 2-year master's programme, special thanks to Erkin Asutay and Börje Wijk for their help during the thesis.

We would also like to thank all the people from A2Zound and test center for providing us with a supportive working environment during the thesis, our friends and well wishers who participated in the jury group evaluations and the final listening test.

Most importantly, we would like to thank our respective families for their constant love and support.

Vijendra Bhat and Yidan Lai

Contents

List of abbreviations

3D	Three-Dimensional
HMI	Human-Machine Interface
IVIS	In Vehicle Information Systems
ADAS	Advanced Driver Assistance Systems
HRTF	Head Related Transfer Function
ITD	Inter-aural Time Difference
IID	Inter-aural Intensity Difference
ITS	Intelligent Transport Systems
BLIS	Blind Spot Information System
FCW	Forward Collision Warning
NVS	Night Vision System
LDW	Lane Departure Warning
DAS	Driver Alert System
ACC	Adaptive Cruise Control
TMS	Tyre Monitoring System
SAM	Self Assessment Manikin
ATIS	Advanced Traveller Information Systems
MIDI	Musical Instrument Digital Interface
GUI	Graphical User Interface

1 Introduction

1.1 Background

Over the past few decades, the importance of safety has become a growing concern within the automotive industry. Advanced Driver Assistance Systems (ADASs) that are proliferating in modern automotive world are known to greatly enhance the driver information (Lindgren *et al.*, 2008). These systems are known to intelligently and interactively tackle different driving scenarios and inform the driver about several events and/or hazards, for example - an imminent collision; blind spot monitoring; lane monitoring; reverse/parking aid description and so on. Apart from events in the external driving environment, the driver has to simultaneously cope with the in-car distractions. Sources of in-vehicle distractions may include manipulating satellite navigation systems or climate control systems, entertainment systems, mobile phones or reaching out for an object (Nevile and Haddington, 2010). Hence, managing with the tasks collectively can distract the driver from his primary task and also put cognitive load on the driver (Dingus and Hulse, 1993). A study about accident data conducted in the USA claimed that it is not just the distracting behaviours that caused accidents, but also the driver characteristics, such as age, that contributed towards it (Stutts and Hunter, 2003). Hence, it has clearly been a challenge to provide the driver with information in a manner such that the driver could readily comprehend it, while remaining undistracted.

At present, vehicles are equipped with advanced telematics technologies for the purpose of improving driver safety/entertainment. Many different modalities are used as a means of presenting in-vehicle information about a potential threat at an early stage to a distracted driver. The three most commonly used modalities are - visual, tactile and auditory. The modality of interest in this project is the auditory modality.

Auditory warnings play a vital role in attracting attention and conveying given information in industries including aviation, automotive, construction etc (Edworthy and Stanton, 1995). In the automotive/aviation world, auditory warning alerts are particularly appealing because they help to reduce the distraction of the driver/pilot from his primary task (driving/flying) and hence enhance the safety (Edworthy, 1994). The auditory modality is generally advantageous for warning signals and a driver's response to situations of danger can be significantly reduced by the use of appropriate early warnings (Edworthy and Haas, 2008; Lee *et al.*, 2002). What is 'appropriate' is decided by the meanings these alerts convey to the driver about a particular event without excessively startling him (Bellotti *et al.*, 2002).

1.2 Problem description

The most common way of alerting the driver has been by the use of earcons. Earcons are brief and distinct musical sounds (e.g. beeps) that convey information about a certain event (Blattner *et al.*, 1989). They are able to convey information instantly, and they may not be as susceptible to information masking in speech-rich environments (Parker *et al.*, 2008). Different parameters like rhythm, pitch, timbre etc are modified to create different messages (Trentecoste, 2004). Studies have shown that the use of a discontinuous sound allows a faster perception and localization of objects than a continuous sound (Bellotti *et al.*, 2002). However, there are limitations. The most important one being that the meanings they represent are not intuitive, they have to be learnt by the user (Trentecoste, 2004).

Using auditory icons has been another means of alerting the driver. The biggest advantage is that they don't have to be learnt, auditory icons contain the inherent meaning of the sound it is representing (e.g. sound of a crash to indicate a collision warning). On the whole, studies have shown that the auditory stimuli are being increasingly preferred in modern human-computer interfaces in vehicles since they lessen the burden for the visual system (Brewster, 2002).

Both ways of conveying information to the driver, be it by intuitive or by non-intuitive sounds, have their common disadvantage - they, by themselves, cannot help the driver locate the direction of the event. Therefore, the use of 3D sounds would help enhance the possibilities of increasing the information content conveyed to the driver about the environment outside the field of view.

1.3 Purpose of the thesis

The purpose of this project was to carry out a study about the effects of using spatialized sounds for in-vehicle use as auditory information signals, an area of research which is still in an early stage of development. 3D sound systems are already being implemented in First Person Shooter PC games like *Half-Life 2* (Valve Software 2004) which aimed at subjecting the gamer to an immersive environment, thus facilitating a much more exciting and realistic gaming experience (Grimshaw *et al.*, 2008).

> "immersion means becoming physically or virtually a part of the experience itself."
>
> (Ermi and Mäyrä, 2005)

3D sound reproduction can be achieved either by the use of loudspeakers with cross-talk cancellation or by the use of headphones. The use of loudspeakers is known to reduce (but not eliminate) the front-back confusion[1] (Lundkvist *et al.*, 2011) which is a common disadvantage by the use of the headphones with generalized HRTFs (Head Related Transfer Function)(Blauert, 1983; Weinrich, 1982). However, in a noisy environment, loudspeakers cannot provide clean binaural signals to the listener's ears the way headphones do, since headphones are unaffected by room acoustics and cross-talk signals (Gehring, 1997; 1992). So, it is clear that

[1]**Front-back confusion** is the result of the human auditory system finding it difficult to correctly distinguish what sound source is in the front of the head and what is in the rear, due to the identical IID and ITD cues for the two sources. The same is also true for up-down confusion and all the cues can be imagined to be a part of a cone which is often referred to as the *Cone of Confusion*(See Chapter 4)

3D sounds should be reproduced in a good manner, but what is also very important is having an approach that helps design and evaluate these sounds. This thesis is an effort to develop a design and evaluation methodology that can be used to test the acceptance of 3D sounds in in-vehicle auditory information systems.

1.4 Thesis goals and tasks

The sound sketching method is assumed to play a vital role in early conceptualisation phases of product sound design. This method is known to strengthen the process of determining the sounds that show the potential of it being usable and functional for a given set of objectives, even before the first prototypes are manufactured. In early conceptualisation phases, simple sketches called thumbnail sketches are created. These sketches serve as a basis for expressing the designers ideas, ideally to a product development team or a group of potential users, and evaluating the possibilities of it being used in the later stages of product development (Nykänen, 2008).

The main scope of this master's thesis is to develop and evaluate a methodology for designing 3D auditory signals for in-vehicle information/warning systems and to test the sound sketching method. Hence, the study broadly aims to investigate the appropriateness of virtually placing sound sources in a vehicle environment i.e. using spatialized warning sounds in relation with non - spatialized warning sounds which are currently used. The principle goals and tasks involved during this master's thesis are :

- Carrying out a literature survey on 3D sounds, auditory warnings, intelligent transport systems, listening tests etc.
- Sketching earcons and auditory icons using a digital audio mixing software, for applicability in different driving events/scenarios (like lane departure warnings, blind spot detection).
- Using a 3D binaural audio plug-in and headphones to create and listen to 3D sounds for the chosen driving context.
- Evaluating these sounds on several jury groups and obtaining feedback from the test subjects with regard to suitability for use in-vehicle auditory systems.
- Performing several iterations of the sound sketches according to the feedback obtained during evaluations, until desirable sounds are designed.
- Testing the designed sound sketches by conducting an individual listening test.

1.5 Delimitations

This thesis is limited by the given binaural plug-in's generic HRTF filters that is used to synthesize 3D sounds. If all the listeners' HRTFs are individually measured, one could greatly improve the binaural spatial hearing experience and possibly eliminate the commonly encountered front-back confusion caused by using generic HRTFs. This is because HRTFs captures the ITDs (Inter-aural Time Difference), IIDs (Inter-aural Intensity Difference) and the spectral coloration produced by the sound's interaction with an individual's pinnae, which are essential for localizing sounds. These measurements, however, are painstaking since it has to be measured for each listener and at every 15° or 30° around the listener (Matsumura *et al.*, 2005). Hence, it was quite hard to implement in the short duration of this project.

The stimuli used in this study are limited to non-verbal auditory signals.
The investigations in this thesis are limited to only 2 driving scenarios (blind spot and lane departure).

1.6 Thesis structure

This report consists of 8 chapters in total.

- The 1st chapter introduces the background and the purpose of this project. The delimitations are also stated.
- The 2nd chapter provides some theory about Intelligent Transport Systems (ITSs). It briefly explains the 2 different kinds - ADASs and In-Vehicle Information Systems (IVISs) and some examples for each kind.
- The 3rd chapter describes some important theory pertaining to auditory signals. More focus is given on earcons and auditory icons and some basic guidelines of how they could be designed.
- The 4th chapter gives an overview of 3D sounds. Different aspects of 3D sounds such as binaural difference cues, spectral cues and HRTFs are explained.
- The 5th chapter explains the methodology adopted for the thesis. It deals with the sound sketching method that was used for the sound design process. Also, the implementation of the jury group evaluations and the final listening test is explained.
- The 6th chapter presents and discusses the results from the jury group evaluations and the final listening test. In addition, the improvements of the sound sketches between evaluations have been summarized.
- The 7th chapter focuses on the discussion of the design methodology from the obtained results and some conclusions of the method is drawn.
- The 8th chapter discusses some future work for the thesis.
- The 9th chapter presents the appendices. It presents some detailed individual participants' results obtained from the evaluations and the final listening test. The questionnaires for the 2 evaluations are also presented.

2 Intelligent Transport Systems

Intelligent Transport Systems (ITS) may be defined as systems utilizing a combination of computer, communication, positioning and automation technologies to use available data to improve the safety, management and efficiency of terrestrial transport and to reduce environmental impact[1]. ITSs aim to tackle the problems of road safety and congestion. Different systems are currently proposed and most of them are already technically feasible. Various systems exist that warn and or/automatically intervene to assist the driver in case of any hazards. However, these systems should be carefully implemented to avoid giving the driver too much or too little to do.[2]

2.1 Advanced Driver Assistance Systems (ADASs)

In an attempt to mitigate the amount of accidents and reduce eventualities, vehicles today are being equipped with Advanced Driver Assistance Systems. ADASs are electronic systems that are designed to support the driver in their driving task and help them manoeuvre through demanding traffic situations. This support ranges from the presentation of simple information to advanced assistance to the driver which might involve taking over the driver's tasks in critical situations, thereby, making him aware of the situation and also aiding in taking the pressure off him during these situations (Stevens *et al.*, 2002).

ADASs can offer support to the driver on four different levels. At the most basic level, the driver is presented with information which enables them to make more informed driving decisions, for example, information about non-visible pedestrians, that the driver encounters during night driving. At the next level, ADASs can give the driver suitable warnings of an imminent and possibly dangerous situation to give them more time to decide and react. The third level of intervention is when the system is involved in not only warning the driver but also advising or guiding them through a given situation. At the highest level of intervention, is it possible for ADASs to either take action independently or override the action of the driver. Regardless of the level of intervention, manufacturers who implement these systems hope to assist the driver before a critical situation arises or atleast, in an attempt to reduce the consequences of a driving error, hence improving the driving safety (Lindgren and Chen, 2007; Lindgren *et al.*, 2008).

[1]See ITU-R (2006) "Intelligent Transport Systems: Handbook on Land Mobile Vol 4", 109 pp, Available at:`www.itu.int/pub/R-HDB-49-2006/en`

[2]*Intelligent Transport Systems*, The Parliamentary Office of Science and Technology, Available at `http://www.parliament.uk/documents/post/postpn322.pdf`. Last accessed on 2012-06-28.

2.1.1 Some of the currently used ADASs in vehicles

As mentioned before, there is a growing need to have better, more intelligent warning systems equipped in modern vehicles. Some of the commonly ADASs in today's automotive industry are:

1. **Blind Spot Information System (BLIS)**
   The BLIS is a sensor or camera-based system used to provide the driver with information about vehicles, pedestrians or cyclists in areas in the blind spot of the car, not visible in the rear-view mirrors. Since the past several years, it has been made available in some new vehicles in Europe but it is on an early stage of deployment.

2. **Forward Collision Warning (FCW)**
   The FCW also uses sensor technology (laser or microwave radars) that measures distance, relative speed and angular position of the car and obstacles in front of the car. If an obstacle is detected, the system decides whether or not the vehicle is in imminent danger of crashing. If there is a risk of crashing, the system alerts the driver with a warning.

3. **Night vision system (NVS)**
   Night vision systems incorporate cameras and near or far infrared lights to flood the environment in front of the vehicle. Thermal readings of the environment are then used to create a visual representation of possible obstacles and present this to the driver with the purpose of improving vision at night and in bad weather.This improved system will make it possible for the driver to see pedestrians outside the headlight beam or obstacles concealed by the light of oncoming traffic.

4. **Lane Departure Warning (LDW)**
   The LDW is a camera based system which recognizes lane markings and is activated when the vehicle is unintentionally drifting away from a lane, that is, when the driver is about to leave a lane without engaging the turn signal. As a new means of alerting the driver for lane departure, haptic information is also gaining prominence. The system lets the driver feel the vibration in the steering wheel so he/she can react accordingly.

5. **Driver Alert System (DAS)**
   DAS use vehicle sensors or in-vehicle cameras to monitor driver behaviour and provide the driver with a warning, for example when the vehicle's lateral position changes in an irrational pattern or when eyelid movements lessen.

6. **Adaptive Cruise Control (ACC)**
   The ACC automatically adjusts the vehicle's speed using radar technology, to maintain a safe distance from a vehicle ahead in the same lane[Siemens 2005]. If the vehicle in front slows down or accelerates, the system either makes the car decelerate or accelerate to maintain the preset distance. The ACC is expected to ensure that there is a certain specified distance between the two cars, even if that driver unexpectedly reduces speed.

7. **Tyre Monitoring System (TMS)**
   The TMS measures a wheel's rotational speed relative to the other wheels which allows the system to detect dangerously low air pressure in the tyres. If there is a critical air pressure loss in any of the tyres, the system notifies the drivers.

2.2 In-vehicle Information Systems (IVISs)

Systems that are incorporated into a vehicle environment, under ITS, with the sole goal of informing the driver about an event, are called In - vehicle Information Systems (IVISs), also known as Advanced Traveller Information Systems (ATIS) (Wilschut, 2009). These systems only provide information to the driver about an event and they are not aimed at taking over the driving task. They can be described as systems that provide relevant real-time in-vehicle information about different aspects of driving like the environment, a nearby vehicle or the state of the driver (Wilschut, 2009). The tremendous advancements in technology applied to gathering and processing traffic information, and parallel advancements in the field of communications have provided new opportunities to the driver to receive information, about the external environment. In-vehicle driver information is one example of the application of information technology and telecommunications in the transport field, more generally known as Advanced Transport Telematics (ATT) or Intelligent Transport Systems (ITS) (Stevens, 2000).

2.2.1 Examples of IVISs

The most commonly implemented IVISs are route guidance, navigation systems, hazard warning and sign information systems (Wilschut, 2009). They include in-vehicle radio traffic reports, from general broadcasters, on queues and accidents; and highway advisory radio (HAR) 'localcasts' on conditions in known congested locations. Recent developments in European countries include the TMC (Traffic Message Channel) system being incorporated within the RDS (Radio Data System). As a further refinement, European companies including iTIS and Trafficmaster (UK) and Mediamobile (France) have developed congestion monitoring and journey time information services that relay information to in-vehicle units.[3]

[3] Chapter 1: What are Intelligent Transport Systems, *ITS Handbook*, 1-2. Available at: `www.itu.int/pub/R-HDB-49-2006/en`. Last accessed on 2012-06-28.

3 Literature on auditory signals

Auditory alarms and warnings have shown their presence as early as the industrial revolution. Mechanical alarms such as klaxon signals and horns used during those times aimed at informing the listener of conditions not directly visible to the naked eye, such as high steam pressure or low oil pressure in a factory process. Modern warnings are omnipresent, found in everything from watches to stoves and microwaves, beeping to warn listeners that time has elapsed or to indicate that the food has finished cooking. Over the years, researchers and practitioners have found that auditory warnings can function as supplements to visual signals or be useful themselves. Studies have confirmed that the auditory modality is more advantageous for warning signals owing to the fact that hearing is omnidirectional and cannot be voluntarily shut off, whereas visual warnings can be voluntarily shut off and must be seen in order to warn the subject (Edworthy and Haas, 2008).

> “the fact that nature did not provide us with earlids is probably due to ... the use of the acoustical channel for warning signal function to which it is exceptionally well matched.”
>
> (Durlach and Colburn, 1978)

The purpose of auditory signals is to draw the driver’s attention to a potential threat. A signal could also be used to draw the attention to a visual warning display. Auditory warnings are able to convey different levels of the subjective impression of urgency. The urgency level should be consistent with the urgency of the potential threat. Higher levels of urgency should indicate more dangerous threats (Durlach and Colburn, 1978). To convey different levels of urgency, one could also for example, vary the speed of the signal, the rhythm, the pulse duration, the intensity and the frequency (Edworthy *et al.*, 1991). One very important issue to consider when using auditory signals is the trade-off between a warning alerting the driver and annoying him/her (Campbell *et al.*, 2007).

3.1 Types of auditory warnings

Auditory warnings are being used in many different environments nowadays. A lot of research work is being done to categorize warning signals to define where and how they can be used. Some studies categorized auditory warnings as incidental or intentional (Wilkins and Martin, 1980; 1987).

> “Incidental or unintentional warnings are not deliberately generated, but arise in conjunction with a dangerous event or process.”
>
> (Wilkins and Martin, 1980; 1987)

This meant that warnings had to occur during a specific context, for example, sounds made by a malfunctioning piece of machinery or an approaching train. He also had opinions that intentional warnings are generated by devices built for the specific purpose of producing warning sounds. Intentional warnings can be categorized as speech or non-speech signals. Another study argued that the unintentional warnings are not warnings at all because it is an artifact intended to represent the status of an event (Edworthy and Adams, 1996). It was also mentioned later in a study about the main features of an auditory warning that it is intended to signal when something significant happens and inform of something critical (Edworthy and Haas, 2008).

It had been previously argued that an alarm can be thought of as a generic term for all forms of sound that attract attention (Edworthy, 1994). This would include animal calls, burglar alarms or car alarms among other sounds. Auditory warnings, however, are sounds that are designed specifically to attract attention but also to provide additional information (Edworthy and Haas, 2008). Speech and non-speech sounds as auditory warnings are discussed below.

3.1.1 Verbal auditory signals

Speech warnings are the sounds which are recorded, digitally or analogously or synthesized from human speech. These warnings are redundant, presenting a repetition of the warning information, which can reduce the confusion to the listeners. The advantage of speech is that it includes lots of information and the detail of the situation in the environment. When listeners need to identify the sound source without training, speech is preferred, for example, a speech warning can tell a driver about the exact distance of the car from an obstacle. Another strength of speech warnings is that when the information need to be a rapid two-way exchange or when the warning message deals with a future time requiring some preparation (like a missile countdown). However, one of the disadvantages of speech warning is that the intelligibility of the sound needs to be taken into consideration. Reproducing speech in a noisy environment is associated with a number of problems in intelligibility. Since the speech warnings take a long while to be enunciated, the time for listener to react and get the information might be longer than the non-speech sound (Edworthy and Haas, 2008). Another downside with speech warning is that it has to be implemented with different languages in different countries.

