Perception and Performance: An Evaluation of Multimodal Feedback for the Assessment of Curve Shape Differences

JACQUELINE HOLLOWOOD, BA (HONS), MSc.

Thesis submitted to the University of Nottingham
for the degree of Doctor of Philosophy

December 2011

Abstract

The EU-funded SATIN project sought to provide a multimodal interface to aid product designers in judging the quality of curved shapes. This thesis outlines a research programme designed to assist in the exploration of fundamental issues related to this project, and provide a means to evaluate the success of such interfaces more generally. Therefore, three studies were undertaken with the aim of exploring the value of haptic and sound feedback in the perception of curve shape differences, and through the knowledge gained provide an evaluative framework for the assessment of such interfaces.

The first study found that visual, haptic, and visual-haptic perception was insufficient to judge discontinuities in curvature without some further augmentation. This led to a second study which explored the use of sound for conveying curve shape information. It was found that sine waves or harmonic sounds were most suited to for this task. The third study combined visual-haptic and auditory information. It was found that sound improved the perception of curve shape differences, although this was dependent upon the type of sonification method used. Further to this, data from studies one and three were used to identify gradient as the active mechanism of curve shape differentiation and provided a model for the prediction of these differences. Similarly performance data (response time, accuracy, and confidence) were analysed to produce a model for the prediction of user performance at varying degrees of task difficulty.

The research undertaken across these studies was used to develop a framework to evaluate multimodal interfaces for curve shape exploration. In particular a ‘discount’ psychophysical method was proposed, along with predictive tools for the creation of perceptual and performance metrics, plus guidelines to aid development. This research has added to fundamental knowledge and provided a useful framework through which future multimodal interfaces may be evaluated.

Acknowledgements

I would like to thank all those in the Human Factors Research Group who have been supportive through this process, and who have made my time within the group enjoyable. In particular, I would like to thank those colleagues who were involved in the SATIN project and who helped in the collection of data for Study 2; Dr Michael Pettitt, Glynn Lawson, Dr Alyson Langley, and Anne Floyde. My thanks also to the project partners who have been generous with their time and greatly increased my understanding, with a special thanks to Dr Simon Shelley. Thanks too to all those who participated in the various studies associated with this research. In addition, I would like to thank the project coordinator, Professor Monica Bordegoni, and all those associated with the creation of this project through which many research opportunities were presented.

I would like to thank Dr Sue Cobb, my supervisor, whose advice and encouragement have been invaluable in the development and completion of this research. I would also like to thank Dr Sarah Sharples, my second supervisor, who has lent a guiding hand throughout, and whose help was particularly appreciated during writing-up. In addition, I would like to thank Dr Alex Stedmon for his brief involvement early on and his good company on site visits.

I have benefited greatly from the support of friends and family throughout this process, and to them all I give a great deal of thanks. In particular, I would like to thank; Dr Helen Hill, who has been a great support and source of advice; Maria Korea, for endless coffee and somewhere to hide; Dr Karen Parkhill, for ensuring that software devilment did not get the better of me in the closing stages, and for supplying Welsh Cakes. I would also like to give a special thanks to Jacky Graining who saw the possibilities, nurtured and encouraged, and whose abiding belief helped me this far.

Finally, I would like to give immense thanks to my partner, Gemma Gregory, without whom none of this would have been possible, and to whom I owe a huge debt of gratitude for her love, support, and forbearance. However, I am sure the guarantee of a return to a normal existence, and walks in the hills together, may in some small way begin to repay that debt.

List of Contents

Abstract i

Acknowledgements ii

List of Contents iii

Glossary xi

Chapter 1: Introduction 1

1.1 Context of the Research 1

1.1.1 The needs of Industrial and Product Designers 1

1.1.2 The Evaluation-Modification Loop 2

1.1.3 The Importance of Curves 4

1.1.4 The SATIN Prototype 5

1.1.5 Preliminary Evaluation of the SATIN Prototypes 7

1.2 Definition of the Problem 12

1.2.1 Perception 12

1.2.2 Performance 12

1.2.3 Interaction 12

1.2.4 Sound 13

1.3 Research Aims and Objectives 13

1.4 Structure of the Thesis 14

Chapter 2: Literature Review 20

2.1 Scope of Review 20

2.2 The use of Multimodal Technologies in Product Design 20

2.2.1 Design and Technology 20

2.2.2 The Perception of Digital Models in the Virtual Environment 22

2.3 The Nature of Haptic Perception 23

2.4 Haptic Perception of Objects 24

2.4.1 Geometric Properties 25

2.4.2 Material Properties 27

2.4.3 Temperature 27

2.4.4 Texture 27

2.5 Perception of Curvature 28

2.5.1 What is Curvature? 28

2.5.2 Relationship of Shape and Curvature 29

2.5.3 Investigating Perception 30

2.5.4 Absolute and Difference Thresholds 33

2.5.5 Mechanism of Curvature Perception 35
2.5.6 Factors affecting Curvature Perception 40
2.6 Multimodality 41
2.6.1 Visual-Haptic 41
2.6.2 Visual-Auditory 42
2.6.3 Haptic-Auditory 42
2.6.4 Issues of Multimodality 43
2.7 Sound and Sonification 44
2.7.1 Properties and Perception of Sound 44
2.7.2 Naturalistic and Abstract Sounds 45
2.7.3 Sonification, Applications, and Human Factors Issues 46
2.8 Evaluation and Guidelines for Multimodal Interfaces 47
2.8.1 Definition, Methods, and Objective Measurement of Usability 48
2.8.2 Guidelines for interface development 49
2.9 Focus of the Thesis in Relation to the Literature 50
Chapter 3: Visual-Haptic Perception and Performance 53
3.1 Introduction and Rationale 53
3.1.1 Shape Assessment, Modification, and the Perception of Curvature Differences 53
3.1.2 Performance Measures and the Development of Metrics 55
3.1.3 Aims 56
3.2 Method (Study 1) 57
3.2.1 Participants 57
3.2.2 Equipment and Set up 57
3.2.3 Procedure 57
3.2.4 Stimuli 58
3.2.5 Design and Hypotheses 59
3.2.6 Data Analysis 62
3.3 Results (Study 1) 63
3.3.1 Perception 63
3.3.2 Performance 64
3.3.3 Performance Characteristics 72
3.3.4 Interaction 76
3.4 Discussion (Study 1) 79
3.4.1 Perception 79
3.4.2 Performance 81
3.4.3 Performance Characteristics 83

3.4.4 Interaction ... 84
3.5 Summary of Chapter ... 85
3.6 Related Chapters ... 86
Chapter 4: Prediction JNDs and Associated Level of Performance ... 87
4.1 Introduction ... 87
4.2 The Value of a Predictive Model ... 87
4.2.1 Ensuring Perceptible Feedback ... 87
4.2.2 Objective Evaluation Criteria ... 88
4.2.3 Provision of a Cost-effective Method ... 89
4.3 The Limitations of using Curvature to Predict Threshold and Performance ... 89
4.4 Alternative Theoretical Approach derived from the Literature ... 90
4.5 Further Analysis of Perception Data ... 92
4.5.1 Data Analysis ... 92
4.5.2 Results ... 92
4.5.3 Discussion ... 94
4.6 Summary of Chapter ... 95
4.7 Related Chapters ... 96
Chapter 5: Perception of Curve Orientation and Magnitude from Sound ... 97
5.1 Introduction and Rationale ... 97
5.1.1 Study Structure ... 97
5.1.2 Exploring shape with sound ... 98
5.1.3 Types of Sound ... 99
5.1.4 Workload ... 100
5.2 Conveying Orientation with Sound (Study 2 - Experiment 1) ... 101
5.2.1 Aims ... 101
5.2.2 Method ... 101
5.2.3 Results ... 106
5.2.4 Discussion ... 112
5.3 Conveying Magnitude with Sound (Study 2 - Experiment 2) ... 117
5.3.1 Aims ... 117
5.3.2 Method ... 117
5.3.3 Results ... 121
5.3.4 Discussion ... 135
5.4 Chapter Summary ... 141
5.5 Related Chapters ... 141
Chapter 6: Sound in an Applied Context ... 142
6.1 Introduction and Rationale ... 142

6.1.1 The Nature of a Curve in Reality ... 142
6.1.2 Aims ... 143
6.2 Method (Study 2 – Experiment 3) ... 143
6.2.1 Participants ... 143
6.2.2 Equipment and Setup ... 143
6.2.3 Procedure ... 144
6.2.4 Task Object ... 145
6.2.5 Sonification ... 146
6.2.6 Design and Hypotheses ... 147
6.2.7 Data Analysis ... 148
6.3 Results (Study 2 – Experiment 3) ... 148
6.3.1 Performance ... 148
6.3.2 User Experience ... 151
6.3.3 Workload ... 152
6.3.4 Background ... 154
6.3.5 Interaction Strategies ... 155
6.4 Discussion (Study 2 – Experiment 3) ... 156
6.5 Chapter Summary ... 160
6.6 Related Chapters ... 161
Chapter 7: The Effect of Sound on Visual-Haptic Perception, Performance, and Interaction ... 162
7.1 Introduction and Rationale ... 162
7.1.1 How this Differs from Studies 1 and 2 ... 162
7.1.2 The Stimuli ... 163
7.1.3 Uniformity of Magnitude Difference ... 165
7.1.4 Performance Characteristics ... 165
7.1.5 Observing Interaction ... 166
7.1.6 Aims ... 167
7.2 Method (Study 3) ... 167
7.2.1 Participants ... 167
7.2.2 Equipment and Set Up ... 168
7.2.3 Procedure ... 169
7.2.4 Stimuli ... 171
7.2.5 Sonification ... 173
7.2.6 Design and Hypotheses ... 175
7.2.7 Data Analysis ... 177
7.3 Results (Study 3) ... 177

7.3.1 Perception ... 177
7.3.2 Performance ... 179
7.3.3 Performance Characteristics ... 192
7.3.4 Interaction ... 195
7.3.5 User Experience ... 196
7.4 Discussion (Study 3) ... 201
7.4.1 Perception ... 201
7.4.2 Performance ... 202
7.4.3 Interaction ... 207
7.4.4 User Experience ... 210
7.5 Chapter Summary ... 212
7.6 Related Chapters ... 213
Chapter 8: Prediction of Perception and Performance ... 214
8.1 Introduction and Rationale ... 214
8.1.1 Experimental Construct ... 214
8.1.2 Aims ... 215
8.2 Further Analysis of Study 3 Data ... 215
8.2.1 Predicted Threshold Gradients ... 215
8.2.2 Design and Hypothesis ... 216
8.2.3 Data Analysis ... 217
8.3 Results ... 217
8.3.1 Perception ... 217
8.3.2 Performance ... 220
8.3.3 Performance Predictability – Multiple Regression Analysis ... 225
8.4 Discussion ... 230
8.4.1 Perception ... 230
8.4.2 Performance ... 232
8.4.3 Predictability ... 233
8.5 Summary of Chapter ... 236
8.6 Related Chapters ... 236
Chapter 9: Evaluation of Multimodal Interfaces for Shape Exploration – An Emergent Framework ... 237
9.1 About this Chapter ... 237
9.2 Evaluation Method ... 238
9.2.1 The Use of Traditional Psychophysics ... 238
9.2.2 Is ‘Discount’ Psychophysics Possible? ... 239
9.2.3 Number of Participants and Trials ... 241