3.1.2 Non-Verbal auditory signals

Non-speech signals consist of tones or other non-speech elements produced by modern devices and consists of just about anything that can be used to signal events to the user. There are two central requirements of an ergonomic non-verbal auditory warning. First, it needs to have suitable loudness. The sound shouldn't be too loud but loud enough to inform the driver. Secondly, it needs to be psychologically appropriate in some way (Edworthy and Haas, 2008).

There are two types of non-verbal warning sounds: *Earcons* and *Auditory Icons*. A fundamental difference between auditory icons and earcons is that earcons can be considered to have arbitrary symbolic representations while auditory icons can be regarded as analogical representations of the events they represent (Edworthy, 1994).

3.2 Earcons

Earcons use abstract simple musical tones to present information to the listener. There are five parameters that can be modified to express different information in earcons, which are rhythm, pitch, timbre, register, and dynamics (Trentecoste, 2004). Earcons can have several definitions:

> “Non-verbal audio messages used in the user-computer interface to provide information to the user about some computer object, operation, or interaction. ”
>
> (Blattner *et al.*, 1989)

> “Abstract, synthetic tones that can be used in structured combinations to create auditory messages.”
>
> (Brewster, 1994)

Earcons are constructed from short building blocks called motives. Rhythm and pitch are the fixed parameters of a motive, while timbre, register, and dynamics are the variable parameters of motives (Blattner *et al.*, 1989)

> “A motive is a brief succession of pitches arranged in such a way as to produce a tonal pattern sufficiently distinct to allow it to function as an individual recognisable entity.”
>
> (Blattner *et al.*, 1989)

The following sections gives a general walk-through of different categories of earcons and how to create them.

3.2.1 Initial earcon research

A few research studies have proposed different ways in which motives can be manipulated to form families of earcons: one-element earcons, compound earcons, transformational earcons and hierarchical earcons (Blattner *et al.*, 1989; McGookin and Brewster, 2011). They can be classified into 4 categories:

- *One-element earcons* are the simplest type and can be used to communicate a single parameter of information.
- *Compound earcons* are formed by joining one-element earcons, or any other form, together to create more meaningful messages.
- *Transformational earcons* families are constructed around a ‘grammar’ or set of rules, where there exists a consistent set of structured symbolic mappings from individual data parameters to individual sound attributes.
- *Hierarchical earcons* are constructed around a set of rules. Different parameters of sound (E.g. timbre, register, intensity, pitch and rhythm) can be manipulated around these rules.

3.2.2 Creating earcons

The effective application of earcons is mostly dependent on the design of earcons. The design guidelines should be followed to create effective earcons. There have been many guidelines proposed as to how earcons should be designed, validated and improved through empirical research studies (Blattner *et al.*, 1989; Brewster *et al.*, 1995). Brewster *et al.* (1995) aimed at designing two families of compound earcons: one representing files in a graphical user interface and another representing operations that could be performed on those files. Three different designs of the two earcon sets were compared. One design was based on the guidelines of Blattner *et al.* (1989) (simple set), using rhythm, pitch structure, register and sinusoidal timbres to encode information about the files and operations. Another design was based on the musical set, but with some modifications, notably the use of musical timbres rather than sinusoidal tones, as these were considered to improve performance. The third design (control set) avoided the use of rhythm to encode information. Instead, only musical timbre and pitch were used. This was considered to be similar to the simple beeps that computer systems made at the time (McGookin and Brewster, 2011).

The earcons were tested by a group of participants and the results showed that the identification of the application family improved significantly when represented using a musical timbre than when the sinusoidal waveforms suggested by Blattner *et al.* (1989) were used. They also found that the identification of the simple control sounds deteriorated for file types, which was represented by pitch rather than rhythm (McGookin and Brewster, 2011).

Based on these results, the musical set of earcons were redesigned to incorporate greater differences between the rhythms used (Brewster *et al.*, 1995). The earcons were evaluated and the identification between the original and revised musical earcon sets showed improvement. Updated general guidelines were later suggested for the effective design of earcons which can be summarized in table 3.1 based on some important auditory attributes (Brewster *et al.*, 1995; Patterson *et al.*, 2010; Bregman, 1990; Trentecoste, 2004).

However, these technical design guidelines are futile if one doesn't take into account the brand identity aspects that are unique to each brand. A set of attributes concerning the brand's core values have to be considered and are vital during a design process, since it is these attributes that build up a brand's personality and aim at communicating with the prospective customers.

3.2.3 Learning and understanding earcons

Earcons are easy to create and they do not have to correspond to the particular object. But the main disadvantage of earcons is that they need to be learned, because the meaning of the simple tones is not inherent (Trentecoste, 2004). It is important to consider how listeners will learn earcons and the data mappings contained within them. Earcons have abstract mappings between sound and what that sound represents, so the training is necessary for the listeners.

In a research study, "intuitiveness" in understanding earcons was compared to

Table 3.1: *General technical design guidelines for creating earcons.*

Auditory Attribute	**Guidelines on use**
Timbre	Using timbres that resemble musical instruments instead of the sinusoidal tones as proposed earlier by Blattner *et al.*, 1989. The timbres of different earcons should come from different musical families (e.g.,brass and organ) so one can subjectively distinguish between two different sounds.
Register	Register can be varied between *high*, *medium* and *low*. Within a family of instruments, the different members are classified based on this quality. Higher register indicates higher pitch, for example violins have higher register than cellos even though they are from the same instrument family. This also means that the perceived timbres vary with change in register. Large differences (of two or three octaves) should be used between the registers used.
Pitch	The use of pitch as the sole means of encoding information and to distinguish between tones should be avoided, unless there is a significant difference between the tones. Suggested pitch ranges are 5 kHz (maximum) and 125 Hz–150 Hz (minimum) to ensure that the sounds are not easily masked and are within the hearing range of the listeners.
Rhythm	Putting a different number of notes in each rhythm is an effective way of distinguishing sounds. By introducing a gap (of at least 0.1 seconds) between two earcons, when played consecutively, one could make sure that they are identified as two, rather than one earcon. However, if in a rhythm, a sequence of high notes of different pitches interleaved with low ones are presented to the listener, there could be a profound effect on the way they are perceived by the listener. One could segregate co-occurring sounds into distinct auditory streams and this could influence the perceived pitch, timbre or loudness of the resulting sounds.
Intensity	Intensity should not be used on its own to differentiate earcons, since listeners are poor at making judgements in loudness. Some suggested ranges for intensity are 20 dB (maximum) above hearing threshold and 10 dB (minimum) above hearing threshold.

that of auditory icons (Garzonis *et al.*, 2009). Auditory icons had high recognition rates compared to earcons. The design of earcons can also influence their learnability. The rules by which earcons are structured can be used as a means of differentiating earcons, hence making it easier for the user to learn. If no training is provided, identification of earcons could be poor, making the listeners feel frustrated and annoyed that all sounds are similar. However, training does not need to be extensive, even a short duration of 5-10 minutes, could help users to achieve good levels of identification (McGookin and Brewster, 2011).

3.3 Auditory icons

Auditory icons are sounds from the everyday environment that help users understand what kind of information they are dealing with. They are caricatures of naturally occurring sounds such as bumps, scrapes, or even files hitting mailboxes (Gaver, 1986). The mappings between data and their auditory representation can be divided into three different types (Blattner *et al.*, 1989): symbolic, iconic, and metaphorical :

- *Iconic auditory icons* use the sound almost alike the object or the action they represent (e.g. sound of a car crash to indicate a collision warning).

- *Metaphorical icons* means they sound like some element of the action they represent. (e.g. sound of a liquid being poured into a container to indicate a *copy* operation in a computer)

- *Symbolic icons* rely on social conventions to symbolize meaning (e.g. sound of a siren to indicate an ambulance approaching).

Auditory icons require an existing relationship between the sound and its meaning, something that may not always exist or the sound is not very typical. In such cases, it may be better to use earcons. Auditory icons and earcons are complementary in an auditory display; both may be useful in the same situations, but with different advantages and disadvantages. Both earcons and auditory icons seek to fill the same role in an auditory display and considerable effort has been made to compare and contrast them over the years. It has largely been found that the advantages of earcons are the disadvantages of auditory icons and vice versa. An understanding of the similarities and differences between these two forms of auditory cue is important in understanding when to use them (Gaver, 1986; Blattner *et al.*, 1989; Garzonis *et al.*, 2009).

Auditory icons have been shown to be more efficient than earcons with good identification performance in some specific urgent situations where quick and accurate recognition of a message is necessary (E.g. for forward collision warning) (Larsson *et al.*, 2009). A carefully designed set of auditory icons requires little training (Garzonis *et al.*, 2009). However, it will be hard to identify suitable sounds for all situations, and thus it becomes difficult to create a good set of auditory icons. It is possible to parameterize auditory icons in a similar way to earcons, but in practice this can be difficult to control a virtual source of a sound along relevant dimensions using digital sampling techniques. (Gaver, 1993).

Although the use of auditory icons may be seen to be the better choice, there is evidence that in some cases earcons may be more appropriate as suggested by Sikora *et al.*, 1995. A study was conducted within the context of a graphical user interface business communications simulation, for which many musical and real world sounds were designed. These sounds were designed to provide auditory signals for specific functions in the graphical user interface and the user's task was to map the sound to it's functions. The sounds, both environmental and musical, that were mapped well to the functions, were compared in terms of agreement of function, appropriateness and pleasantness. They found that the musical sounds were rated as more appropriate and pleasant than the real world sounds. The real world sounds were consistently rated poor in terms in pleasantness and appropriateness and were seldom preferred as the sound for the given function. In this study, they found that the type of auditory feedback preferred by the subjects were not the same as the sounds that were believed to provide the best mapping to the functions, rather the abstract musical sounds (Sikora *et al.*, 1995; McGookin and Brewster, 2011).

3.3.1 Designing auditory icons

Auditory icons can be created in different ways. The simplest method is to record the particular everyday sound and store it as a sound file for use in the application where it has to be used. There are some aspects of generating auditory icons. It is possible to create and use *parametric auditory icons*, e.g. changing sizes of objects, or the rate of pouring a liquid into a container (Brazil and Fernström, 2011). Simple forms of parameterization includes changing the loudness, changing the pitch or low pass filtering of sounds, or varying the playback rate. A term called *cartoonification* has been discussed in some studies which essentially means improving the recognition of the auditory icons by exaggerating the features that distinguish it from the real world sounds. Therefore, the sound models would be made more computationally efficient than fully realistic models, so they could be used in portable devices like mobiles (Brazil and Fernström, 2004).

The recognition problem when choosing auditory icons for the interface can also be discussed:

> "It is an art with many hidden dangers and dependent upon the skills of the designer."
>
> (Mynatt, 1994)

A set of guidelines for designing auditory icons can be based on four factors: identifiability, conceptual mapping, physical parameters, and user preference that influence the usability of auditory icons (Mynatt, 1994).

In addition, another study (Brazil and Fernström, 2011) summarized some steps in choosing sounds to create auditory icons:

1. Choose short sounds that have a wide bandwidth and where their length, intensity, and sampling quality can be controlled. The set of sounds chosen should represent the variety and meaning needed for the anticipated design space.

2. Free-form answers should be used to evaluate the identifiability of the auditory cues and the learnability of those auditory cues that are not readily identified, should also be evaluated.

3. Evaluate the possible sets of auditory icons that show potential problems with masking, discriminability and conflicting mappings.

4. Conduct usability experiments with interfaces using to test the resulting auditory icons.

These steps, as mentioned before, are meaningless if the brand identity aspects are not considered.

4 Overview of 3D Sounds

Sound provides an important channel of feedback that either can be helpfully redundant to a visual cue, or can provide feedback for actions or situations that are out of the field of view of the listener (Grosse, 2009; Gehring, 1997). If one wants to improve the quality and the ease of interaction within a human machine interface of a communication system, or office or a home environment, one should not ignore the sense of hearing. To control the audio sensations of a human user in a positive way requires examining all aspects of human sensitivity to sound (Begault, 1994). One of the bigger advantages of 3D sound compared to vision is the ease of synthesis of multiple virtual sound sources anywhere in the 360° space around the listener. Visual display involves eye or head movement to focus on the sound sources, but with an audio display, on the other hand, one can switch the focus of attention between virtual sound sources at will.

Several efforts are already being made to utilize 3D sounds in military applications, for example, the Australian Defence Science and Technology Organisation said that "it has been working on ways to enhance the performance of aircrew in combat situations by giving them sound cues, that help to rapidly detect the direction and types of threats and targets. Synthesized 3D sound, which mimics the sounds of the real world is developed from recordings taken from within the ears of individual crew members and then reproduced digitally to provide the critical directional information. 3D sound technology will reduce the high visual workloads experienced by the aircrew and provide them with more time to respond to threats or targets, giving them an operational advantage and a safety edge."[1]

In the next section, we will put forth the theoretical foundations of 3D sounds, that is, binaural hearing and spectral cues and the principles of localization of sound.

4.1 Perceptual basis of 3D audio

3D sounds help in gathering perceptual information about the location of different events happening around a listener. A study briefly stated the meaning of perception.

> "As human beings we gather information about the state of the world around us in order to understand what is happening in our world and to help guide our actions. We use our sensory organs (i.e. our eyes, ears, etc) to gather this information, and the process of interpreting the information we gather is called perception."
>
> (Parker *et al.*, 2008)

In order to create an appropriate perceptual experience and to provide a complete picture of the world around a listener, one requires to understand the requirements of a 3D audio system which replicates the information and simultaneously helps in

[1] Sixty Years of Aeronautical Research in Australia: Research Overview.Australian Defence Science and Technology Organisation, March 2003. Available at `http://www.dsto.defence.gov.au/`.

balancing the sensory information (Begault, 1994; Parker *et al.*, 2008).

The underlying design philosophy of 3-D audio lies in the fact that one should recreate this information in exactly the same form as if a real sound were present in the external environment. To this, the brain would therefore respond in exactly the same manner and perceived when a sound is heard out in the real environment (Parker *et al.*, 2008).

In the following subsection, we will throw some light on binaural difference cues and spectral cues that influence the perception of 3D sounds.

4.1.1 Binaural Difference Cues

'Binaural' as the name suggests, deals with the double auditory channel, that is, the two ears of a listener. When a listener is subjected to spatialized sounds, these sounds are processed by the auditory system that integrates information from the two ears into a single sound image in space (Begault, 1994). In psychological studies, the reference point for describing angles and distances of sound images is considered to be approximately at the origin point between the two ears, at the eye level in the center of the head. The two general conventions used to describe the angular perception of a virtual sound source are *azimuth* and *elevation*. Azimuth refers to angle of a sound source on a plane parallel to he ground surface. Azimuth can be measured from 0 to 180 degrees counter-clockwise (left) or from 0 to 180 degrees clockwise (right). On the other hand, elevation refers to the angle of the sound source from the head on any other plane other than the horizontal plane. It can be measured from 0 to 90 degrees **up** or from 0 to 90 degrees **down**. 0 degrees azimuth and elevation refers to a source source positioned at the origin point. In order to obtain a more complete description of a sound sources' position in space, the attribute *perceived distance* should be taken into account (Begault, 1994). Figure 4.1 gives an idea of the parameters that are used to describe the localization of 3D sounds as single sound images in space.

Binaural difference occurs because human beings have two ears. When a sound is played on one side of the head, the perceived intensity of that sound on the other ear is reduced due to the sound attenuation caused by the head. Hence the closest ear tends to perceive the sound louder than the furthest ear (Cheng and Wakefield, 2001). The extent to which this intensity difference occurs depends on the location of the sound source, which means, if the sound source is placed directly in front of the listener, the difference is significantly lesser and the sound is perceived to be at the same intensity level on both ears, than if the sound source is placed on either side of the listener's head. The same is also true if the sound is directly behind the listener. The term used to describe this difference in perceived sound intensity level between the two ears is Inter-aural Intensity Difference (IID or also sometimes called Inter-aural Level Difference, ILD) (see figure 4.2).

Another term, ITD (Inter-aural Time Difference), is used to describe the difference in time taken for the sound to arrive at each of the right and left ears of the listener. This occurs just like the IID, that is, ITD is not evident when the sound source is placed directly in front or behind the listener. But when the sound

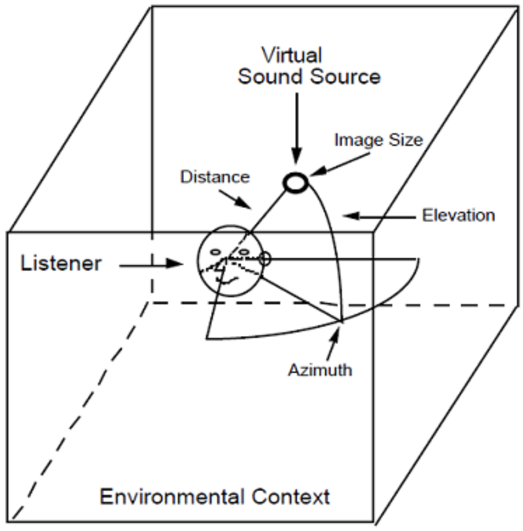


Figure 4.1: *Different parameters used to describe a virtual sound source in space. Adapted from Begault (1994).*

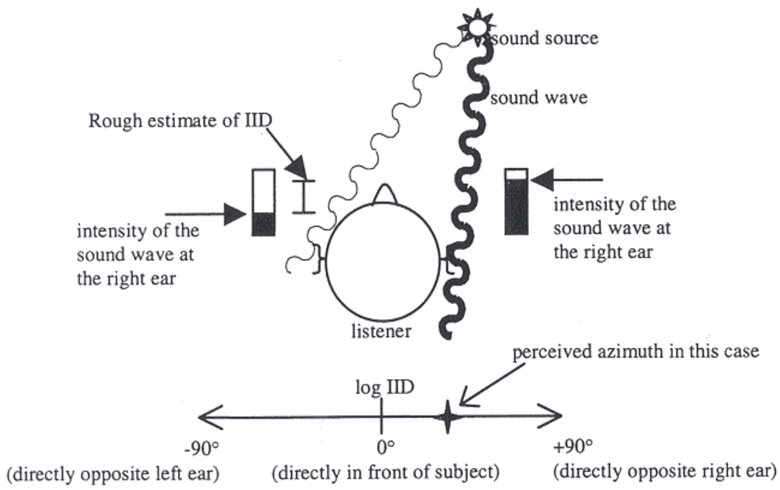


Figure 4.2: *Change of Inter-aural Intensity Difference (IID) with location. Sound is perceived to be at a higher intensity level on the right ear than the left ear. Adapted from Cheng and Wakefield, 2001*

is placed on one side of the head of the listener, the furthest ear experiences a time delay in the arrival of sound (See figure 4.3).

Theoretically, neither the IIDs nor the ITDs can be completely identical, even if a sound is placed directly in front or behind the listener. This is because it depends on the shape of the head. Therefore, the differences can be disregarded only if a spherical head is assumed and if one also disregards the effect of facial features and the pinnae[2] (Begault, 1994). The following subsection describes the effects of taking into account the spectral cues provided by the pinnae.