9.1.1 Choosing between Traditional and Discount Methods 245
9.3 The Evaluation of Multimodal Interfaces using Difference Threshold and Performance Metrics 246
9.3.1 Choice of Variables 246
9.3.2 Type of Evaluation, Comparative Data, and Predictive Models 247
9.4 Characterising Performance and its use in Evaluating Multimodal Interfaces 248
9.4.1 Performance Characteristics 248
9.4.2 Feedback & User Error 249
9.5 Observing Interaction 251
9.5.1 Direct Observation 251
9.5.2 Mediated Observation 252
9.5.3 Recommendations for Observing Multimodal Interaction 253
9.6 Guidelines for the Development of Multimodal Interfaces 254
9.7 Chapter Summary 255
9.8 Related Chapters 256
Chapter 10: General Discussion 257
10.1 Perception & Multimodality 257
10.1.1 Unimodal versus Bimodal Exploration of Curvature Differences 257
10.1.2 Bimodal versus Multimodal Exploration of Curvature Differences 258
10.1.3 Differences between Intra- and Inter-Object Judgements 260
10.1.4 Effective Stimulus for Difference Perception and Prediction of JNDs 261
10.1.5 Implications for interface design 262
10.2 Performance, Multimodality and other Affective Factors 263
10.2.1 Modality 264
10.2.2 Stimulus Properties 266
10.2.3 Magnitude Difference 267
10.2.4 Sound Type and Sonification 268
10.2.5 Prediction of Performance 270
10.3 Interaction 271
10.3.1 Styles 271
10.3.2 The Effect of Sound 272
10.3.3 Haptic Duality 273
10.4 Sound and Sonification 274
10.4.1 Perceptual Capabilities 274
10.4.2 Musical and Cognitive Abilities 275
10.4.3 Training 278
10.5 The value of this research for SATIN and future projects 278

Chapter 11: Conclusions and Further Work ... 281
11.1 Conclusions ... 281
11.1.1 Identify the effect of combining haptic, visual, and sound feedback on perception, performance, and interaction. ... 281
11.1.2 Inform the Theory and Develop a Predictive Model based on JNDs ... 282
11.1.3 Assess the Appropriateness of Sound for Communicating Curve Shape and Curvature Information ... 283
11.1.4 Develop Appropriate Performance Metrics for Evaluation of Curvature using Unimodal and Multimodal Feedback ... 283
11.1.5 Provide a Framework for the Evaluation of Virtual Interfaces for Curve Shape Exploration ... 285
11.2 Research Contribution ... 285
11.3 Limitations and Future Work ... 287
11.3.1 Perception ... 287
11.3.2 Performance ... 288
11.3.3 Interaction ... 288
11.4 Final Comments ... 289
References ... 290
Appendix A : Geometric Relationships between Various Stimulus Dimensions ... 301
Appendix B : Study 1 Experiment Documents ... 302
Appendix C : Stimuli Specification (Studies 1 and 3) ... 303
Appendix D : Summary Results ... 305
Study 1 ... 305
Perception (H1-H3) ... 305
Study 2 – Experiment 1 (Chapter 5) ... 306
Performance ... 306
Practice ... 306
User Experience ... 306
Workload (H24-H30) ... 306
Study 2 - Experiment 2 (Chapter 5) ... 306
Performance ... 306
User Experience Issues ... 307
Workload (H25-31) ... 307
Study 3 - Experiment 3 (Chapter 6) ... 308
Performance (H1-H3) ... 308
User Experience Issues ... 308
Workload ... 309

Workload ... 309
Study 3 (Chapter 7) ... 309
Perception ... 309
Performance (H4-H21) ... 309
Performance Characteristics ... 309
Interaction ... 310
User Experience ... 310
Workload (H56-H62) ... 310
Study 3 (Further Analysis – Chapter 8) ... 310
Perception (H1-H2) ... 310
Performance ... 310
Predictability (H12-H20) ... 311
Appendix E : Psychophysical data for Haptic, Visual, and Visual-Haptic Modalities in Low and High Curvature Conditions ... 312
Calculation of Threshold ... 312
Limitations of the Method ... 312
Regression and Psychophysical Data ... 313
Appendix F : Non-Normal Data and ANOVA ... 315
Appendix G : Study 2 Experiment Documents ... 316
Appendix H : Eowave Sensor Data Sheet ... 317
Appendix I : Study 3 Experiment Documents ... 318
Appendix J : Regression and Psychophysical Data (Study 3) ... 319
No Sound ... 319
Curve Shape ... 320
Curvature (original) ... 321
Curvature (adjusted) ... 322
Appendix K : Performance Characteristics – Correlation Statistics ... 323
Appendix L : Participants Subjective Views about Conflict of Senses ... 324
Appendix M : Regression and Psychophysical Data for Individual Thresholds (Study 3 – Further Analysis) ... 325
Appendix N : ANOVA Weber Fractions (gradient) Stimulus B and C (Study 1) ... 326
Appendix O : The Effect of Trial Group Position on Percentage of Correct Judgements ... 328
Background to Data ... 328
Data Analysis ... 328
Results ... 328

Glossary

Absolute Threshold	The level at which a physical event can be perceived (this is also known as 'detection' threshold).
Acuity	The level of sensitivity to a given stimulus.
Attitude Difference	This is a term used by Pont et al (1997), and is the sum of 'Local Attitude' (see below for definition) within the stimuli. As a stimulus has only two local attitudes, and these have the same angle, then the attitude difference is twice the local attitude.
Augmented Reality	Refers to the over-laying of real objects with computer-generated information in order to aid or enhance our experience. This may be delivered in a visual or auditory form, and tends to be contextual to the task being undertaken.
Base-to-Peak Height	The measure from the chord mid-point to the apex (of a circle). This is illustrated in Figure 2-4.
CAD	Computer-Aided Design.
Curvature	The mathematical expression of the amount of curvedness of a line. This is given as $k=1/r$, where k is the curvature at a given point and r is the radius of an imagined circle through that point. This is illustrated in Figure 2-2.
Curvature Continuity	The curvature at the point where two surfaces join is the same.
Curvature Gradient	This is more usually referred to as the tangent and should not be confused with gradient (see below for a definition).

	Both these are illustrated in Figure 2-4, and should be viewed to appreciate the difference in these measures.
Difference Threshold	The amount beyond which an increase (or decrease) in the stimulus will be perceived 75% of the time.
Gradient	The ratio of the height over the width. That is, it gives the amount by which the height increases for each unit of width. The gradient was identified by Davidson (1972) as the mechanism of curvature detection. For any given curvature a particular gradient can be calculated between two points. The gradient is calculated by dividing ½ chord length by the base-to-peak height (from the chord mid-point to the circle apex). This is illustrated in Figure 2-4.
JND	A ‘Just Noticeable Difference’ is the amount by which a stimulus must change in order for it to be perceived as different.
Local Attitude	A term defined by Pont et al (1997). It refers to the angle formed between the chord and the tangent at the intersection of the curve and chord. This is best understood diagrammatically and is illustrated in Figure 2-4.
Magnitude Difference	The extent to which a dimension varies.
Modality	A sense through which humans perceive; touch, vision, hearing, smell, and taste.
Multimodal Interface	An interface that uses more than one mode (or sense) to provide a means of communication between the computer and user. Typically there is provision of visual, haptic, and auditory input and output devices which form the interface.
Perception	The ability to receive sensations from stimuli within the environment.

Psychophysics	An area of science that seeks to quantify the relationship between physical events and psychological (subjective) responses.
SATIN	Sound And Tangible Interfaces for Novel product design (SATIN) project. This developed a multimodal interface for the exploration and modification of curved shapes.
Shape	The geometric boundary of object in two or three dimensions. For example, four straight lines whose ends intersect at right angles to each other on a flat surface (2 dimensions) demarcate a square. So, the shape, a square, is the boundary created by a particular geometric configuration.
Sonification	This relates to the mapping of a sound to a property or parameter of an object in order to convey information about it.
Stimulus	This takes the form of a physical event or entity that might be perceived by the human senses. In relation to this thesis the stimulus of primary interest is a physical entity; curve shape.
Stimulus Gradient	This is the gradient associated with the stimulus. It is calculated by dividing ½ chord length by the base-to-peak height of the curve shape. This is illustrated in Figure 4-2. The stimulus itself is associated with a curved block of various dimensions – one of these being the gradient. Other dimensions of the stimulus are curvature, and width, and would give rise to the phrases such as stimulus width and stimulus curvature.
Stimulus Width	This has two meanings. The first is a measure of the extent of the stimulus from side to side along the widest length. The second is used when calculating gradient and

	is taken as the distance between the stimulus end and mid-point.
Threshold Gradient	This is the gradient associated with the threshold. The threshold is a mathematically calculated point at which 75% of the time a comparison stimulus is perceived as different to the originally presented stimulus. It is therefore the gradient at which a difference from the original stimulus gradient is perceived. It is calculated by dividing ½ chord length by the base-to-peak height of the comparison curve shape. This is illustrated in Figure 4-2.
Virtual Environments	Are computer-generated surroundings which mimic the real-world or an imagined space.
Virtual Prototyping	A prototype is a draft version of an envisaged product. A virtual prototype is one that is computer-generated, and does not exist in a physical form.
Weber Fraction	This is the JND as a percentage of the original stimulus.
Weber's Law	This asserts that the amount by which a stimulus needs to change, in order to be perceived as different, is a fixed proportion of the initial stimulus.

Chapter 1: Introduction

1.1 Context of the Research

This research has been undertaken as part of the EU funded Sound And Tangible Interfaces for Novel product design (SATIN) project (FP6-IST-5-054525), which developed a multimodal interface for the exploration and modification of curved shapes. There was a need within the project to understand fundamental aspects of perception, performance, and interaction. This chapter outlines the background context that informed the research focus and the aims and objectives that were derived as a result of this. It then goes on to show how the research has contributed to human factors knowledge, and finally describes the structure of this thesis.

1.1.1 The needs of Industrial and Product Designers

The SATIN project included three end-user partners; Alessi, ItalDesign (IDG), and Steklarna Hrastnik (SH), all of whom are involved in various types of product design. The artefacts produced by these end-users range from small decorative bottles to large vehicles, and are manufactured from a wide range of materials including glass, plastic, and steel. The aim of SATIN was to produce a prototype interface for free-hand haptic exploration of shape that would allow assessment of geometric qualities through auditory feedback.

During the early stages of the SATIN project, a number of activities were undertaken by project partners in order to elicit the needs of these end-users in relation to their design activities and the proposed multimodal interface. As part of these activities the author, along with a colleague from the University of Nottingham, undertook a visit to each end-user site. The purpose of these visits was to characterise the users, identify their tasks and understand the context in which these tasks took place. In addition, the multimodal nature of these activities and the proposed interface were explored.

The findings of these visits were included SATIN deliverable D3 and can be found on the accompanying CD (SATIN Consortium 2007). In summary, it was found that despite the range of activities and contexts there were many similarities amongst the end users. The difference between them was largely in the extent to which various characteristics were applied. For example, each would produce prototypes, however the number, type, and production time of these varied from one end-user to another. The use of prototyping was found to be central to all end-users' processes and was an increasing requirement. It was used for aesthetic and technical judgements about the quality of the object's shape. However, whilst useful, the production of physical prototypes was time consuming and expensive. There was therefore interest in the idea of the SATIN interface for the production of virtual prototypes upon which these types of assessment could be made more economically. It was also found that, whilst design processes varied in detail, a generic process could be discerned within these activities (see Figure 1-1 below). In relation to multimodality, it was found that visual and touch inspection of prototypes was considered important. However, most end-users found it difficult to envisage the use of sound other than in a naturalistic way, for example tapping an object to hear the noise made and thus inform an understanding of its density.

1.1.2 The Evaluation-Modification Loop

The end-user site visits discussed above revealed that similar design processes were employed by all companies involved in the SATIN project, and it was therefore possible to propose a generic model that encompasses these. Figure 1-1 shows a high-level model of the task flow and is common to all the SATIN end-user companies.