[2] *Pinna*, plural *Pinnae*, in zoology stands for the part of the ear that projects like a little wing from the head. Available at `http://www.medterms.com/script/main/art.asp?articlekey=4910`, last accessed 2012-05-08

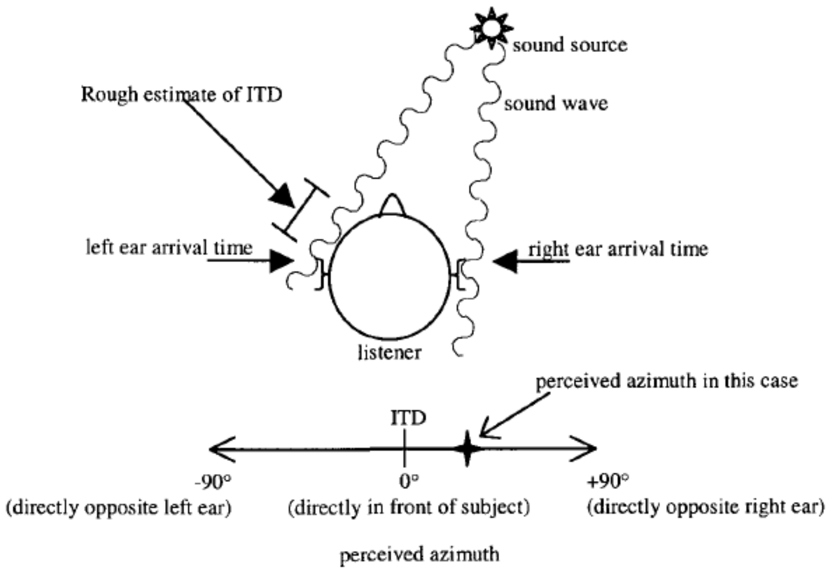


Figure 4.3: *Change of Inter-aural Time Difference (ITD) with location. Sound is perceived to be at a higher intensity level on the right ear than the left ear. Adapted from Cheng and Wakefield, 2001*

4.1.2 Spectral cues provided by the pinnae

In order to provide accurate localization of sounds, only the binaural differences are not enough. There is a distribution of individual binaural difference with respect to more than a single unique location. Since we humans are generally accurate in localizing sound sources, we can assume that there must be additional information available which enables the resolution of the ambiguities of binaural differences. This additional information is referred to as spectral cues (Parker *et al.*, 2008).

For example, if we imagine an axis drawn between the two ears that divides front from back (referred to as the interaural axis), then a sound 15 degrees in front of the interaural axis (Point “a” in 4.4) will have the same pattern of binaural differences as a sound 15 degrees behind the interaural axis (Point “b” in Figure 4.4). This is also true for a sound 15 degrees above the interaural axis (Point “x” in Figure 4.4) and 15 degrees below (Point “y” in Figure 4.4). In fact if the head is assumed to be spherical, then the range of possible locations that are specified by any single binaural difference form a cone with the apex at the centre of the head (See Figure 4.4). This is referred as the ’cone of confusion’ that is used to describe this distribution of separate locations, each providing identical binaural differences. This facilitates calculation of identical IIDs and ITDs at any two opposite points on the surface of the cone for a spherical model of a head. If the only cues we used to localize sounds were binaural differences, then we might expect that we would often hear sounds that originally had to be in front as coming from behind and vice versa (Parker *et al.*, 2008).

As sound travels towards the ears of a listener, it interacts with the body parts such as head, torso and the pinnae. It is this interaction which causes the direct and reflected sound wave front to fall on the listeners ears. This as a result, causes the amplification of certain frequencies (resulting in spectral peaks) more than the

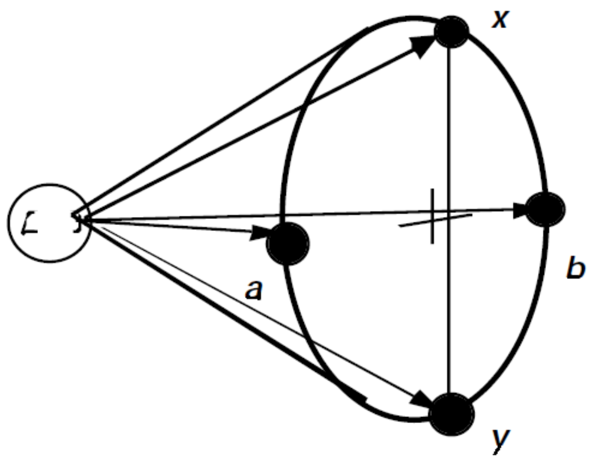


Figure 4.4: *The Cone of Confusion. Here,* ***a-b*** *show front-back ambiguity and* ***x-y*** *shows elevation ambiguity. Adapted from Begault, 1994.*

other, in addition causing attenuation of other frequencies (resulting in spectral notches). Spectral cues are direction dependent. The spectral variation changes with change in source location. It can be seen that the spectrum is very similar below frequencies lower than first spectral peak. This is because the wavelength is higher for lower frequencies and hence the sound waves can propagate around the head of the listener. Therefore, the shadowing of the head has lesser influence at lower frequencies, causing less attenuation, with minor changes in the direction. However, the influence of the head is evident at higher frequencies (0.5 - 1.6 kHz) with change in source location (Begault 1994; Grosse 2009). The HRTF spectral cues can be used to provide information if the sound has been presented in the front or the back, and where it lies in the vertical plane (i.e. elevation). Thus, the spectral cues along with binaural difference cues can be used to provide sufficient information about a unique location of sound in space (Parker *et al.*, 2008).

In the next section, we will continue by explaining about the key component of a 3D sound system, Head Related Transfer Function (HRTF).

4.2 Head Related Transfer Functions (HRTFs)

Many researchers have made efforts to physically model, empirically measure and more recently, computationally simulate the direction dependent frequency response related to ear directly, as a first step towards comprehending directional hearing. These efforts collectively summarize the direction-dependent spectral filtering of free-field sound caused by the interaction of sound with the head, torso and pinna and ear canals and are referred to as HRTF measurements (Begault, 1994; Cheng and Wakefield, 2001). Binaural HRTF refers to both left and right ear HRTFs. By knowing the HRTF, one could, in principle, reconstruct the exact pressure waveforms that would reach the ears of the listener, for any arbitrary source waveform arising from the particular location (Duraiswami *et al.*, 2001). It is interesting to know that the small folds on the pinnae are big enough to cause minute time delays (in the range of 0-300μsec) that cause the spectral content at the eardrum to vary significantly from that at the sound source, for example, as if it were measured by an omni-directional microphone. The underlying theory of

HRTF spectral shaping is that it is the most accurate means to produce a spatial sound cue which is done by transforming the spectrum of a sound source at the ear canals as closely as possible to the way it would be transformed under normal spatial hearing conditions (Begault, 1994). Hence, it is the HRTFs that helps provide the data required to construct filters necessary to produce 3D audio.

4.2.1 Measurement of HRTFs

The usual technique for empirical HRTF measurements is by placing tiny microphones into the subject's ear drums and recording the response at the ears by playing a stimulus signal at a particular azimuth and elevation. The stimulus signal could be either impulses over loudspeakers or *Golay codes*[3](also called *Complementary codes*). Golay codes were used as stimulus signals as a method of measuring HRTFs on a human subject in an anechoic chamber, during a research study at the University of Michigan. These signals were presented from loudspeakers positioned 1.5 m from the subjects' heads and binaural HRTF measurements were made at several different direction of arrival (DOA) - at about 400 different spatial locations to (10 to 15 degrees apart in azimuth and elevation directions)(Duraiswami *et al.*, 2001; Cheng and Wakefield, 2001).

An alternate way of measuring HRTF is by using artificial geometric models of the upper human body. Some example of models used for these measurements are Knowles Electronic Manikin for Acoustic Research (KEMAR), Neumann KU-100, Bruel and Kjaer Type 4100, Head Acoustics, Cortex Electronic Mankin MK1(Parker *et al.*, 2008). These geometric models aim at compensating for the anthropometric differences in fundamental head-torso-ear features that are characteristic to each individual. These measurements results are more 'generic' and the accuracy with which this approach depends on the match between the geometric model and the physical features of the subject/listener under study. One method for generating HRTF sets that are more generic in nature is by taking into account the signal changes caused by the facial features such as head, torso and pinnae by mathematically modelling them, in an attempt to extract those fundamental features that is common to all potential users. Thus, the generic HRTF set ideally contains the most critical features that are crucial to accurately localize a spatial sound source (Parker *et al.*, 2008; Begault, 1994).

4.2.2 Problems with use of HRTF for synthesis of spatial audio

Resuming from the last sub-section, one has to say that the use artificial geometric models for HRTF measurements results in a reasonable spatial sound experience and the front-back confusion still occurs. However, this method is much better when compared to making individual HRTF measurements due to the fact that they are painstakingly time consuming and not pragmatic. With further refinements in

[3]In the coding theory of applied mathematics and electronics engineering, **Golay codes** refers to a particular kind of linear error-correcting code used in digital communications. Error-control codes help prevent the miscommunication of a message by working to correct any mistakes or scrambling of the message that occurs during data transmission. Available at `http://www-math.mit.edu/phase2/UJM/vol1/MKANEM~1.PDF`. Last accessed 2012-07-17.

the geometric models, it can be expected that the sound experience can be better (Parker *et al.*, 2008).

5 Methodology

The primary methodology that was adopted during this thesis was an iterative sound sketching method. The sound sketches designed were for different driving scenarios and the goal was to evaluate it on several jury groups and improve them between evaluations and also before they were tested in the final listening test. Initially designed warning sounds were improved with the help of feedback obtained from each evaluation. A final listening test was conducted after two evaluations which aimed at investigating the final sound designs and also the efficiency of the sound sketching method, which will be discussed more in the next section.

5.1 Sound sketching method

Traditional sketch rendering using pencil and paper is one of the most effective technique of developing many design ideas (Johnson et al., 2009). This way, one has the ability of making a large number of sketches and quickly assessing what sketches to dispose off and what sketches to save, based on the evaluations by a panel of experts in that field (for e.g. automotive design experts to assess several automotive designs). Hence, one readily has the chance to improve the initially made sketches and produce a set of final designs.

Based on this, a sound sketching method is used to test different sound design ideas (Nykänen, 2008). The initial stage of the process consisted of identification of design rules laid down in the literature; identifying and understanding some of the currently used auditory signals used as information/warning sounds in cars. The iterative sound design procedure was the next and the most important stage of the project that consisted of jury group evaluations and analysing their responses. The "final" results were obtained after a series of iterations and improvements of the sound sketches. Figure 5.1 shows a basic approach to the iterative design procedure as an underlying basis for this project. A more detailed explanation of the sound design procedure is shown in the next section.

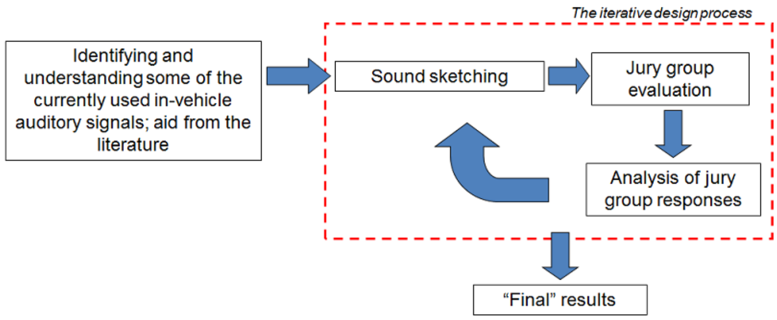


Figure 5.1: *The iterative sound sketching method*

5.2 Sound design

Equipments used

- TASCAM FW1884 Digital Mixing Console
- RME A/D - D/A Converter ADI-2
- M-Audio Keystation 88ES MIDI Keyboard
- Native Instruments ®Komplete 6
- Samplitude Pro X
- Longcat 3D Audio Technologies' ®H3D binaural plug-in
- Behringer Miniamp AMP800 4-Channel Stereo Headphone Amplifier
- Sennheiser HD650 Headphones

The software synthesizers (Native Instruments ®Komplete 6) could be used with Samplitude Pro X as a plug-in effect to treat the incoming audio. They offered a library of sounds that could be used for sound design. The H3D binaural spatializer was used to generate the 3D audio.

The maximum number of tracks chosen for creating sound sketches on Samplitude (see Figure 5.2) were set to 8. This number can vary between sound designers. There was a wide range of sounds offered by each synthesizer, so the next task was to find the sounds which could be used as in-vehicle auditory information/warning sounds. Some of the currently used in-vehicle auditory information signals were identified and understood and the sounds were designed by basing it on those sounds. Most of those signals (earcons) sounded like the sound from the piano keys. Therefore, the designs were mostly based on synthesized sounds for the piano instrument that could be chosen in the MIDI synthesizer browser.

Once a stereo sketch was created, the H3D Binaural Spatializer was used to add generic HRTFs for binaural reproduction via headphones. Figure 5.3 shows an example of the use of the H3D plug-in effect. The listener was assumed to be at the origin point and the sound sources could be moved around the head of the listener, also above and below the head. This plug-in also facilitated the use of multiple sound sources (if needed) while generating the 3D audio. In addition, it offered the use of different room parameters like Presence, Reverberation, Damping. These parameters could be varied from 0% to 100%. The *stereo width* was set to zero to avoid any cross-talk effect on the individual (left-right) channels. Besides, the plug-in employed helped pan the sound sources too.

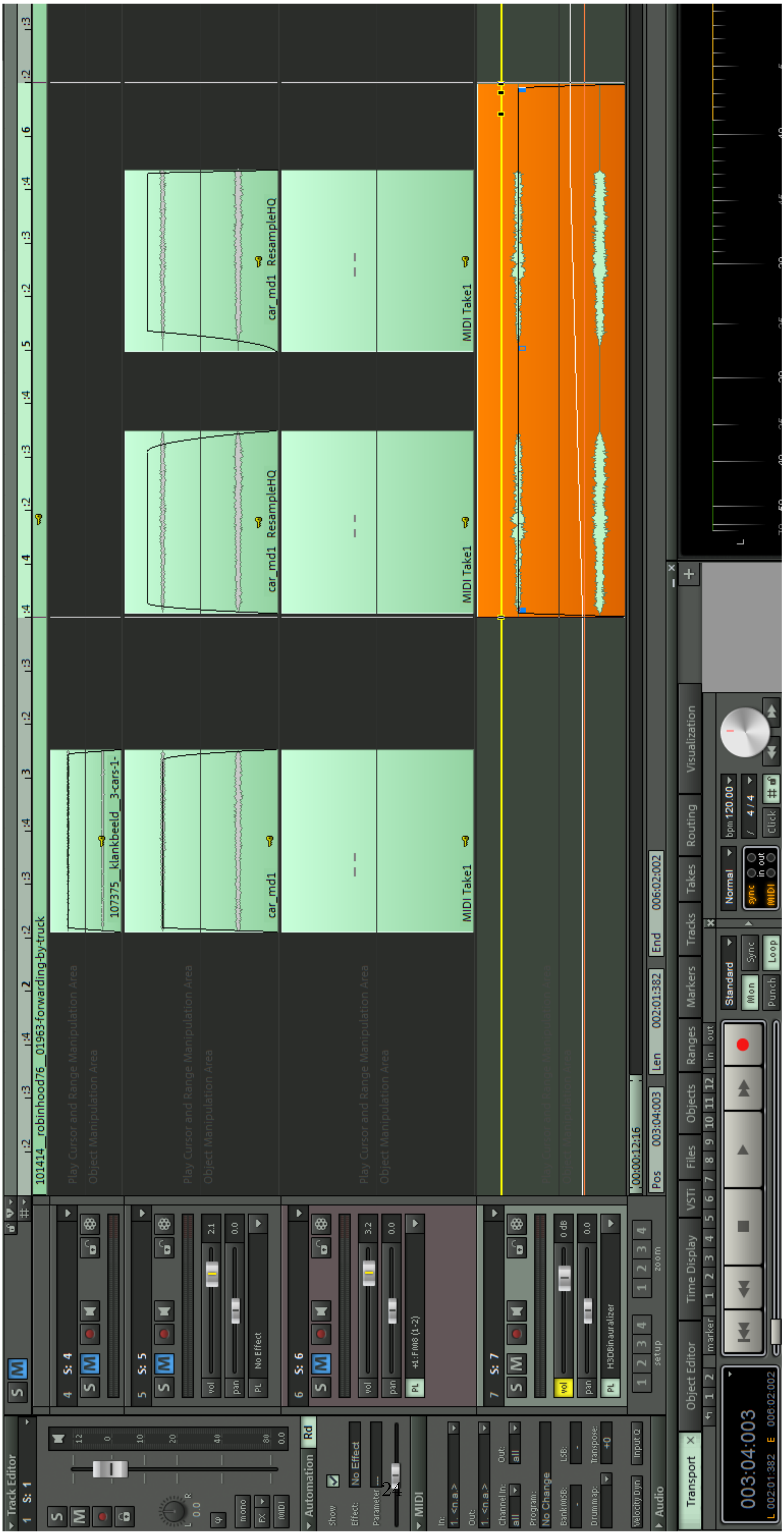


Figure 5.2: *An example of the sound sketching using Samplitude Pro X.*

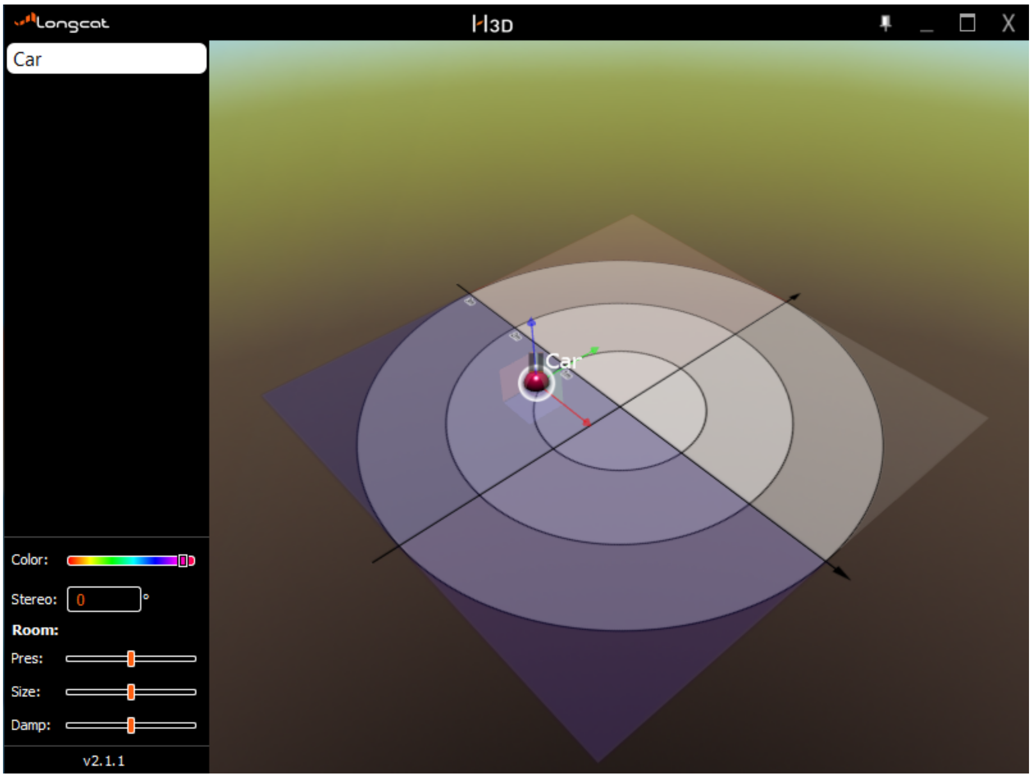


Figure 5.3: *An example of the use of the H3D plug-in effect to create 3D audio. In this case, the car sound is on the left-back side of the listener's head.*

5.3 Choice of driving scenarios/events

There are different driving scenarios (E.g. forward collision, low fuel, pedestrian etc) that one can study as a part of designing sounds for in-vehicle auditory information systems. For the studies carried out in this thesis, two driving scenarios were selected - *blind spot* and *lane departure*, the reasons being that:

1. These two events form a critical part of the early warning systems used in cars today.

2. There is directionality involved in these two scenarios, meaning, these events would occur to the left or right or any other direction (except front) of the vehicle, as opposed to Forward Collision or Adaptive Cruise Control for example, where the events are occurring only in front of the car. Hence, in a demanding traffic situation where the driver doesn't get any information about the direction of events and the environment around his vehicle, 3D sounds could greatly help in giving the necessary information with direction and the driver could benefit from it.