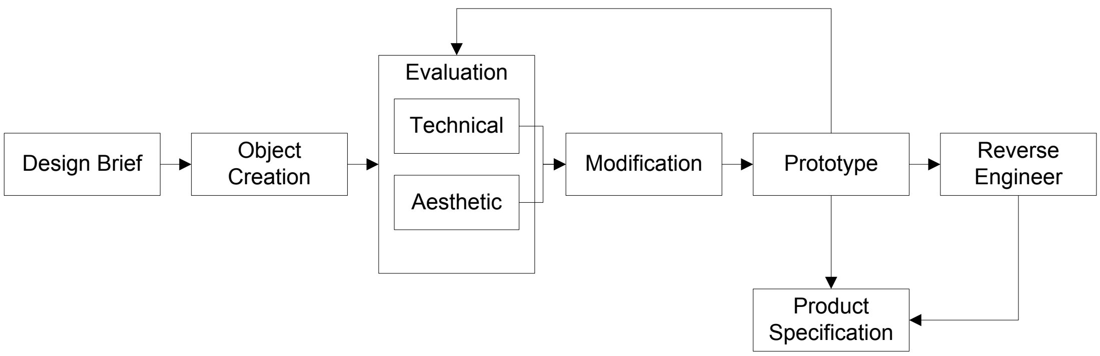


Figure 1-1: Generic Model of Design Workflows

The main focus of this task flow was the evaluation-modification loop. This was the point at which the object created was evaluated, modifications made, then re-evaluated in the light of those modifications. The object remained in this modification-evaluation loop until it was deemed satisfactory; that is it had achieved the desired objectives. The evaluations undertaken were both technical and aesthetic. It is therefore possible for some of these evaluative judgements to be made on a CAD model (virtual object) whilst other judgements require that the object is physically rendered. The process of providing a physical object makes the evaluation-modification loop time consuming and inefficient.

This is because each time a modification is made to a physical model the modified shape needs to be digitised in order to produce a revised virtual object. The SATIN project identified a more time and cost effective approach; to present a virtual object that could be directly manipulated by the designer. So by providing a haptic interface and sound feedback it was hoped that evaluation and modification could be undertaken purely on a virtual object without a need to produce a physical model.

The looping process undertaken as described above maps onto Norman's (1988) 'Seven Stages of Action' (see Figure 1-2). This model seeks to describe goal oriented action within the real world, so appropriately describes the actions that a designer might undertake. For the product designer, their aim is to achieve a particular form with certain properties. Their actions upon a given shape seek to modify the object to conform to this preconceived goal. However progress towards this goal can only be achieved if the results of their actions can be perceived and evaluated.

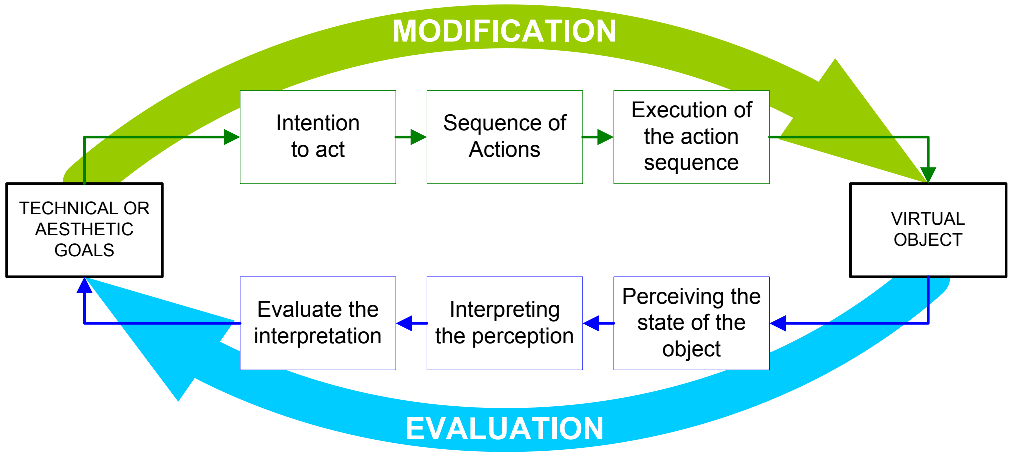


Figure 1-2: The Evaluation-Modification Loop. Norman's Seven Stages of Action applied to Product Design

How well the evaluation-modification loop works will be dependent to some extent on what Norman (1988) terms the 'gulfs of execution and evaluation'. In order to execute the envisaged modifications the interface must be able to support the user's desired action. Likewise, in order for evaluation to be successful the interface must give perceptible feedback.

1.1.3 The Importance of Curves

Automotive and consumer products are designed using Computer-Aided Design (CAD) software utilising a method known as surface modelling. Using this method the CAD technician or designer constructs the desired object from a series of surface patches. Each of these surfaces has its shape controlled by a number of points (control points) that define the curve path about that point. The points use a number of different mathematical formulations in order to control the path of the curve about these points; one such is known as B-splines. This type gives its name to the Non-uniform rational B-spline or NURBS surface that is predominately found in the CAD software used by product designers. By controlling curvature the shape of the surface is defined. This is illustrated in Figure 1-3 where it can be seen that the two principal curvatures define the shape of the surface.

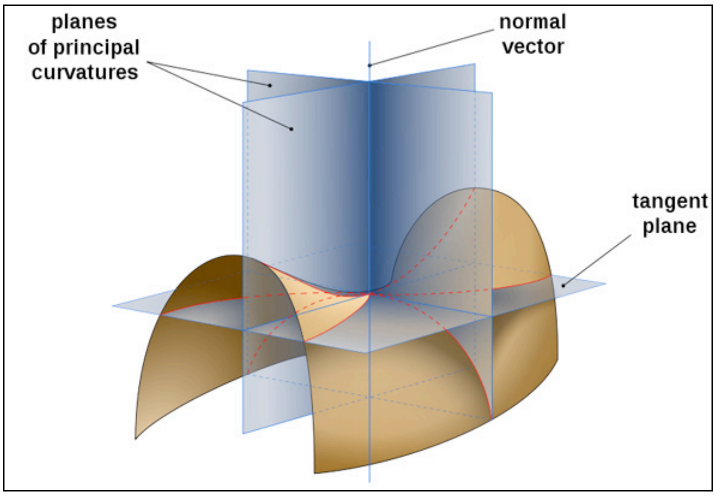


Figure 1-3: Relationship of curvature to shape

In order to create an object, a number of these surface patches are joined together. The way in which these patches are joined is important as it can affect the quality of the resultant object's surface in terms of their aesthetic and reflective properties. Automobiles and other high quality products require a Class-A surface. This means that the surface patches are joined in such a way that they have what is known as G2 or Curvature Continuity. This is best described in relationship to the other types of continuity as follows:

- Positional Continuity (G0) – the end points of two curves coincide, that is they share the same position (within 0.001mm).
- Tangential Continuity (G1) – the points have G0 continuity, plus the curves at the point of intersection have parallel tangents (within 0.01°).