5.4 First jury group evaluation

5.4.1 Implementation

The 1st evaluation was performed on the 26th of April, 2012 and it lasted for about an hour. 6 people participated in this evaluation, all the six being male employees of Semcon (4 of them from A2Zound) and their average age was 44 years. Some of the participants reported some hearing problems - One had some problems with hearing higher frequencies in his left ear, other had a small hearing damage at one of the ears. But it was likely that impairments didn't influence the results since none of the sounds were in extremely high frequency range. Sennheiser HD414 headphones were used and the playback was done using VLC Media Player. This evaluation was broadly divided into two parts:

1. **Introduction** : The 1st part consisted of a short presentation and introduction our thesis work and the background of ADAS to all the participants. The goal of this thesis was briefly stated and four events of ADASs were introduced with pictures to give the listeners some background knowledge for answering the questionnaire.
2. **Questionnaire** : There were 20 different sound sketches in total (8 earcons and 2 auditory icons each for blind spot warning and lane departure warning) and the playlist was randomized. The length of each sound sample was about 6 seconds. The listeners first listened to a general 3D sound demonstration of a wasp flying around the listener's head. This was done in an attempt to give the listener a basic idea of what 3D sounds were. This was followed by the listeners answering the questionnaire. Appendix B shows the format of the questionnaire. The participants were asked to answer 4 questions per sound sketch.

    - The first question was about the choice of events. The listener had to make the choice of what event the sound sketch was best suitable for. This aimed at knowing what our sound sketches reminded the listener of, if they could relate it to the two events we designed it for or something else. Therefore, the options were kept fairly open, the options being *blind spot warning*, *lane departure warning* and *Other early warning systems* with some space for them to write their ideas.
    - The second question consisted of 6 scales ranging from -4 to 4. The first two scales were *Calm - Alert* and *Unhappy - Happy* aimed at knowing the emotional responses of sounds. The other 4 scales were *Not likable - Likable*, *Inappropriate - Appropriate*, *Worthless - Assisting* and *Insignificant - Significant*. The last question was asked to know if the sound reminded them of something significant.
    - The third question was to know if the listeners could localize the direction of the sound sources, they were asked to mark a point on the circumference of a circle implying the direction.
    - The fourth question was asked to know the appropriateness of the given sound sketch as a warning sound. The listener was free to write and

express what he thought could be changed in the sound to make it more appropriate in a vehicle environment.

- The evaluation was then concluded with a short discussion of the sounds.

5.5 Second jury group evaluation

5.5.1 Implementation

The second evaluation was conducted on the 23rd of May, 2012 and it lasted for about an hour and 20 minutes. 9 people totally participated in this evaluation, all 9 being male employees of Semcon (5 of them from A2Zound) and their average age was 37 years. The average experience of the participants with driving was 1-3 times a week. This evaluation was done with all people in the same group. The questionnaire had more examples of how to answer the questions in the questionnaire in an attempt to eliminate any more doubts in the minds of the listeners. Appendix B shows the format of the questionnaire.

The basic approach followed for this evaluation was the same as the first evaluation. But some things were different. It can be broadly classified into 2 sections - introduction and the actual test.

1. **Introduction** : The evaluation started with a short introduction to our work and the questionnaire, some questions that had to be made clearer to the participants.
2. **Questionnaire** : This evaluation consisted of a total of 13 sound sketches (5 earcons and 2 auditory icons for Blind Spot Warning; 5 earcons and 1 auditory icon for Lane Departure Warning) that were tested. The length of each sound sample was about 6 seconds, as before. 2 auditory icons(car engine sound for blind spot; and rumble strips for lane departure) were synthesized.

There were 4 sections : The first section consisted of 13 sound sketches and the listeners had to answer 3 questions per sound sketch.

- **1st Section** : In the first question, the listener had to make the choice of what event the sound sketch was best suitable for, between 6 different choices. The goal was to see how many sounds fell into the more urgent (lane departure, forward collision and overspeed warnings) category and how many fell into the less urgent (blind spot, low fuel and seatbelt warnings) category and also to check how many listeners could identify it correctly. The second question was to gauge the emotional response of our sound sketches. The listeners had to mark their responses on Self Assessment Manikin (SAM) scales. This time, pictures were used instead of words (*Calmness* - *Arousal* and *Unhappy* - *Happy* used in the first jury group evaluation) and their meanings were explained in the beginning, to validate Bradley and Lang's (1994) pre-tested scale and to not focus this study on semantic evaluation using words. The third question was aimed at knowing if the listeners could localize the sound source. The approach was the same as before except that there were letters used this time - R, L, F, B, LF, RF, LB, RB to signify right, left, front, back, left front, right front and the other directions respectively. The listeners had to mark their responses only on one of these directions and were not given the freedom to mark at points on the circle other than these directions. They had to answer these questions for all 13 different sound sketches.

- **2nd Section** : The same 13 sound sketches were used but in a different order than for the 1st section. In the first question, the listener was asked to rate his responses based on the conditions - if the sound sketch was designed for blind spot warning and the same for lane departure warnings on 4 scales ranging from -4 to 4 for the four adjectives (*Not likable - Likable*, *Inappropriate - Appropriate*, *Worthless - Assisting* and *Not urgent - Urgent*) used for test the usefulness of the designed sound sketches. *Not urgent - Urgent* was used this time instead of *Insignificant - Significant* to eliminate any confusion caused by the latter. Therefore, the listeners had to answer the 8 scales for all 13 sound sketches.
- **3rd Section** : In the first jury group evaluation, some sounds with more reverberation were preferred to the other i.e, some sounds inherently were more reverberant than the other. Hence, it was interesting to know the acceptance of having a more reverberant sound as a warning sound in a vehicle environment. Hence there were a total of 20 sketches (10 pairs of less/no and more reverberant earcons) and 10 questions that the listeners had to answer in this section. They had to mark which sound they preferred in each pair, as a warning sound in a vehicle environment.
- **4th Section** : The 4th section consisted of about 20 minutes discussion of some of the sound sketches. This discussion was recorded for transcribing purpose.

5.6 Listening test

Hardware used

- Mac Mini (Serial No.:C07DT1FXDD6L)
- LG Monitor Flatron W2753VC
- Lucid D/A Converter DA9624
- Sennheiser HD650 Headphones

5.6.1 Test Setup

Figure 5.4 shows the setup of the hardware components for the controlled listening test. The D/A Converter was used to convert the digital signal from the computer to analog signal for headphones. The amplification on the converter was adjusted until an ‘optimum’ level was reached. The sampling rate was set to 88.2 *kHz*.

5.6.2 Implementation

The final listening test was conducted in the listening test lab at the department of applied acoustics, Chalmers University of Technology. This room was chosen as

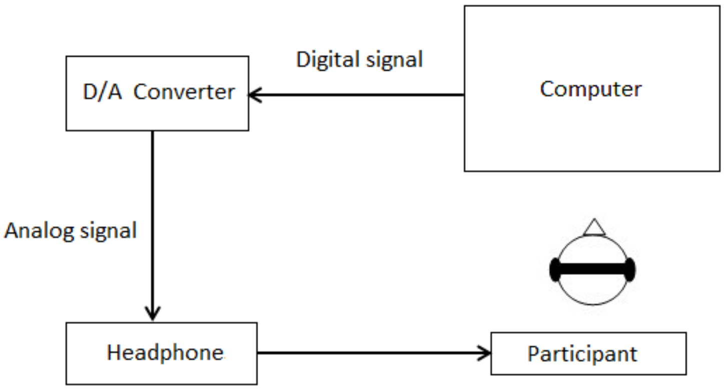


Figure 5.4: *Block diagram of the test setup. The D/A converter was used to provide analog output from the headphones with a sampling rate of 88.2 kHz*

it was semi - anechoic. 24 participants (average age of 26 years; standard deviation of 3.98 years) participated in this final listening test. 1 person did the test at a time and it took around 40 minutes. The average experience with driving was 1-3 times a month. Before the test, each participant was tested for normal hearing using *Audio Console*. Each listener was asked to wear the headphones and press on the click button only when they heard any tones from the headphones. The results showed that none of the participants had any significant hearing problems.

This listening test had 24 different sound sketches in total, which included 1 3D demo sound, 3 simple earcon sounds used as examples, 10 final designed sound sketches and 10 sounds sketches that were used for either one of the jury group evaluations. The length of all the sound samples were reduced to about 3 seconds. The 10 final sketches included 5 sketches(3 earcons, 2 auditory icons) for blind spot and 5 sketches(3 earcons, 2 auditory icons) for lane departure. Since there were 5 sound sketches chosen for use as a final sound designs for each event, there were 5*(5-1)/2 = 10 pairs, that could be used for comparisons in each event. There was a constant stimuli to which every other stimuli was compared to. This meant that, in a given driving context, and in any given pair, it was possible that the earcon sketches were compared to the auditory icon sketches in terms of listener's preference as a warning sound. The test GUI was built using MATLAB, which also helped in randomizing the sound order in each section for each participant.

MATLAB GUI

A MATLAB GUI was built based on the questionnaires from the previous two evaluations. Though still similar, this idea was incorporated to allow for better flexibility in usage between the listener and the interface.

The first part of the GUI was the background questionnaire(See Figure 5.5), where the participants had to answer questions like, 'How often do you drive?'(with the options being *Everyday*, *1-3 times a week*, *1-3 times a month*, *Less than 1-3 times a month*, *Never*) and 'How actively do you listen to sounds (Apart from music)?' (with the options being *Very actively*, *Actively*, *I don't know*, *Not so actively*, *Not actively at all*). For those participants who were not familiar with blind spot and lane departure warning systems, simple images of the event were shown along with some explanation that was given to make them clear about the background.

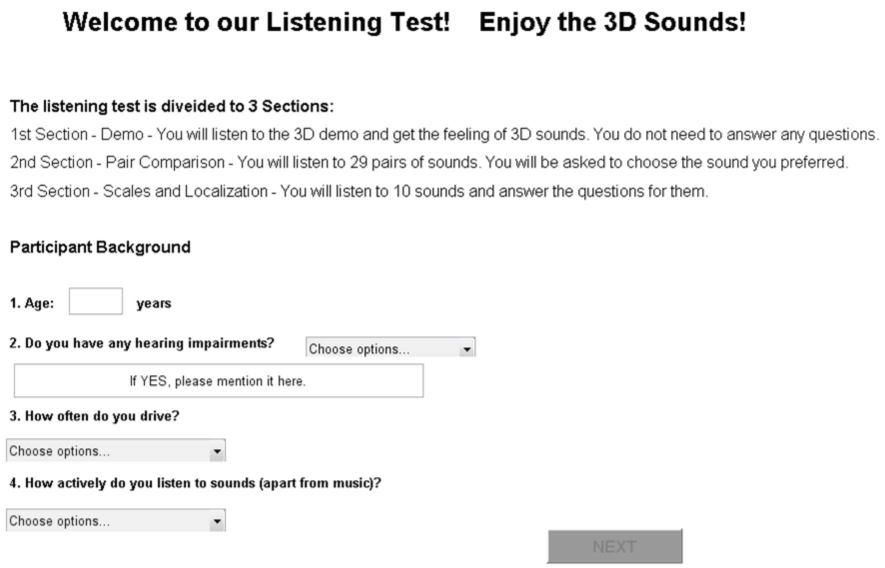


Figure 5.5: *Participant background questionnaire on the MATLAB GUI.*

The listening test was divided into 3 sections:

1. The listeners were asked to listen to a general 3D sound demonstration of a wasp flying around the listener's head. This was done in an attempt, as before, to give the listener a basic idea of what 3D sounds were and also to test the sound from the headphones. In this section, at least one of us sat near the participants to explain the 3D movement to them.

2. Before this section (see figure 5.6), an example was given to let the participants understand how to use the GUI. After the example, the participants started the real test alone. A total 29 pair-comparisons were done in this section - 10 pair comparisons for each of the 2 events; and 9 pair comparisons for original-improved sketches to test the efficiency of the sound sketching method. The listeners were asked 'Which sound do you prefer as a warning sound?' in each pair. The participant was presented with the description of the event by means of some text on the computer screen. After this section, the participants were asked to take a 5 minutes break.

3. Before this section(see figures 5.7 and 5.8), an example was given and SAM scales along with other scales were explained. After the participants were clear about the questions and how to use the GUI, they started the real test independently. 10 final designed sound sketches were played to the listeners in

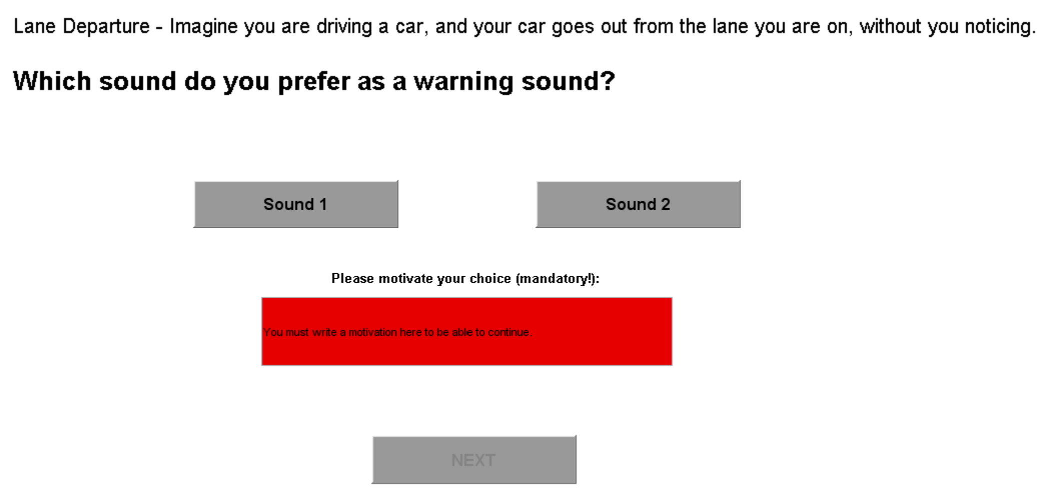


Figure 5.6: *An example of GUI for pair comparison with 2 lane departure sound sketches.*

this section. For each sound, 3 questions were asked. The first question was about the emotional response of the sound. The listeners had to mark their responses on SAM scales. Before the second question, one of the descriptions of the events was given to give the context of the sound sketches. In the second question, the listener was asked to mark his response, based on the context, on 4 scales ranging from -4 to 4 for the four adjectives as before (Not likable - Likable, Inappropriate - Appropriate, Worthless - Assisting and Not urgent - Urgent) (see figure 5.7). The third question was 'where do you think the sound came from?'. R, L, F, B, LF, RF, LB, RB signified right, left, front, back, left front, right front and the other directions respectively (see figure 5.8).

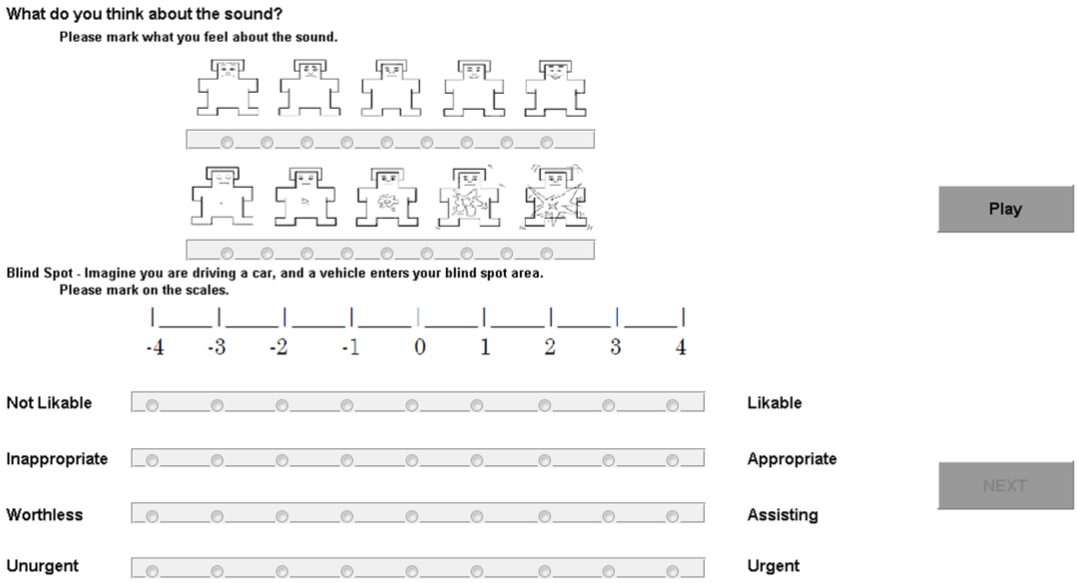


Figure 5.7: *An example of the GUI for scale tests in the context blind spot warning.*

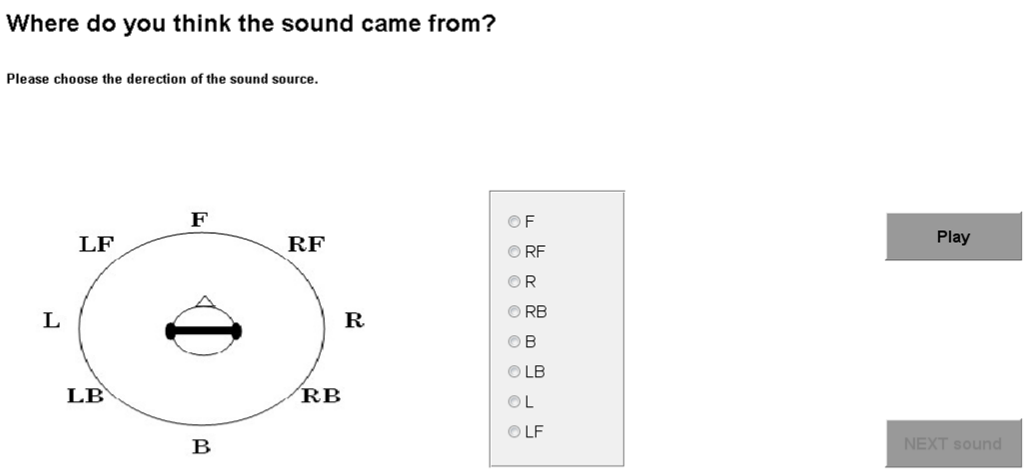


Figure 5.8: *An example of the GUI for the sound localization test.*

6 Results

6.1 First jury group evaluation

In this section, the results from the first jury group evaluation is presented. The results can be shown in four subsections. The first subsection presents the results from the valence - activation scales of the sound sketches designed for both events. The second subsection shows the results of the scale ratings for the two driving scenarios. The third subsection presents results for the questions where the listeners were asked which event they thought the sound was best suitable for. The final subsection presents results for the localization of the 3D sound sources.

6.1.1 Section 1 : Emotional responses from Valence - Activation scales

The emotional responses could be investigated by calculating the mean values of the Arousal scales (*Calm - Alert* and the valence scales *Sad - Happy*) and presenting them for all the sound sketches. The sound sketches 1-10 were designed for blind spot warning and the sound sketches 11-20 were designed for lane departure warning.

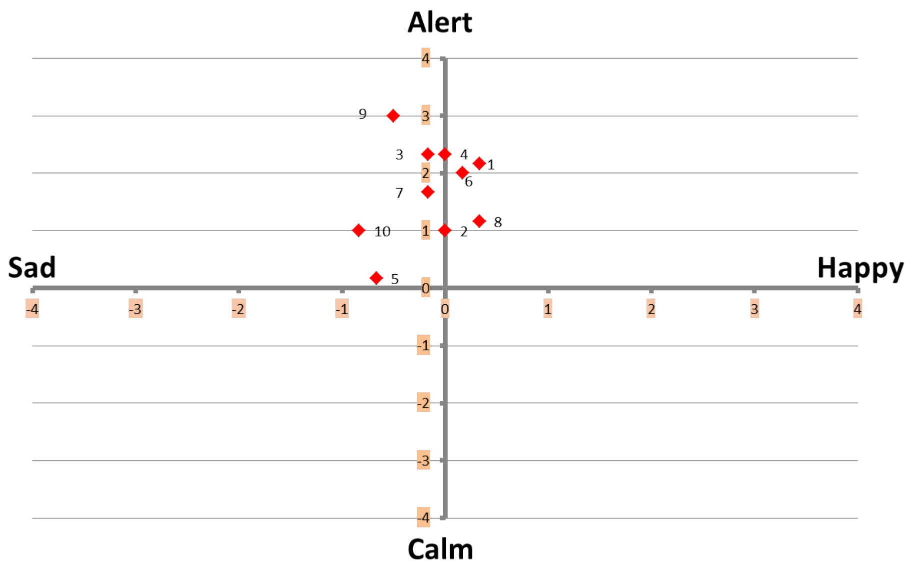


Figure 6.1: *Valence - Activation mappings for 10 different blind spot sound sketches.*

Figure 6.1 shows the Valence - Activation mappings of sound sketches designed for blind spot warning. Sketches 9 and 10 were auditory icons and the rest were all earcons.

Figure 6.2 shows the Valence - Activation mappings of sound sketches designed for lane departure warning. Sketches 11 and 12 were auditory icons and the rest were all earcons.

If one looks at figures 6.1 and 6.2, one can see that, in general, many sound

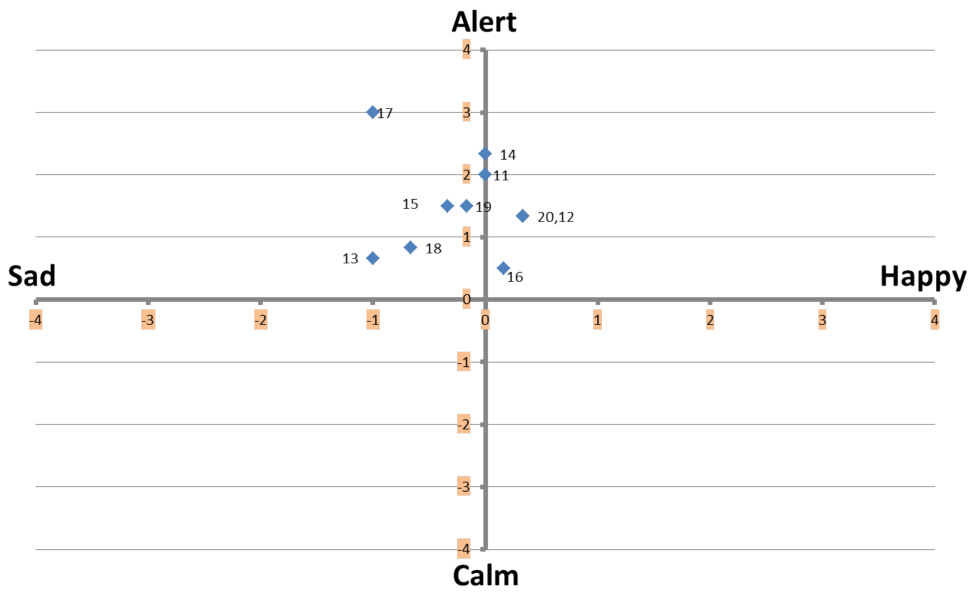


Figure 6.2: *Valence - Activation mappings for 10 different lane departure sound sketches.*

sketches for blind spot warning gave a more pleasant feel when compared to that for lane departure warning (See figure 6.2) which were more arousing. But there were a few sound sketches (5, 9 and 10) that sounded quite unpleasant to the listeners. Sketches 9 and 10 out of these were the auditory icons. For lane departure warning, the sounds were mostly unpleasant. Sound sketches 13, 17 and 18 were very much on the unpleasant side. However, 3 sound sketches for both the events (1, 6 and 8 for blind spot; 12, 16 and 20 for lane departure) sounded quite pleasant to the listeners. On the arousal side, the blind spot warning sounds were more arousing compared to lane departure warning sounds. The goal of the mappings using these valence scales was to see how many sounds landed on the pleasant and arousal quadrant - to measure the emotional response of the designed sound sketches.

6.1.2 Section 2 : Scale ratings for the two driving scenarios

Figure 6.3 shows the mean values with standard deviation for 10 different sound sketches designed for Blind Spot warning. Sketches 9 and 10 are auditory icons and the rest are earcons.

Figure 6.4 shows the mean values with standard deviation for 10 different sound sketches designed for lane departure warning. Sketch 11 and 12 were auditory icons and the rest are earcons. As one can see from the above two figures, the auditory icons generally had higher standard deviation compared to the earcons. Sketches 3, 4, 17 and 20 were among the ones that showed least standard deviation.

6.1.3 Section 3 : Choice of events

Table 6.1 and Table 6.2 shows the individual results from the choice of events for the sound sketches designed for blind spot and lane departure warnings respectively. Sketches 9 and 10 were auditory icons and the rest are earcons. In both the tables,

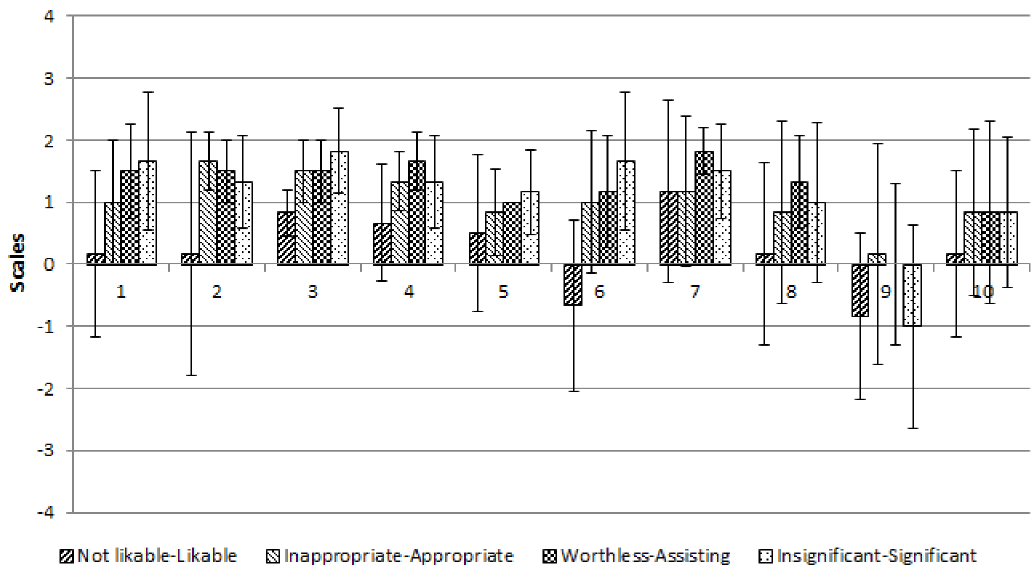


Figure 6.3: *Mean with standard deviations for 10 different Blind Spot Sound Sketches (marked 1 to 10) for all the 4 scales.*

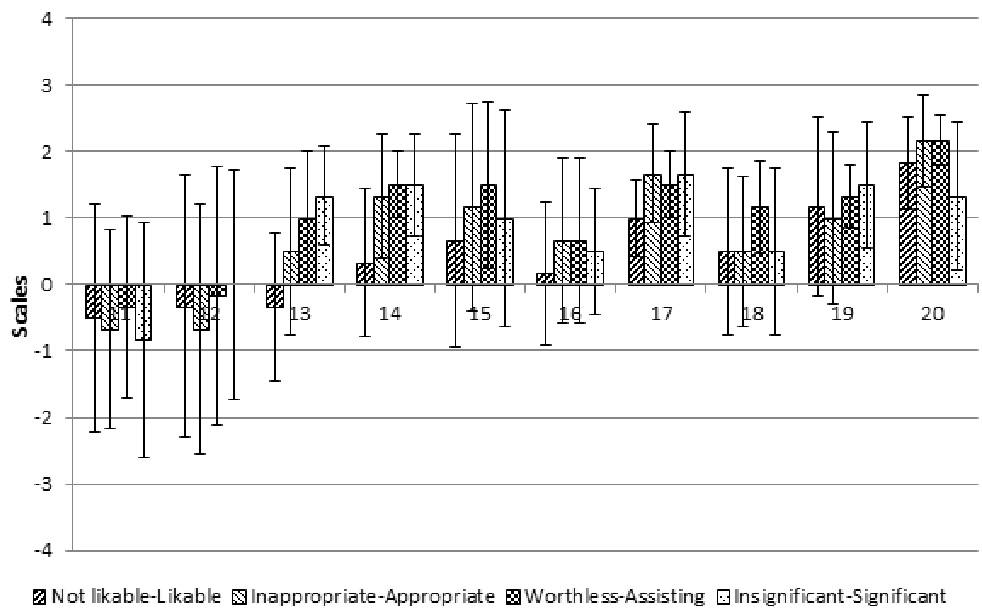


Figure 6.4: *Mean with standard deviations for 10 different Lane Departure Sound Sketches (marked 11 to 20) for all the 4 scales*

letters L, O, BS, FC, LF, A, S, SMS stand for Lane Departure, Overspeed, Blind Spot, Forward Collision, Low Fuel, Alarm, Seatbelt Cellphone SMS warnings/alert sounds respectively. P1, P2, P3, P4, P5, P6 refers to the participants numbered from 1 to 6.

It can be seen that sketch number 8 was the best chosen earcon for the blind spot warning, whereas the other earcons failed to give them an impression of a vehicle in the blind spot. Sketch 9 which was an auditory icon was fairly good in giving them an idea of a vehicle. There was however sketch 7 which was chosen by all

Table 6.1: *Results from the choice of events (designed for blind spot warning).*

Sketch No.	1	2	3	4	5	6	7	8	9	10
P1	L	O	O	L	O	L	L	BS	-	L
P2	BS	L	BS	L	BS	L	L	BS	BS	-
P3	BS	L	L	L	L	L	L	BS	-	BS
P4	FC	BS	BS	BS	L	-	L	BS	-	-
P5	-	-	SMS	S	-	-	L	SMS	BS	BS
P6	L		L	L	L	FC	L	L	BS	-
% Correct	33.3	16.7	33.3	16.7	16.7	0	0	66.7	50	33.3

participants to be suitable for lane departure warning. Sketches 11, 15, 17 and 20 were chosen to be suitable for lane departure warnings. But sketches 16, 18 and 19 were considered suitable for blind spot than lane departure. Therefore, subjectively the sound sketches were more urgent and more suitable for lane departure warning.

Table 6.2: *Results from the choice of events (designed for lane departure warning).*

Sketch No.	11	12	13	14	15	16	17	18	19	20
P1	L	BS	BS	L	L	BS	L	BS	BS	L
P2	BS	-	L	L	BS	L	L	L	BS	L
P3	BS	BS	BS	BS	L	BS	L	BS	BS	L
P4	L	-	BS	BS	L	BS	L	L	L	BS
P5	L	L	-	S	-	SMS	O	-	LF	A
P6	-	-	L	-	-	L	L	BS	L	L
% Correct	50	16.7	33.3	33.3	50	33.3	83.3	33.3	16.7	66.7

6.1.4 Section 4 : Sound localization

Figure 6.5 shows the results for the sound localization from the final listening test. The front-back confusion was 28.3%.

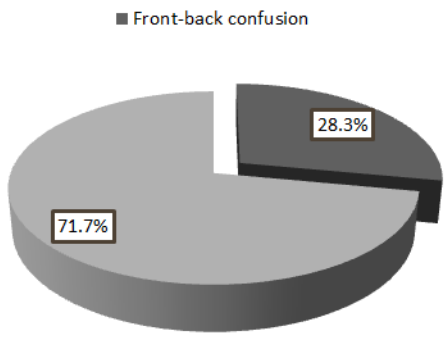


Figure 6.5: *Front-back confusion from the first jury group evaluation.*

See appendix A, table 9.1 for individual participant results.

6.2 1st improvement of sound sketches

Based on the results of the first evaluation, the improvement of the sketches was done. From figure 6.3, one can see that sound sketch 3 has high mean values and low standard deviations for the blind spot warning. This sound sketch was kept for the next evaluation without any changes. From table 6.1, one can see that sound sketch 7, which is designed for blind spot warning, was considered more appropriate for lane departure warning. Also, the 'Worthless - Assisting' value was high when the sound was designed for lane departure warning. This sketch was then used as a lane departure warning sound in the next evaluation. From table 6.1, one can also see that sound sketch 8 was considered more suitable for blind spot warning than others. But the 'Not likable – Likable' value was low if one looks at figure 6.3. Then the timbre of sketch 5, which had higher 'Not likable – Likable' value, was chosen to replace the timbre of sketch 8. High-pass filtering was applied to reduce the reverberation of the sound.

The two auditory icons for Blind Spot warning were: one with the recording of a heavy vehicle engine sound; and the other was a processed V8 engine sound. Both of them didn't show good responses from the evaluation. One reason might be that the sound qualities of one of the icons was not very good and the sketches didn't remind the listeners much of the events. The idea of using the recorded engine noise was saved and the new sketch was designed based on this idea by using a recording of a Mini Cooper engine. The sound quality was improved by filtering some low frequencies. The tyre noise of the car was synthesized and used as a 2nd auditory icon in the 2nd evaluation. The enhanced auditory icons were used in the next evaluation. The auditory icon was combined with a simple earcon. The enhanced auditory icon could help recognize the warning sound and the earcon could keep the sound from being masked by the potential background noise. On the other hand, for lane departure warnings, figure 6.4 shows that the auditory icons (sketches 11 12) were not preferred and the standard deviation of them were quite high as discussed. The idea of sketch 11 was the synthesized rumble strip sound and sketch 12 is the recording of the sound of a real car that went on a rumble strip. The improved plan was to make the synthesized sound closer to the real sound and add some amplitude and volume modulation to try to make the sound more urgent. From figure 6.4, one can also see that sketch 20 had the highest value and the standard deviation was low. This sound was preferred as a lane departure warning sound (more than 50%. See table 6.2). In the discussion part, some listeners gave some valuable suggestions to improve this sound. So the time interval between the earcons was increased to make the sound less urgent.

From the table 6.2, one can see that sketch 17 was chosen as lane departure warning sound, but sounds 13, 18 and 19 were more suitable for blind spot warning sounds. High-pass filtering was applied to sound sketch 17 and the sound was made smoother and less aggressive. The timbre of sound sketch 13 was changed and this sound was used for blind spot warning sound. For sound sketch 18, the numbers of notes in the sound were decreased from 6 to 4 to make the sound less aggressive. This sound was chosen for blind spot warning sound as well, so the low - pass filtering was applied to make the sound less urgent. The interval between earcons in sketch 19 was increased to reduce the urgency level. From figure 6.4, one can see that the scale ratings of sketch 14 were high except the ‘Not likable - Likable’ value, so the pitch of this sound was increased to make it sound more comfortable and sharp. Sound sketch 16 showed lower scale ratings, but it had a ‘desirable’ rhythm. The timbre of blind spot sketch 4 was used and the new combination was considered better than the previous one.

In general, some sounds were disposed off in this iteration and some promising sounds were saved and/or improved. Only one sound went into the next evaluation without changes. Most of the auditory icons showed average results in the evaluation compared to the earcons, and the enhanced auditory icons were tried in the next evaluation. The short conclusion of this evaluation can be seen in table 6.3.

Table 6.3: *Summary of 1st improvement*

Sound Sketch	**1st iteration**
BS Sketch 1	-
BS Sketch 2	-
BS Sketch 3	This sketch was saved.
BS Sketch 4	-
BS Sketch 5	-
BS Sketch 6	-
BS Sketch 7	Used for lane depature; filtered the frequency content at around 5000Hz.
BS Sketch 8	Used blind spot sketch 5’s timbre; filtered.
BS Sketch 9	New design was based on this idea.
BS Sketch 10	-
LD Sketch 11	New design was based on this idea.
LD Sketch 12	New design.
LD Sketch 13	Timbre changed; used for blind spot.
LD Sketch 14	Pitch was made higher.
LD Sketch 15	-
LD Sketch 16	Used blind spot sketch 4’s timbre; filtered.
LD Sketch 17	Filtered.
LD Sketch 18	Frequencies above 2000 Hz brought down by 60 dB and used for blind spot.
LD Sketch 19	Filtered; used for blind spot.
LD Sketch 20	The time interval between the notes was bigger to sound less urgent and adaptive.

6.3 Second jury group evaluation

In this section, results from the second jury group evaluation is presented. The results can be shown in four subsections. The first subsection presents the results from the valence - activation scales of sound sketches designed for both events. The second subsection shows the results of the scale ratings for the two driving scenarios. The third subsection presents results for the questions where the listeners were asked which event they thought the sound was best suitable for. The final subsection presents results for the localization of the 3D sound sources.

6.3.1 Section 1 : Emotional responses from Valence - Activation SAM scales

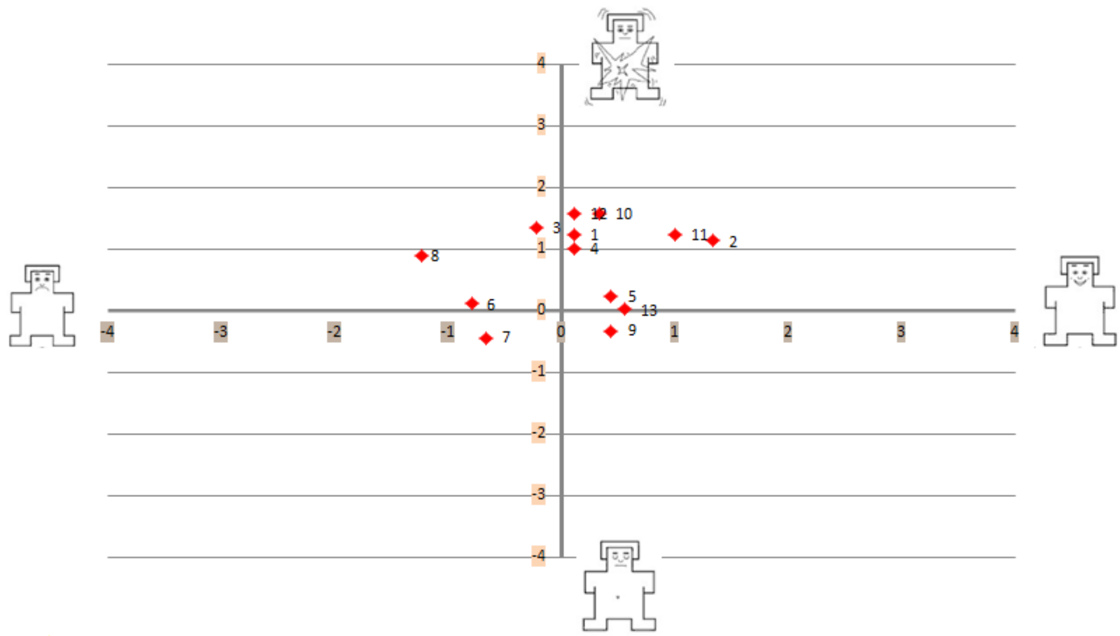


Figure 6.6: *Mean value of the SAM scale values for all the 13 sound sketches.*

Figure 6.6 shows the mappings for the Pleasantness - Arousal ratings for all the 13 sound sketches. Sketches numbered from 1-7 were for blind spot whereas 8-13 were for lane departure warnings. It can be seen that, when compared to Figures 6.1 or 6.2, the above figure shows the mappings were subjectively on the pleasant side for most of the sound sketches. However, auditory icons numbers 6, 7 and 8 were on the not arousing - less pleasant side of the figure. This could possibly question the use of auditory icons as warning alerts in these situations of blind spot warnings or lane departure warnings.