- Curvature Continuity (G2) – the points have G0 and G1 continuity, plus the curvature for both points at this position should be the same (within 0.01/m).

Designers or technicians evaluate the CAD models to ensure that the objects produced have G2 continuity. This assessment is made by means of visual inspection tools such as ‘zebra’ lines. It was envisaged within the SATIN project that, in addition to visual assessment, it may be possible to perform an audio-tactile evaluation of G2 continuity. This would replicate natural interaction with the haptic sense but give additional information about the curve through sound.

As mentioned above, principal curvatures help to form the shape of objects. Shapes are assessed in relation to these curvatures and these curvatures may be adapted to form more pleasing or exact shapes. Usually this is normally not only achieved through manipulation of CAD models, but also in relation to physical prototypes. The images in Figure 1-4 show a designer assessing the curvature of an object using a deformable strip.

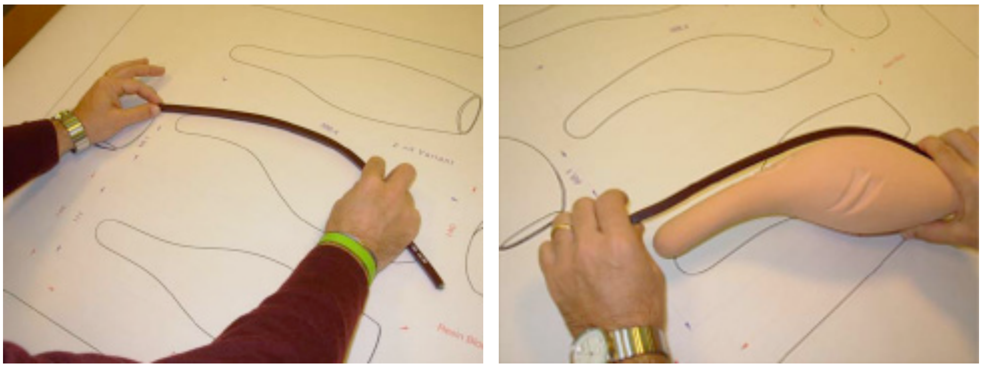


Figure 1-4: Designer assessing the curvature of an object

(Photographs courtesy of Alessi)

The strip is matched to a desired shape and then the physical model is evaluated against this. It is these types of actions that informed the ‘strip’ metaphor that was used in the SATIN interface (SATIN Consortium 2006) It is clear from the discussion here how fundamental an assessment of curvature is to the manufacture of everyday objects. It is because of this that the research undertaken within this thesis will focus upon curvature discrimination.

1.1.4 The SATIN Prototype

The concept of the SATIN system was to allow designers to interact intuitively with digitally created shapes (see Figure 1-5). This would allow them to ‘feel’ the shape and understand its aesthetic qualities. In addition, sound feedback would be provided in order that they could further investigate the shape's geometric qualities, such as curvature continuity, minimum and maximum curvature, and points of inflexion. A detailed description of the final SATIN prototype system can be found in Bordegoni (2010). It is therefore not proposed to give a detailed technical description of the prototype here. However a brief description of this may provide a useful background to the research reported within this thesis.

The SATIN prototype underwent a number of iterations during the course of the project, and an intermediate prototype can be seen in Figure 1-5. The key aspects of the device interface were; 3D visualisation (A), haptic strip (B), and sound feedback (C).

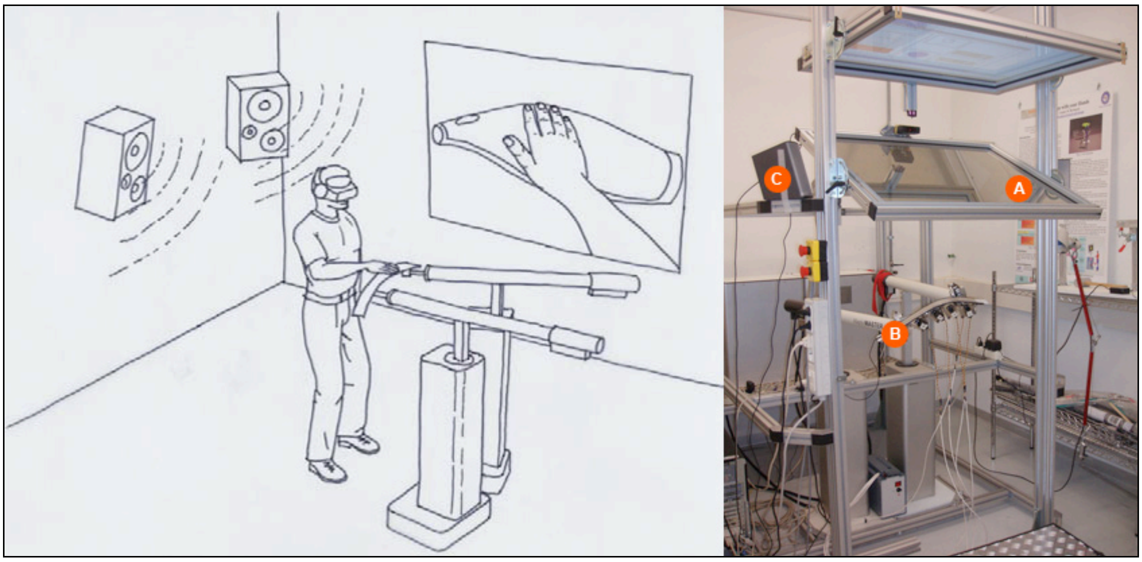
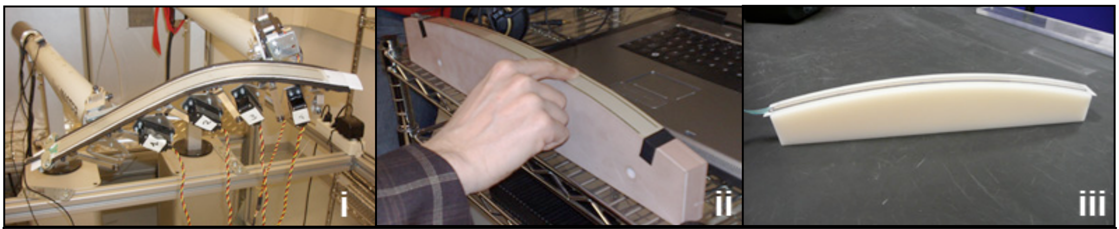


Figure 1-5: The SATIN concept (left) and SATIN prototype (right)
(Illustration courtesy of Politecnico di Milano)

The visual and haptic representation of the object was achieved through the use of the Direct Visuo-Haptic Display System (DVHDS). The visual aspect of this system was comprised of a digital light processing (DLP) projector, mirror, rear-projection screen, and finally a half-silvered mirror (labelled 'A' in Figure 1-5). It was used to co-locate a three-dimensional visual representation of the object over the haptic strip. This meant that as the user (wearing stereoscopic glasses) looked down towards the haptic strip through the half-silvered mirror, they would see the virtual object but not the strip. Therefore, as they touched the strip what they actually saw was their hand traversing the virtual object; this created the illusion of touching a real object.