6.3.2 Section 2 : Scale ratings for the two driving scenarios

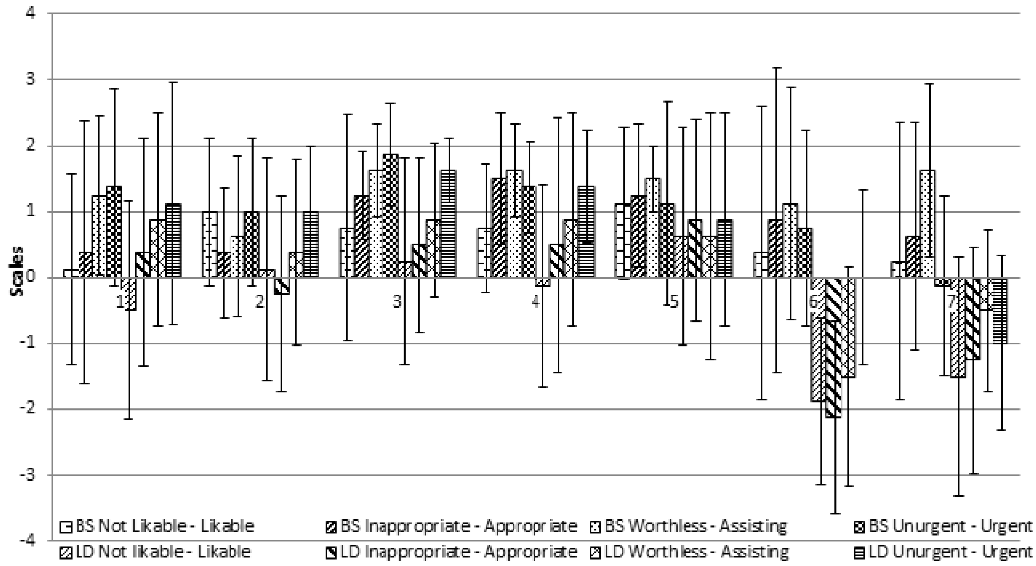


Figure 6.7: *Mean value with standard deviation for the scale values for all the blind spot sound sketches.*

Figure 6.7 shows the mean values with standard deviation for 7 different sound sketches designed for blind spot warning. Sketches 6 and 7 were auditory icons and the rest were earcons. It can be seen that, in general, sketches 3, 4 and 5 show better mean values and lower standard deviation than other sketches. Sketches 6 and 7 which were the auditory icons and showed higher standard deviations. But they were very good in helping the listeners segregate the sketches between blind spot and lane departure. It can be seen that sketches 6 and 7 were less usable for lane departure warnings but could be used for blind spot warnings.

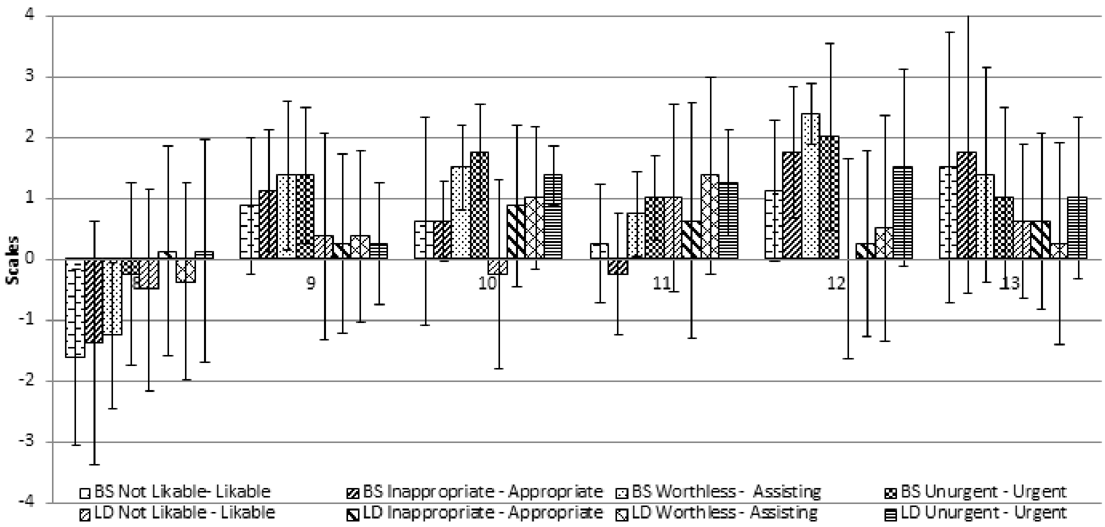


Figure 6.8: *Mean value with standard deviation for the scale values for all the lane departure sound sketches.*

Figure 6.8 shows the mean values with standard deviation for 6 different sound sketches designed for lane departure warning. Sketch 8 was an auditory icon and the rest were earcons. Sketch 8 was perceived as unsuitable both for lane departure or blind spot warnings and showed quite high standard deviations. Sketches 9, 10, 12 and 13 were among the candidates which could be modified and chosen for the next evaluation. Sketches 9 and 12 were among the 'best' sound sketches that helped the listeners distinguish between the lane departure warnings and the blind spot warnings.

6.3.3 Section 3 : Choice of events

Table 6.4: *Results from the choice of events for all the 13 sound sketches.*

Sketch No.	P1	P2	P3	P4	P5	P6	P7	P8	P9	Correct event	% Correct
1	L	L	BS	O	O	O	-	BS	S	BS	22.2
2	L	L	BS	S	BS	BS	LF	S	LF	BS	33.3
3	BS	L	BS	O	BS	O	BS	S	O	BS	44.4
4	L	L	S	FC	BS	BS	S	O	BS	BS	33.3
5	BS	L	S	BS	BS	S	O	LF	LF	BS	33.3
6	BS	BS	BS	BS	BS	LF	BS	L	L	BS	66.7
7	LF	BS	O	FC	BS	L	L	LF	BS	BS	33.3
8	S	-	L	L	BS	L	LF	O	LF	L	33.3
9	S	BS	S	L	LF	L	L	L	L	L	55.6
10	L	L	BS	LF	BS	S	L	BS	BS	L	33.3
11	BS	L	BS	BS	L	FC	LF	S	O	L	22.2
12	L	L	BS	FC	BS	O	O	FC	S	L	22.2
13	L	L	BS	BS	BS	BS	L	BS	BS	L	33.3

Table 6.4 shows the results from 9 participants numbered from P1 to P9 for the choice of events for all the 13 sound sketches designed for both the events. Here L, O, BS, FC, S, LF stands for Lane Departure, Overspeed, Blind Spot, Forward Collision, Seatbelt and Low Fuel warnings/alert sounds respectively.

6.3.4 Section 4 : Sound localization

Figure 6.9 shows the results for the sound localization from the final listening test. The front-back confusion was 13.3%.

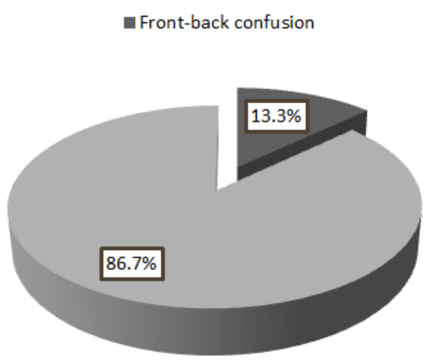


Figure 6.9: *Front-back confusion from the second jury group evaluation.*

See appendix A, table 9.2 for individual participant results.

6.3.5 Section 5 : Pair Comparisons of reverberant/less or not reverberant sounds

Subjectively, the reverberant sounds were less preferred than the not or less reverberant sounds. The percentage win for non-reverberant sounds in each case was shown mostly with big margins, except in pairs 2 and 5 (see appendix A, tables 9.4 and 9.5) where reverberant sounds were more preferred. Pairs 6, 7 and 8 were among the pairs where the percentage win of the non-reverberant sounds was big. On the whole, one could see that the less or not reverberant sound sketches were preferred 80% of the time and hence should be used for the next evaluations.

6.4 2nd improvement of sound sketches

From figure 6.7, one can see that the scale ratings of sound sketch 3, 4 and 5 were higher for blind spot warning sounds than sketches 1 and 2. So sketches 3 and 5 were kept for the final test. Sketch 4 was filtered to sound more gentle and the number of the notes was reduced from 6 to 4. The new rhythm was slower and the sound didn't sound aggressive.

For the auditory icons, one can see that the scale ratings for sketch 6 and 7 were better for blind spot warning than the lane departure warning from Figure 6.7. These two sounds were kept for the final test. Based on some suggestions from the discussion, each sound sketch was made shorter, so the sound could remind the listener of the event while minimizing annoyance.

Sketches 10 and 11 had higher scale ratings than other sketches when they were considered as design for lane departure warning sound. So sketch 11 was kept for the final test. The 'not likable-likable' value for sketch 10 was not high. So the timbre of sketch 12, which had the higher value in 'not likable-likable', was used for sketch 10. For sketch 13, the rhythm of sketch 9 was used and the sketches were filtered to make the sound more urgent. Also, the time interval between the notes was decreased to increase the urgency.

From Figure 6.8 , one can see that the rating for sketch 8 was very low and the standard deviation was high. From the discussion, most of the participants didn't get the idea of the rumble strip and thought the sound was too 'floating'. Hence, it was desirable that the sharpness of sketch 8 was increased. The pitch of this sound was increased to higher the urgency level. The other auditory icon for the blind spot scenario was designed based on sketch 8 and sounded sharper than the other icon. The reverberation of the sound was decreased and sound of the tyre going through the rumble strip was enhanced in an attempt to increase the listener's feeling about the car vibrating.

Summary of the 2nd improvements can be seen in table 6.5.

Table 6.5: *Summary of the 2nd improvement.*

Sound Sketch	**2nd iteration**
BS Sketch 1	-
BS Sketch 2	-
BS Sketch 3	This sketch was saved.
BS Sketch 4	Filtered; the notes were closer to each other.
BS Sketch 5	This sound was saved.
BS Sketch 6	Filtered.
BS Sketch 7	New design based on this idea.
LD Sketch 8	New design based on this idea.
LD Sketch 9	-
LD Sketch 10	Used the timbre of sketch 12.
LD Sketch 11	This sound was saved.
LD Sketch 12	-
LD Sketch 13	Removed higher frequencies (above 3000 Hz, reduce around 20 dB). Used the rhythm of lane departure sketch 9.

6.5 Listening test

This section presents the results from the final controlled listening test. The first subsection shows the SAM scale mappings for the final 10 sound sketches.

6.5.1 Section 1 : Emotional responses from Valence - Activation SAM scales

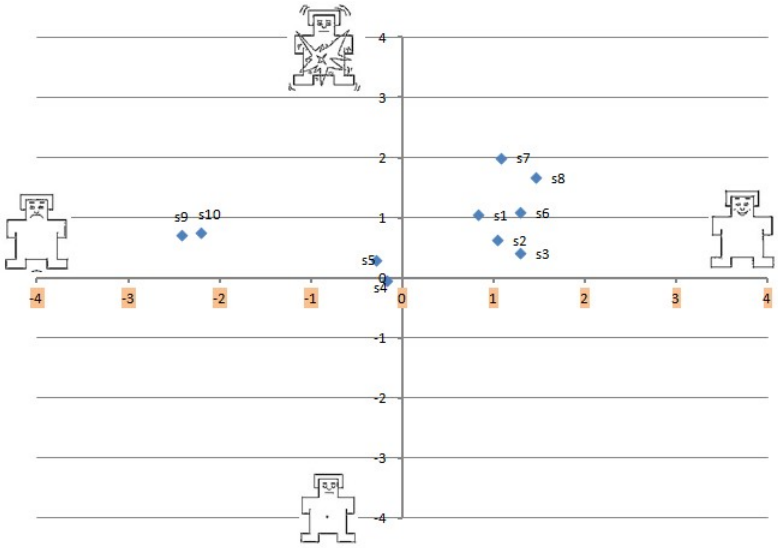


Figure 6.10: *SAM scale mappings for all the 10 sketches (referred to as S1 to S10). S1, S2, S3, S4, S5 were for blind spot warnings; S6, S7 and S8, S9, S10 were all sketches for lane departure warnings. S4, S5, S9 and S10 were auditory icon sketches and the remaining were earcon sketches.*

Figure 6.10 shows the SAM scale mappings for the 10 sound sketches used in the final listening test. It can be seen that most sketches were placed in the first quadrant (Pleasant - Arousing). The sketches that were not placed in this quadrant were the auditory icon sketches (S4, S5, S9 and S10). S9 and S10 were not perceived to be pleasant at all but were arousing to some extent. S6, S7 and S8 were among the most arousing and pleasantness-emotion inducing sound sketches. These were sketches for lane departure warnings. The earcon sketches (S1 to S3) for blind spot warnings, as desired, were generally at lesser arousal ratings than those for lane departure warnings.

6.5.2 Section 2 : Scale ratings for the two driving scenarios

Figure 6.11 shows the mean values with standard deviations of the scale values for all the 10 sound sketches. S4, S5, S9 and S10 were auditory icon sketches and the remaining were earcon sketches. S1, S2, S3, S4, S5 were for blind spot warning; S6, S7 and S8, S9, S10 were all sketches for lane departure warning. Subjectively, lane departure earcon sketches showed greater scale ratings than blind spot earcon sketches. Among lane departure sound sketches, sketch 7 showed highest mean values. It was followed by sketches 8 and 6. Among blind spot sound sketches, sketch 1 showed the highest mean value and it was followed by sketches 2 and 3. All the auditory icons, showed high standard deviations and average or negative

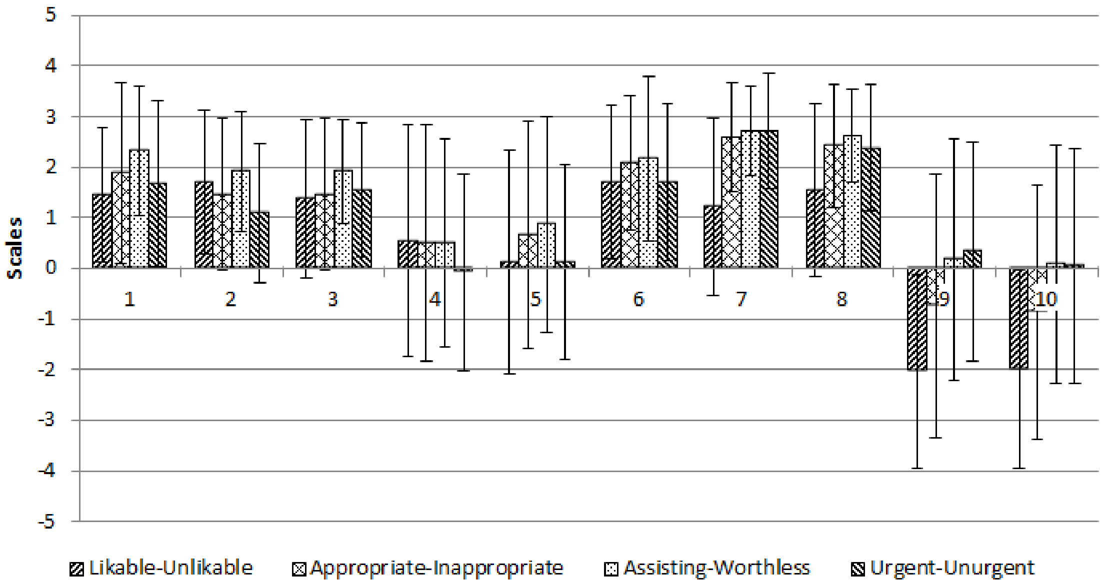


Figure 6.11: *Mean value with standard deviation of the scale values for all the 10 sound sketches.*

mean values.

6.5.3 Section 3 : Pair comparisons of final sound sketches

There were 5 sound sketches for each event. The goal was to arrange the 10 pairs for each event such that a proportionality matrix could be created, where in, alongside the left side of the diagonal, the proportion with which the first sound was chosen over the second was written down. The values on the other side of the diagonal can be given by one minus the values on the left of the diagonal. Table 6.6 shows an example of how the proportionality was created.

Table 6.6: *An example proportionality matrix used for the pair comparisons. It shows the proportion which 1st sound in each pair (sketch number shown in rows) was chosen over the second sound (sketch number shown in columns).*

Sketch No.	1	2	3	4	5
1	0	$1\text{-}p_{2>1}$	$1\text{-}p_{3>1}$	$1\text{-}p_{4>1}$	$1\text{-}p_{5>1}$
2	$p_{2>1}$	0	$1\text{-}p_{3>2}$	$1\text{-}p_{4>2}$	$1\text{-}p_{5>2}$
3	$p_{3>1}$	$p_{3>2}$	0	$1\text{-}p_{4>3}$	$1\text{-}p_{5>3}$
4	$p_{4>1}$	$p_{4>2}$	$p_{4>3}$	0	$1\text{-}p_{5>4}$
5	$p_{5>1}$	$p_{5>2}$	$p_{5>3}$	$p_{5>4}$	0

Based on this, tables 6.7 and 6.8 were created, that summarize the results for blind spot warning (sketches 1 to 5) and lane departure warning(sketches 6 to 10).

These proportionality values were then transformed to Z values from the Z scores tables shown in Gescheider (1997). Following his methodology, the final scale values were also computed.

Tables 6.9 and 6.10 show the Z-scores and final scales values for the blind spot and lane departure warnings respectively. The final scales values were calculated

Table 6.7: *Proportionality matrix (proportions rounded off to the nearest hundredths decimal place) for blind spot pair comparisons.*

Sketch No.	1	2	3	4	5
1	0	0.54	0.42	0.75	0.67
2	0.46	0	0.42	0.75	0.67
3	0.58	0.58	0	0.62	0.75
4	0.25	0.25	0.38	0	0.5
5	0.33	0.33	0.25	0.5	0

Table 6.8: *Proportionality matrix (proportions rounded off to the nearest hundredths decimal place) for lane departure pair comparisons*

Sketch No.	6	7	8	9	10
6	0	0.29	0.33	0.79	0.92
7	0.71	0	0.75	0.92	0.87
8	0.67	0.25	0	0.79	0.83
9	0.21	0.08	0.21	0	0.5
10	0.08	0.13	0.17	0.5	0

Table 6.9: *Z score matrix for blind spot pair comparisons.*

Sketch No.	1	2	3	4	5
1	0	0.1	-0.2	0.67	0.44
2	-0.1	0	-0.2	0.67	0.44
3	0.2	0.2	0	0.31	0.67
4	-0.67	-0.67	-0.31	0	0
5	-0.44	-0.44	-0.67	0	0
Mean	-0.202	-0.162	-0.276	0.33	0.31
Final	0.074	0.114	0	0.606	0.586

Table 6.10: *Z score matrix for lane departure pair comparisons.*

Sketch No.	6	7	8	9	10
6	0	-0.55	-0.44	0.81	1.41
7	0.55	0	-0.25	1.41	1.13
8	0.44	0.25	0	0.81	0.95
9	-0.81	-1.41	-0.81	0	0
10	-1.41	-1.13	-0.95	0	0
Mean	-0.246	-0.568	-0.49	0.606	0.698
Final	0.322	0	0.078	1.174	1.266

by first setting a zero value to the lowest mean value and then adding the absolute of that mean value to the desired mean value, i.e,

Final scale value$_j$ = Absolute value of the least mean value + The given mean value$_j$.