The haptic element of this display was a flexible strip mounted on a pair of haptic masters (labelled 'B' in Figure 1-5). The haptic masters were used to position the object within the virtual space, whilst the flexible strip deformed to take the shape of the object (more specifically the curve shape along a particular plane). The haptic strip was a key part of the prototype interface which allowed the user to explore the curve shape. It also tracked their position along it in order to provide sound information with regard to the curvature at that point. Because of this, the haptic strip or representative blocks were central to evaluations undertaken for the SATIN project and studies within this thesis (see Figure 1-6).



An early version of the haptic strip is shown in photograph (i). Preliminary SATIN evaluations of sound feedback using this type of strip were undertaken on a representative block as shown in photograph (ii). The experimental work reported in this thesis also used blocks that were designed to be representative as shown in photograph (iii).

Figure 1-6: The SATIN haptic interface and representative blocks used in evaluations and experimental research.

The final aspect of the SATIN prototype was the sound interface. The haptic strip itself is limited in the extent to which it can represent geometric properties e.g. exact curvature, and so sound is used to convey this type of information. As the user moves their hand along the strip sound feedback informs the user of curvature values (a close-up of the haptic strip is shown in Figure 1-6i). This is experienced as the rising and falling of the sound's pitch, related to increases and decreases in the curvature of the curve shape, as the user's hand travels along the strip. The sound is also used to convey discontinuities in curvature. These were represented by an audible click as the user's hand passed the point of discontinuity. The user could then hold their finger at this point to hear an alternating sound between two frequencies, which represented the curvature at either side of the discontinuity. In this way the quality of the geometric shape could be assessed. The sound feedback was produced from speakers located either side of the prototype labelled 'iii' in Figure 1-5.

1.1.5 Preliminary Evaluation of the SATIN Prototypes

Prior to the commencement of the research detailed within this thesis, two preliminary evaluations were undertaken on the SATIN prototype. These evaluations helped to guide some of the focus of this research, and so a brief review of these will provide a helpful background. In addition to these preliminary evaluations, a further two evaluations were undertaken on later prototypes; these are reported in 'Heuristic Evaluation Report' (SATIN Consortium 2009) and SATIN technical report D9.2 (SATIN Consortium 2009) which are on the accompanying CD. Figure 1-7 gives a timeline of all the evaluations undertaken on the SATIN prototype system and the fundamental studies reported in this thesis. Whilst the later SATIN evaluations did not inform the focus of the studies reported here, they did contribute to an overall understanding of multimodal interaction issues and so helped in the formulation of the guidelines presented in Chapter 9.