The higher the final scale value, the less preferred the sound sketch was. From the final scale values, one can see that sketch 3 was the most preferred sound for blind spot warning and sketch 7 was the most preferred sound for lane departure warning. Auditory icons when compared to the earcons, showed higher final scale values and hence were less preferred by the subjects.

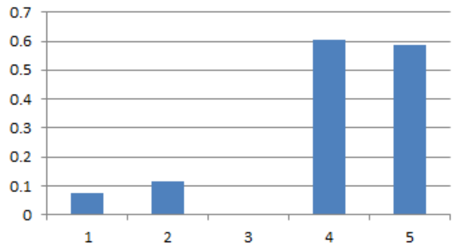


Figure 6.12: *Final scale values for blind spot sound sketches.*

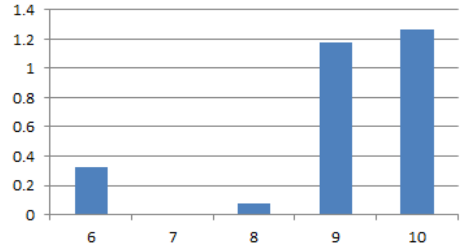


Figure 6.13: *Final scale values for lane departure sound sketches.*

6.5.4 Section 4 : Sound localization

Figure 6.14 shows the results for the sound localization from the final listening test. The front-back confusion was 15.8%.

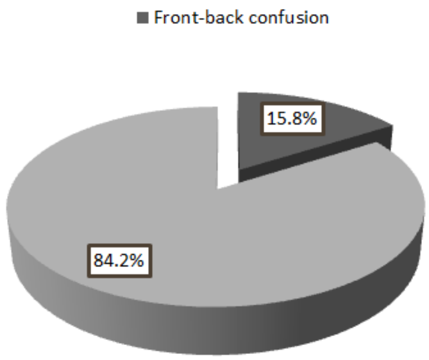


Figure 6.14: *Front-back confusion from the final listening test.*

See appendix A, table 9.3 for individual participant results.

6.5.5 Section 5 : Pair comparisons of original and improved sound sketches.

The methodology followed here was not similar to that followed for previous pair comparisons (see section 5.6.2). Here, each pair consisted of one original sound sketch (used during previous jury group evaluations) and one improved sound sketch. One sound sketch in a pair was compared only once to the other sound in the same pair.

Table 6.11 shows the results from the pair comparisons of original and improved sound sketches. The significance values were calculated by performing paired sampled t-tests using the IBM SPSS software for 95% confidence interval. Therefore, it was desired to have significance values below 0.05 to consider the answers significant. Pairs 4, 5, 6 and 8 were auditory icon pairs and the rest were earcon pairs. It can be seen that only 3 values were below (from Pairs 5, 6 and 9) which makes these results quite open for discussions.

Table 6.11: *Results from the sketching method pair comparisons for all the 9 pairs obtained from all the participants. LD and BS stand for Lane Departure and Blind Spot. Significance is shown for 95% confidence interval.*

Pair No	Winner	Significance
Pair 1	Old BS	0.426
Pair 2	New LD	0.692
Pair 3	New LD	0.103
Pair 4	Old LD	0.426
Pair 5	New BS	0.000
Pair 6	New BS	0.038
Pair 7	Old LD	0.228
Pair 8	Old LD	0.103
Pair 9	New LD	0.011

7 Discussions and conclusions

In general, it was desirable to have pleasant warning sounds since it is the pleasantness in the sounds that could potentially influence the end users' positive emotional reactions. If one looks at the valence - activation scale mappings from the 1st and 2nd evaluations' results and compare them to the SAM scale mappings from the listening test results, it can be seen that the subjective perception of the sketches changed from less pleasant to more pleasant which was desirable. However, the auditory icon sketch mappings from the final listening test are on the less pleasant and less arousing side, lane departure sketches being less pleasant than blind spot. The ideal mappings would be high pleasantness and medium to high arousal ratings for the blind spot warnings; high arousal and less pleasantness ratings for the lane departure warnings. But that was the case as seen from the results from previous evaluations too. Therefore, it is of interest to know if listeners really prefer auditory icons as in-vehicle warning sounds at all. Further discussions could throw more light on this topic.

From the results of the scale ratings after each evaluation, it is evident that the scale ratings seemed to increase along with lower standard deviations in the final listening test than that from the previous evaluations. This is probably because of the reason that in the final evaluation, the driving context was presented to the listener which helped reduce variations in the answers. From the scale ratings in the second evaluation, it can be seen that there are relatively high deviations in most answers. This could have been because of the design of the questionnaire. The 1st section of the questionnaire consisted of the choice of events, where the listeners had to match the designed sounds to one of the 6 choices of events according to its suitability. The 2nd section consisted of the subjects listening to the same sounds (in a different order) and evaluating their 'usefulness' in 4 scales for the 2 chosen driving contexts. This design of the questionnaire could have provided the listeners the freedom to differ in their opinions in each of the sections and it was reflected by a relatively high standard deviation in the scale ratings. This methodology, however, consisted of sounds that showed good potential of its usability for the chosen driving context.

In the 2nd jury group evaluation, an investigation of the acceptance of reverberant sounds (in relation to non-reverberant sounds) was carried out. It could be seen that 80% of the time, non-reverberant sounds were preferred over the less/not reverberant sounds. This evinces that reverberant sounds could have been pleasant (at times, according to the discussions after 2nd jury group evaluation) but were not urgent enough to give the right information to the driver. This investigation was performed because some sounds were inherently more reverberant than the other in the first evaluation, and were chosen to be more 'premium' sounding than the rest. Therefore it was interesting to investigate the incorporation of reverberant sounds in relation to non-reverberant sounds.

The pair comparisons of the final designed sounds helped rank the sounds. According to the final scale values, sketch 3 was the most favourite blind spot warning sound, followed by sketches 1 and 2. For the lane departure warnings, sketch 7 was the most favourite warning sound, followed by sketches 8 and 6. The auditory icons for the both the events, again were not preferred as much as the earcons (see figures 6.12 and 6.13). Therefore from all the evaluations, it could be concluded that the designed auditory icons were less preferred compared to the designed earcons, as in-vehicle warning sounds for the two chosen driving contexts. However, the comparison between these two icons is applicable only for certain urgent situations where there is an immediate need to inform the driver of the environment, for example forward collision, in which case auditory icons would probably work better than earcons. Earcons are more preferred probably due to the fact that they are most commonly used today in interfaces. Therefore, the users could be more used to them than everyday sounds. If the listeners undergo some 'training' by listening to a few everyday sounds before the tests, their acceptance might increase.

The generic HRTFs were used by the plug-in, it was one of the obvious reasons for the front-back confusion. But it was desirable to reduce them. The results from the sound localization show that the front-back confusion was reduced from about 28% to 13% after the 2nd jury group evaluation, which was a desirable progress. This shows that the slight panning of the sound sources (around 20 - 25 degrees in the 2nd evaluation; 10 - 15 degrees in the listening test) in the respective direction helped improve the sound localization. This confusion increased slightly to about 16% in the final listening test but evidently it wasn't a significant increase. The hardware used in this listening test was different from the previous evaluations - better quality headphones and a D/A converter was used. Also, the fact that the length of the sound samples was reduced from 6 seconds to 3 seconds perhaps gave the listener less time to locate the sound source correctly. Therefore, it is likely that this could have influenced the localization.

One can also pinpoint the limitations of the chosen sound design method. Since only one method has been used for this study, it could be of interest to test other methods for designing sounds, not by the use of the audio mixing software and the 3D plug-in, but by the use of binaural synthesis methods. For example, by measuring binaural transfer functions from the known source position to the listener's position and then auralising sources signals with the binaural transfer functions, one could essentially allow artificial positioning of sound sources in free field conditions, i.e, create 3D sounds (Nykänen, 2008).

Paired sample t-tests were performed to test the significance of the results from the 9 pair comparisons. From the results of the original-improved pair comparisons from the listening test, one can see that only 3 out of 9 pairs had significance values the desired value of less than 0.05. Two of these pairs were the auditory icon sketches (pairs 5 and 6) and the third pair was the earcon sketch (pair 9). This shows that the results for the final auditory icon designs improved in general and were more significant compared to the final earcon designs. Only 1 earcon pair comparison yielded a significance value of less than 0.05. However in pair 4, the older auditory icon for lane departure was preferred over the newer auditory icon, which was designed to be more sharper. The listeners also preferred the older auditory rumble strip sound over the newly designed auditory icon in the pair 8. The newer auditory icon was designed to be more sharper compared to the older sound, which was a recording of a car passing over a rumble strip. Therefore, this sound was expected to be more suitable for an urgent event such as lane departure. But it didn't give the listeners an impression of that. In pair 1, the old blind spot earcon was the winner according to the whole consensus. But it is vital to mention that among the 4 experienced listeners (listeners with better knowledge of sounds / acousticians), the results differed. 3 out of the 4 experienced listeners chose the new blind spot earcon over the old.

Therefore, these results probably can be attributed to the fact that 20 out of the 24 participants were naive listeners. It would probably be beneficial to have equal number of experienced listeners and naive listeners so the differences in their results could be evaluated. Further, it could also be better to have people with extensive driving experience over people with fair driving experience (1-3 times a month on a average as per the data from the participant background). This could also call for more experience of the designer in the field of sound design, so bigger problems could be identified and fixed at an early stage.

8 Future work

Since the acceptance of the auditory icons as 3D warning sounds in this project was not as high as earcons, it could be of interest to focus more on the design of more suitable auditory icons.

From the sound designer's perspective, there are certain things to investigate. During the initial design stages, the sketches were informally tested in a driving simulator by using a game (Driving Simulator 2011). The sketches were manually played to the driver driving the car in the simulator game. A vast majority of these sketches seemed less urgent during the try out in the driving simulator than during the sound design process. Thus, this apparent change in the urgency level could be important to take into account if this experiment is implemented in a simulator.

It could also be interesting to evaluate the sound sketches in a real-time driving simulator to investigate if they work better with visual display. One could check if the visual modality helps reduce the front-back confusion, which could be a good supplement to this thesis. Further on, certain things that can be taken into account while doing this study are - if 3D sounds make the driver turn his head towards the direction of the sound source, if the front-back confusion is likely to cause serious problems during the driving task or even to check, if 3D sounds could distract the driver while driving. The head movement of the listener is not considered in this thesis. The listener is assumed to be facing to the front. It could be interesting to track the head movement since the location of the sound source changes with reference to the axis (driver facing front).

Eventually, 3D sounds could also be tested for other different warning systems, for instance, cross traffic information or pedestrian crossing etc.

Bibliography

Begault, D. R. (1994). *3-D Sound for Virtual Reality and Multimedia.*, Academic Press Professional, Cambridge, MA.

Bellotti, F., Berta R., Gloria, A. D., and Margarone, M. (2002). "Using 3D Sound to Improve the Effectiveness of the Advanced Driver Assistance Systems," *Personal and Ubiquitous Computing*, Department of Electronics and Biophysical Engineering, University of Genova, Italy, Springer UK, 155-162.

Blattner, M. M., Sumikawa, D. A. and Greenberg, R. M. (1989). "Earcons and icons: Their structure and common design principles," *Human Computer Interaction*, 4(1), 1989, 11-44.

Blauert, J. (1983). *Spatial Hearing: The Psychophysics of Human Sound Localization*, MIT, Cambridge, MA.

Bradley, M. M. and Lang, P. J. (1994). "Measuring Emotion: The Self-Assessment Manikin And The Semantic Differential," In *Journal of Behavior Therapy and Experimental Psychiatry*, Vol. 25, 49-59.

Brazil, E. and Fernström, M. (2004). "Human-Computer Interaction Design based on Interactive Sonification – Hearing Actions or Instruments/Agents," *Proceedings of the 2004 International Workshop on Interactive Sonification*, Bielefeld University, Germany, 1-4.

Brazil, E. and Fernström, M. (2011). "Chapter 13: Auditory Icons," In Hermann T., Hunt A., Neuhoff J. G. (Eds),*The Sonification Handbook*,Logos Publishing House, Berlin, Germany, 325-338

Bregman, S. A. (1990). "Auditory scene analysis," In Smelzer, N.J. and Bates, P.B.(Eds.), *International Encyclopedia of the Social and Behavioral Sciences*, MIT Press: Cambridge, MA, 3-5.

Brewster, S. A. (1994). *Providing a structured method for integrating non-speech audio into human-computer interfaces.* PhD thesis, Human-Computer Interaction Group, Department of Computer Science, University of York,

Brewster, S. A. (2002). "Non-speech auditory output," *The Human Computer Interaction Handbook*, J. Jacko and A. Sears, Eds. USA: Lawrence Erlbaum Associates, 220-239.

Brewster, S. A., Wright, P. C. and Edwards, A. D. N. (1995). "Experimentally derived guidelines for the creation of earcons," In *Proceedings of BCS-HCI 95*, volume 2, Huddersfield, UK, Springer, 155-159.

Campbell, J. L., Richard, C. M., Brown, J. L. and McCallum, M. (2007). "Crash warning system interfaces: Human factors insights and lessons learned," *U.S Department of Transportational National Highway Traffic Safety Administration*, No. HS 810697.

Cheng, C. I. and Wakefield, G. H. (2001). "Introduction to Head-Related Transfer Functions (HRTFs): Representations of HRTFs in Time, Frequency, and Space," *Journal of the Audio Engineering Society*, Vol 49, No 4, 231-249.

Dingus, T.A. and Hulse, M.C. (1993). "Some human factors design issues and recommendations for automobile navigation information systems," Transportation Research, 1C(2), 119–131.

Duraiswami, R., Zotkin, D. N., Davis, L. S. (2001). "Rendering Localized Spatial Audio in a Virtual Auditory Space," *Perceptual Interfaces and Reality Laboratory*, Institute for Advanced Computer Studies, University of Maryland, College Park, MD 20742, 4-7.

Durlach, N. I., Colburn, H. S. (1978). "Binaural phenomenon," In E. C. Carterette and M. P. Friedman (Eds), *Handbook of perception, Vol. IV Hearing.* New York: Academic Press.

Edworthy, J. (1994). "The design and implementation of non-verbal auditory warnings," In *Applied Ergonomics 1994*, Department of psychology, University of Plymouth, UK, 25(4), 202-210.

Edworthy, J. and Adams, A.S. (1996). *Warning Design: A Research Prospective.* London : Taylor and Francis.

Edworthy, J. and Haas, E. (2008). "Chapter 14: An introduction to auditory warnings and alarms," In Wogalter, M (Ed) *Handbook of Warnings* 189-191.

Edworthy, J., Loxley, S. and Dennis, I. (1991). "Improving auditory warning design : The relationship between warning sound parameters and perceived urgency," *Human Factors*, 33, 205 - 231.

Edworthy, J., and Stanton, N. (1995). "A user-centred approach to the design and evaluation of auditory warning signals: 1. Methodology," *Ergonomics*, Vol. 38 No. 11, 2262-2280.

Ermi, L. and F. Mäyrä. (2005). "Fundamental Components of the Gameplay Experience: Analysing Immersion," Paper read at *Changing Views – Worlds in Play*, June 16—16, Toronto.

Garzonis, S., Jones, S., Jay, T. and O'Neill, E. (2009). "Auditory icon and earcon mobile service notifications: intuitiveness, learnability, memorability and preference," In *Proceedings of the CHI 2009*, Boston, MA USA, Vol. 1, 1513–1522.

Gaver, W. (1986). "Auditory icons: Using sound in computer interfaces," *Human-Computer Interaction*, 2, 167-177.

Gaver, W.W. (1993). "Synthesizing auditory icons," *In Proceedings of INTER-CHI'93*, volume 1, Amsterdam, The Netherlands, ACM Press, 228–235.

Gehring, B. (1992). "Interactive Entertainment With Three Dimensional Sound," Presented at the *134th SMPTE Technical Conference*, 1-4.

Gehring, B. (1997). "Why 3D Sound Through Headphones?", Presented at the *Computer Game Developers' Conference*, 1-7.

Gescheider, G. A. (1997). "Chapter 9 : The Measurement of Sensory Attributes," In G. A. Gescheider, *Psychophysics: The Fundamentals*, Mahwah, NJ, Lawrence Erlbaum Associates, 198-206.

Grimshaw, M., Lindley, C.A., Nacke, L. (2008). "Sound and Immersion in the First-Person Shooter : Mixed Measurement of the Player's Sonic Experience," *Development*, AudioMostly, 1-7.

Grosse, K. (2009). *Audio-Visual Perception in Interactive Virtual Environments*, Master's Thesis, Institute of Computer Graphics and Algorithms, Vienna University of Technology, 10 - 15.

Johnson, G., Gross, M. D., Hong, J. and Do E. Y. -L. (2009). "Computational support for sketching in design: a review," In *Foundations and Trends in Human-Computer Interaction*, Vol. 2, No.1, 1–4.

Larsson, P., Opperud, A., Fredriksson, K. and Västfjäll, D. (2009). "Emotional and behavioural response to auditory icons and earcons in driver-vehicle interfaces," *Proceedings of the 21st International Technical Conference on the Enhanced Safety of Vehicles*, Stuttgart, Germany, June 15-18, 2009. Paper No. 09-0104-O, 1-12.

Lee, J. D., McGehee, D. V., Brown, T. L., and Reyes, M. L. (2002). "Collision Warning Timing, Driver Distraction, and Driver Response to Imminent Rear-End Collisions in a High-Fidelity Driving Simulator," *Human Factors and Ergonomics Society*, Vol. 44, No. 2, 2002, 314–318.

Lindgren, A. and Chen, F. (2007). "State of the art analysis: An overview of advanced driver assistance systems (ADAS) and possible human factors issues," In C. Weikert (Ed.), *Proceedings of the Swedish Human Factors Network (HFN) Conference*. Linköping : Swedish Network for Human Factors, 38-40.

Lindgren, A., Chen, F., Jordan, P. W. and Zhang, H. (2008). " Requirements for the design of advanced driver assistance systems - The differences between Swedish and Chinese drivers," *International Journal of Design*, 2(2), 41-44.

Lundkvist, A., Nykänen, A. and Johnsson R. (2011). “ 3D-Sound in Car Compartments Based on Loudspeaker Reproduction Using Crosstalk Cancellation,” *Audio Engineering Society Convention 130, London, UK*, 1-11

Matsumura, T., Iwanaga, N., Kobayashi, W., Onoye, T., Shirakawa, I. (2005). “ Embedded 3D Sound Movement System based on Feature Extraction of Head-Related Transfer Function,” *IEEE Transactions on Consumer Electronics*, Vol. 51, No. 1, 262-263.

McGookin, D. and Brewster, S. (2011). “ Chapter 14: Earcons,” In Hermann T., Hunt A., Neuhoff J. G. (Eds),*The Sonification Handbook*,Logos Publishing House, Berlin, Germany, 339-361.

Mynatt, E. D. (1994). “Designing with Auditory Icons”, In Kramer G. and Smith S. (Eds), *Second International Conference on Auditory Display (ICAD ’94)*, Santa Fe, New Mexico, Santa Fe Institute, 109–119.

Nevile, M. and Haddington, P. (2010). “ In-car distractions and their impact on driving activities,” *Department of Infrastructure and Transport*, Australian National University, 2-6.

Nykänen, A. (2008). *Methods for Product Sound Design*, Doctorate Thesis in Division of Human Work Sciences, Division of Sound and Vibration, Luleå University of Technology, ISSN: 1402 - 1544, 1-17.

Parker, S. P. A., Eberle, G., Martin, R. L., and Mc Anally K. I. (2008). “ Construction of 3-D Audio Systems: Background, Research and General Requirements,” *Air Operations Division Defence Science and Technology Organisation*,DSTO-TR-2184, 1-12.

Patterson, R. D., Gaudrain, E. and Walters, T. C. (2010). “ Chapter 2 : The Perception of Family and Register in Musical Tones,” In Jones, M. R., Fay, R. R., and Popper, A. N. (Eds). *Music Perception.*, Springer Handbook of Auditory Research 36, Springer Science+Business Media, LLC, New York, 13-15.

Sikora, C.A., Roberts, L.A., and Murray, L.T. (1995). “ Musical vs. real world feedback signals,” *In Proceedings of CHI 95*, volume 2, Denver, USA, ACM Press, 220–221.

Stevens, A. (2000). “ Safety of driver interaction with in-vehicle information systems,” *Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering 2000*, 639-640.

Stevens, A., Quimby, A., Board, A., Kersloot, T. and Burns, P. (2002). “ Design guidelines for safety of in-vehicle information systems,” Project Record prepared for *TTT Division, DTLR*, TRL Limited, 1-2.

Stutts, J. C., and Hunter, W. (2003). " Driver inattention, driver distraction and traffic crashes," *Institute of Transportation Engineers Journal.* 73(7), 34.

Trentecoste, M. F. (2004). " Chapter 6: The auditory presentation of in-vehicle information," *In-Vehicle Display Icons and Other Information Elements Volume I: Guidelines*, Publication No. FHWA-RD-03-065, 80-83.

Weinrich, S. (1982). *The problem of front-back localization in binaural hearing*, Scand. Audiol. 11, Suppl. 15.

Wilkins, R. A., Martin, A. M. (1980). " The effects of hearing protection on the perception of warning sounds," In P. W. Alberti (Ed.), *Personal hearing protection in industry*, New York: Raven, 339-369.

Wilkins, P. A., Martin, A. M. (1987). " Hearing protection and warning sounds in industry-A review," *Applied Acoustics, 21*, 267-293.

Wilschut, E. S. (2009). *The impact of in-vehicle information systems on simulated driving performance : effects of age, timing and display characteristics : effects of age, timing and display characteristics*, Doctorate Thesis in Behavioural and Social Sciences, University of Groningen, Netherlands, 6-8.

Ziefle, M., Pappachan, P., Jakobs, E. M., and Wallentowitz, H. (2008). " Visual and Auditory Interfaces of Advanced Driver Assistant Systems for Older Drivers," In K. Miesenberger et al. (Eds.): ICCHP 2008, *Lecture Notes in Computer Science*, Vol. 5105, Springer 2008, 62-69.

9 Appendices

9.1 Appendix A : Tables

9.1.1 Sound localization - 1st jury group evaluation

Table 9.1: *Results from the localization of sound sources for all 20 sound sketches.*

Sketch No	Correct direction	P1	P2	P3	P4	P5	P6
1	LB30	***LF30***	LB30	L	***LF10***	LB20	LB20
2	LF20	LF10	***LB30***	LF20	LF20	LF10	L
3	RF20	RF45	***RB30***	R	R	R	R
4	RB30	***RF45***	R	R	RB20	RB30	R
5	RB60	***RF60***	RB30	***RF20***	***RF10***	RB30	RB20
6	LF20	LF45	L	LF10	LF10	***LB60***	L
7	LB30	***LF45***	***LF45***	***LF30***	***LF30***	LB30	L
8	LB60	L	LB60	***LF10***	LB10	LB45	LB45
9	RF20	RF45	***RB10***	RF30	RF10	***RB20***	R
10	LF20	LF45	***LB20***	L	L	LF10	L
11	R	-	RB20	R	R	R	R
12	L	LF45	LB30	L	LF10	L	L
13	L	LF45	L	L	L	LF10	LB20
14	R	RF20	RB10	RF20	R	RB20	RF30
15	R	RF30	RB45	RF20	R	RF10	R
16	L	L	LB30	LF20	LF10	LB45	L
17	R	RF30	R	RF30	RF30	R	R
18	L	LF45	L	L	LF10	LB45	L
19	L	LF30	LB30	L	LF10	L	L
20	R	R	RB10	RF45	R	RF20	R

Table 9.1 shows the results of the localization of the sound sources for 6 participants numbered from P1 to P6. Here, R, L, F, B, LF, RF, LB, RB represent right, left, front, back, left front, right front and the other directions, the numbers followed by these letters represent the angle in degrees from the reference axis i.e, the x-axis. For example, if the direction was selected to be LF20, the sound sources were assumed to be 20 degrees on the left-front side of the listener. The directions highlighted in bold and italics show the front-back confusion. Sketches 11 to 20 are not considered for front-back confusion because the sounds were not designed for front or back directions. Hence, the total percentage of front-back confusion was found to be $\frac{17}{60} = 28.3\%$

9.1.2 Sound localization - 2nd jury group evaluation

Table 9.2 shows the results for the localization of sound sources from the 9 participants (P1 - P9). Sketches 10, 11 and 12 were designed for left or right directions and hence were not considered for front-back confusion calculation. The total front-back confusion was found to be $\frac{12}{90} = 13.3\%$.

Table 9.2: *Results from the localization of sound sources for all the 13 sound sketches.*

Sketch No.	P1	P2	P3	P4	P5	P6	P7	P8	P9	Correct
1	R	R	R	R	R	***RF***	R	R	RB	RB
2	L	L	***LF***	LB	LB	L	L	L	L	LB
3	R	R	RB	R	R	***RF***	RB	***RF***	R	RB
4	L	LB	LB	L	LB	LB	LB	L	LB	LB
5	RB	R	R	RB	R	RB	RB	R	RB	RB
6	R	R	RB	RB	RB	RB	RB	R	RB	RB
7	L	LB	L	L	LB	L	L	L	L	LB
8	L	***LB***	L	L	L	LF	L	L	L	LF
9	L	L	***LB***	***LB***	***LB***	LF	***LB***	L	***LB***	LF
10	L	L	L	L	L	L	L	L	L	L
11	R	R	R	R	RF	RF	R	R	RF	R
12	R	R	RF	RF	RF	R	R	R	RF	R
13	L	L	L	L	***LB***	L	L	L	***LB***	LF

Table 9.3: *Results from the localization of sound sources for all the 24 participants.*

Sketch No.	1	2	3	4	5	6	7	8	9	10
P1	R	LB	RB	LB	RB	R	LF	R	L	L
P2	RB	***LF***	RB	L	RB	R	LF	R	LF	LF
P3	RB	LB	***RF***	L	***RF***	L	LF	RF	LF	LF
P4	***RF***	LB	***RF***	LB	R	RF	***LB***	RF	LF	LF
P5	R	LB	R	L	R	R	L	***RB***	L	L
P6	R	L	R	LB	R	R	L	***RB***	LF	L
P7	R	LB	RB	L	R	RB	LF	***RB***	LF	***LB***
P8	R	L	RB	LB	RB	R	L	R	***LB***	L
P9	R	LB	RB	LB	R	RF	LF	R	**LB**	***LB***
P10	R	LB	RB	LB	RB	RB	***LB***	RF	***LB***	***LB***
P11	R	L	R	LB	RB	RF	L	***RB***	L	LF
P12	R	***LF***	R	***LF***	RB	R	LF	R	L	L
P13	R	LB	RB	LB	RB	R	L	R	L	***LB***
P14	RB	LB	RB	LB	RB	RB	***LB***	R	***LB***	***LB***
P15	RB	***LF***	R	L	***RF***	R	L	R	L	LF
P16	R	***LF***	R	L	RB	RF	LF	R	LF	LF
P17	RB	L	RB	LB	RB	R	LF	R	L	***LB***
P18	R	LB	RB	***LF***	RB	RB	LF	***RB***	L	LF
P19	***RF***	***LF***	R	LB	RB	RF	LF	RF	L	LF
P20	R	L	RB	LB	R	R	L	R	L	L
P21	RB	LB	R	LB	R	R	L	R	L	L
P22	RB	LB	RB	LB	R	RB	L	R	L	L
P23	***RF***	LB	***RF***	LB	R	RF	LF	RF	L	LF
P24	R	L	RB	LB	***RF***	RF	LF	R	L	L
Correct	RB	LB	RB	LB	RB	R	LF	RF	LF	LF

9.1.3 Sound Localization - Final listening test

Table 9.3 shows the results for the localization of sound sources from the 24 participants (P1 - P24). Sketch 6 was designed for the right direction and was not considered for the calculation. The total front-back confusion was found to be $\frac{34}{24*9} = 15.8\%$.

9.1.4 Pair comparison of reverberant and non-reverberant sounds

Tables 9.4 and 9.5 shows the results from the choice between reverberant and less or not reverberant sounds. The letters in brackets represent the order of reverberant-not reverberant sounds in each pair; number 1 stands for the first sound, and number 2 stands for the second sound. Therefore, the first sound could either be a reverberant sound (represented by R) or a not reverberant sound (represented by NR) and the number represents the answer selected.

Table 9.4: *Results from the pair comparisons of reverberant/less reverberant sounds for the first 5 earcon pairs.*

Participant No.	Pair 1 (R-NR)	Pair 2 (R-NR)	Pair 3 (NR-R)	Pair 4 (R-NR)	Pair 5 (R-NR)
1	2	2	1	1	1
2	2	2	1	1	2
3	1	1	1	2	1
4	2	2	1	2	2
5	1	1	1	1	2
6	1	1	1	2	1
7	2	1	2	2	1
8	2	1	2	1	1
9	2	1	1	2	2
Winner	2	1	1	2	1
% win	66.7	66.7	77.8	55.6	55.6

Table 9.5: *Results from the pair comparisons of reverberant/less reverberant sounds for the remaining 5 earcon pairs.*

Participant No.	Pair 6 (R-NR)	Pair 7 (NR-R)	Pair 8 (NR-R)	Pair 9 (R-NR)	Pair 10 (NR-R)
1	2	1	1	2	2
2	1	1	1	1	2
3	2	1	1	2	1
4	2	1	1	2	1
5	2	1	1	2	1
6	2	1	1	2	1
7	2	2	1	1	2
8	2	1	1	1	2
9	2	1	1	2	1
Winner	2	1	1	2	1
% win	88.9	88.9	100	66.7	55.6

9.2 Appendix B : Questionnaires

9.2.1 First jury group evaluation

Sound evaluation for Master's thesis by Vijendra and Yidan
Thanks for participating in our sound evaluation!

Instructions:
This sound evaluation is divided into three sections. First is the 3D sound demonstration. In this section, you don't need to answer any questions. In the second and third section, you will have 10 sounds each played for you and you need to answer four questions for each sound. Each sound will be played once first, and if you want to listen to it again, just let us know. Between the second and third section, you will have five minutes break. Please feel free to write any comments for each sound.

Participant background questionnaire:
Gender: Male/Female
Age: _______
Do you have any hearing impairments? Yes/No
If yes, please mention here: __________

First Section
3D sound demonstration: Please listen to the sound and feel the 3D effect.

Second Section/Third Section

Please listen to each sound and answer the questions below:

Sound 1

1. What event do you think this sound is best suitable for?

- Blind Spot Information System
- Lane Keeping Warning
- Other early warning systems (Please state why)

  ____

2. Please mark on these scales based on the questions. Please do **not** mark between the numbers.

How does the sound make you feel?

Calm |_____|_____|_____|_____|_____|_____|_____|_____| Alert
-4 -3 -2 -1 0 1 2 3 4

Unhappy |_____|_____|_____|_____|_____|_____|_____|_____| Happy
-4 -3 -2 -1 0 1 2 3 4

What do you think about the sound?

Not Likable |_____|_____|_____|_____|_____|_____|_____|_____| Likable
-4 -3 -2 -1 0 1 2 3 4

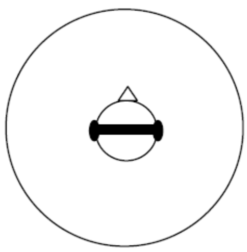
Inappropriate |_____|_____|_____|_____|_____|_____|_____|_____| Appropriate
-4 -3 -2 -1 0 1 2 3 4

Worthless |_____|_____|_____|_____|_____|_____|_____|_____| Assisting
-4 -3 -2 -1 0 1 2 3 4

Does the sound remind you of a significant event? Please mark on the scale.

Insignificant |_____|_____|_____|_____|_____|_____|_____|_____| Significant
-4 -3 -2 -1 0 1 2 3 4

3. Where do you think the sound came from? Please put a 'X' mark anywhere **on** the circle. The center of the circle represents the listener's head (top view).



4. What do you think can be changed in the sound to make it appropriate in a vehicle environment?

____

Comments: ____

9.2.2 Second jury group evaluation

Sound evaluation 2 for Master's thesis by Vijendra and Yidan
Thanks for participating in our sound evaluation!

Instructions:
This sound evaluation is divided into 4 sections.
1st section: 13 sounds, 3 questions for each sound.
2nd section: 13 sounds, 2 questions for each sound.
3rd section: 10 pair of sounds, 1 question per sound, 20 sounds in total.
4th section: General discussion of 13 sounds.

Each sound will be played once first, and if you want to listen to it again, please let us know.

Participant background questionnaire:

Gender: Male/Female
Age: _______

Do you have any hearing impairments? Yes/No
If yes, please mention here: ____________________

How often do you drive a car?

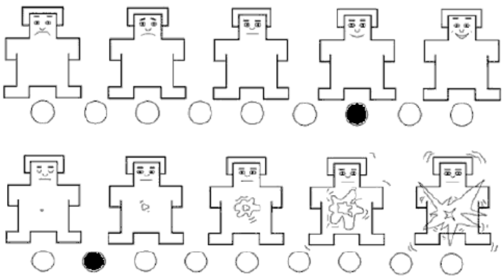
- O Everyday
- O 1-3 times a week
- O 1-3 times a month
- O 1-3 times per 6 months
- O 1-3 times a year
- O Less than 1 time a year

Example Questions:
What event do you think this sound is best suitable for? Please mark on the event.

○ Seatbelt Warning
○ Blind Spot Warning
● Low Fuel Warning
○ Forward Collision Warning
○ Over Speed Warning
○ Lane Departure Warning

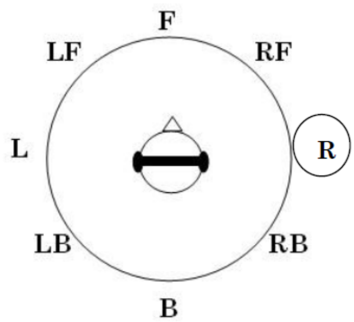
Please mark on these scales based on the questions. Please do **not** mark between the circles.
How does the sound make you feel?

- The first scale refers to human emotions ranging from **sad** to **happy**. The face to the extreme left means you are sad, disappointed (something makes you feel unpleasant). The face to the extreme right means you are happy (something makes you feel pleasant). The neutral face in the middle means a neutral expression.
- The second scale refers to emotions ranging from **calm** to **aroused.** The face to the extreme left means you are "not activated" (E.g. sleepy, calm). The face to the extreme right means you are "activated" (E.g. frightened or surprised by something). The face in the middle stands for a neutral expression.



Where do you think the sound came from? Please mark on the picture. The center of the picture stands for the top view of the listener (with headphones).

L - Left
R - Right
F - Front
B - Back
LF - Left side Front
LB - Left side Back
RF - Right side Front
RB - Right side Back



What do you think about the sound?

Not Likable	-4	-3	(-2)	-1	0	1	2	3	4	Likable
Inappropriate	-4	-3	-2	-1	0	(1)	2	3	4	Appropriate
Worthless	-4	-3	-2	(-1)	0	1	2	3	4	Assisting
Not Urgent	-4	-3	-2	-1	0	1	2	(3)	4	Urgent

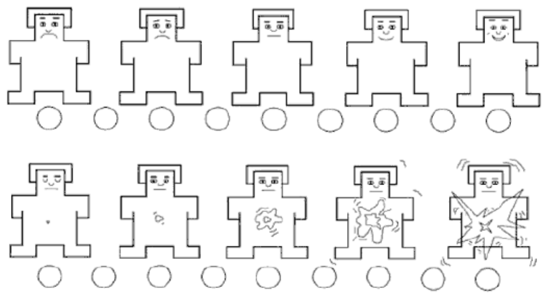
--SOUND CHECK ON HEADPHONES--

Please turn to the next page to start the sound evaluation.

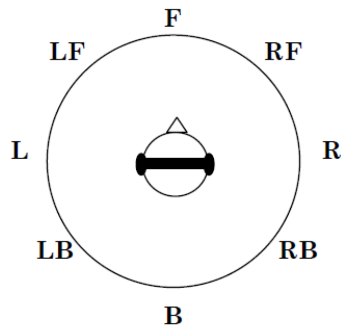
Section 1
Please listen to each sound and answer the questions below:

Sound 1
1. What event do you think this sound is best suitable for? Please mark on the event.
- O Seatbelt Warning
- O Blind Spot Warning
- O Low Fuel Warning
- O Forward Collision Warning
- O Over Speed Warning
- O Lane Departure Warning

2. How does the sound make you feel? Please do **not** mark between the circles.



3. Where do you think the sound came from? Please mark on the picture.



Section 2
Please answer the questions based on the event.

Sound 1

1. If this sound is designed for Blind Spot Warning, please answer the following questions:

What do you think about the sound?

	-4	-3	-2	-1	0	1	2	3	4	
Not Likable										Likable
Inappropriate										Appropriate
Worthless										Assisting
Not Urgent										Urgent

2. If this sound is designed for Lane Departure Warning, please answer the following questions:

What do you think about the sound?

	-4	-3	-2	-1	0	1	2	3	4	
Not Likable										Likable
Inappropriate										Appropriate
Worthless										Assisting
Not Urgent										Urgent

Comments: ____.

Section 3
Please listen to two sounds in each question and **choose the one you preferred as a warning sound in vehicle environment.**

Pair 1
Which sound do you prefer to be used as a warning sound in a vehicle?
- Sound 1
- Sound 2

Pair 2
Which sound do you prefer to be used as a warning sound in a vehicle?
- Sound 1
- Sound 2

Pair 3
Which sound do you prefer to be used as a warning sound in a vehicle?
- Sound 1
- Sound 2

Pair 4
Which sound do you prefer to be used as a warning sound in a vehicle?
- Sound 1
- Sound 2

Pair 5
Which sound do you prefer to be used as a warning sound in a vehicle?
- Sound 1
- Sound 2

Pair 6
Which sound do you prefer to be used as a warning sound in a vehicle?
- Sound 1
- Sound 2

Pair 7
Which sound do you prefer to be used as a warning sound in a vehicle?
- Sound 1
- Sound 2

Pair 8
Which sound do you prefer to be used as a warning sound in a vehicle?
- Sound 1
- Sound 2

Pair 9
Which sound do you prefer to be used as a warning sound in a vehicle?
- Sound 1
- Sound 2

Pair 10
Which sound do you prefer to be used as a warning sound in a vehicle?

- ○ Sound 1
- ○ Sound 2

This is the end of the evaluation. Please turn to section 2 of this questionnaire so we can start with the discussion of sounds and you can review your answers.