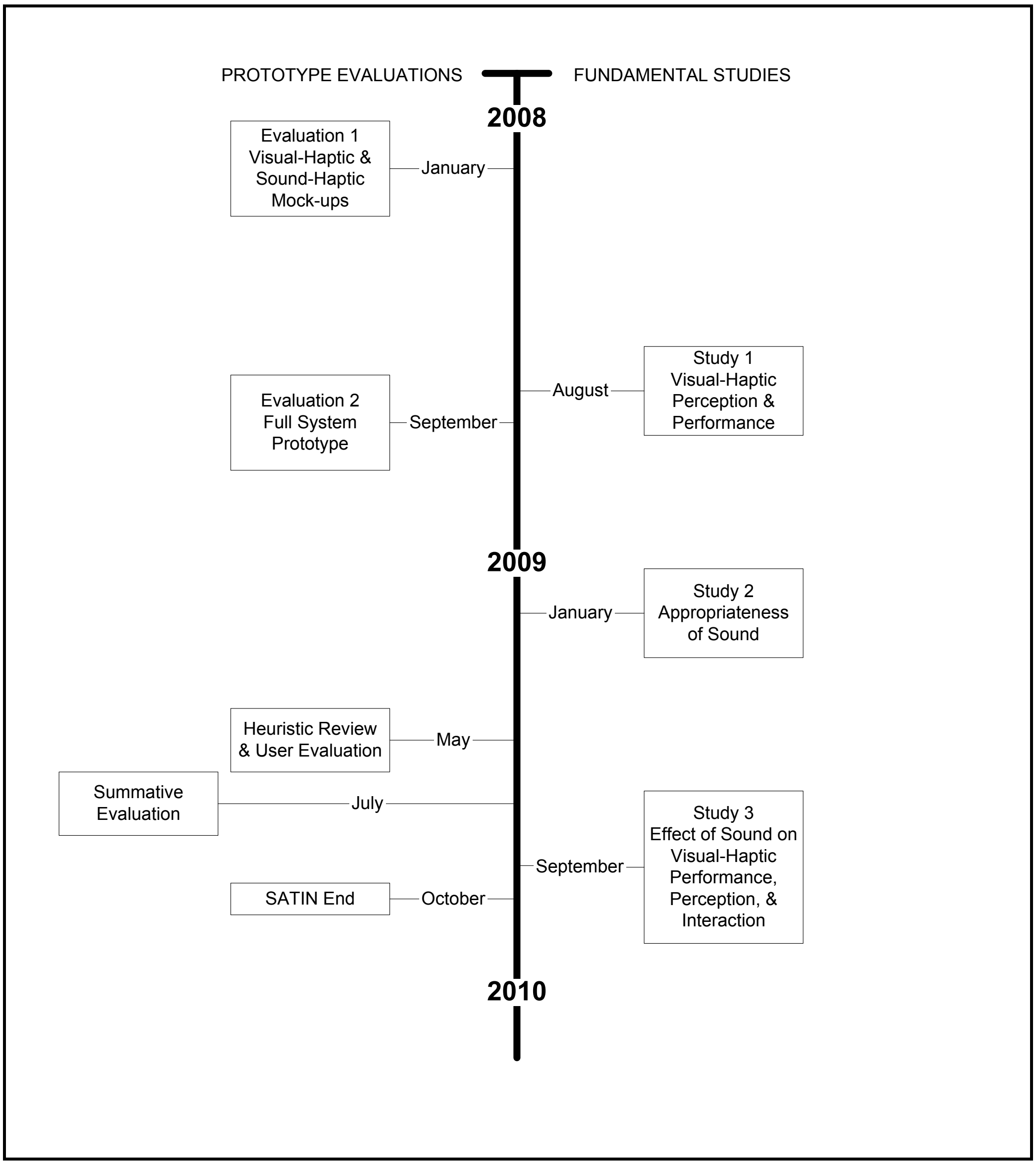


Figure 1-7: Timeline of Evaluations and Studies

1.1.5.1 Evaluation 1: Visual-Haptic and Sound-Haptic Mock-ups (January 2008)

The first evaluation was undertaken on two early stage mock-ups of the SATIN system. The mock-ups represented different facets of the complete system. The first was a visual-haptic prototype that enabled the participants to interact with the virtual object, but did not allow modification of the shape. It can be seen in Figure 1-8 that a wooden template stands in for the haptic strip that would be provided at a later date. The second prototype was a haptic-auditory mock-up that would allow users to touch the object and get sound feedback about various curve characteristics.

The aim of the evaluation was to gather formative feedback about the usability of the different interfaces and elicit feedback concerning the conceptual design from end-users. The evaluation consisted of a series of tasks to be undertaken on the interface, during which the participants were encouraged to give their views on the system. This was then followed by a questionnaire concerning usability, engagement, and user experience. The key findings for the visual-haptic mock-up are given in Table 1-1 and those for the sound-haptic mock-up are given in Table 1-2. A detailed description of the evaluation method and results can be found in the SATIN technical report D9.1 version 1 (SATIN Consortium 2008)(see CD for a copy of this).

Figure 1-8: Visual-haptic Prototype

Inset image courtesy of Politecnico di Milano

The mock-ups were generally liked by the end-users and the concept was thought to be valuable. In particular the use of sound to provide feedback on curve shape and curvature characteristics, such as discontinuity, was believed to work well. However, about 30% of participants reported that they had felt a mismatch between the visual and haptic representations of the object. In addition, about half of the participants also reported a mismatch between the sound and haptic representations. It was these observations in particular that led to a focus on modality integration and its effect on performance within Study 1 and Study 3 (see Chapter 3 and Chapter 7 respectively).

Table 1-1: Key Findings Visual-Haptic Mock-up

A. User Evaluation – Observations & User Comments
• Participants were good at positioning virtual object in achievable positions • Participants struggled with some positions that required objects to have 360 degree movement (not currently possible) • When position required system to be moved to extreme positions, some participants had problems with reaching the control device • Some participants appeared to perceive that the device was fragile • Participants varied in their perception of object scale • The majority of participants found it useful to feel the object they could see and felt that the haptic and visual interfaces matched • Some participants perceived a lack of alignment between the visual and haptic interfaces • Some participants commented that the [flat] sides of the haptic device did not match the [curved] shape of the object
B. Usability Questionnaire
• Participants felt the display of the objects was realistic, and that the movement of the object was natural • The visual display was comfortable and not distorted, but there were some negative comments about the display quality • Participants felt that they performed the task well • Participants experienced some problems with positioning the object, and thought that the system would be uncomfortable to use for a long period of time in its current configuration
C. Engagement Questionnaire
• Participants generally reported a high level of engagement with the system • Some participants felt that the input devices and other system aspects caused a distraction
D. Overall User Experience
• Participants liked the SATIN concept and had a generally positive feeling towards the system • Participants thought that a free rotation option would improve the system • Participants commented on the problems caused by the visual reflection • Participants were interested in the possibility of adding textured images to the display

Table 1-2: Key Findings Sound-Haptic Mock-up

B. Usability Questionnaire
• Participants generally found the sound useful and pleasant • Some participants perceived a mismatch between the sound and other displays • Sound type 1 (sine wave) was preferred • Participants thought it would be a good idea to use sound to represent the texture of the object surface
D. Overall User Experience
• Participants liked the concept of sound, particularly for perceiving discontinuity and curvature • The interface of the system was intuitive • Participants would like the sound to be used for representation of thickness, material properties and texture • There was no large preference for either headphones or speakers

Notes: Parts A and C were not applicable to the sound-haptic mock-up

1.1.5.2 Evaluation 2: Full System Prototype (September 2008)

The second evaluation was undertaken on a full system prototype. This was the first time that the visual, haptic, and sound interfaces had been integrated for evaluative purposes. The aim of the evaluation was therefore to test the usability of the integrated system and obtain end-user feedback about their experience. The evaluation consisted of a series of tasks that involved the use of haptic, visual, and sound feedback in order to identify curvature characteristics within the virtual object. This was then followed by a questionnaire concerning usability, engagement, and user experience. The results for task performance are summarised in Table 1-3. The key findings of the prototype evaluation are given in Table 1-4. A detailed description of the evaluation method and results can be found in the SATIN technical report D9.1 version 2 (SATIN Consortium 2008)(see CD for a copy of this).

Table 1-3: Task Performance

Identification Task	Response Time	Accuracy
Max. Curvature	33	84
Min. Curvature	17	45
Inflexion	47	72
Discontinuity	27	66
All	31	67

Table 1-4 Key Findings SATIN Prototype System

A. User Evaluation – Observations & User Comments

- Participants may be working at high-end of workload capacity. Further investigation is required.
- Memorability of point sounds (inflexion and discontinuity) needs to be improved
- Point sounds (inflexion, discontinuity) should be clearly distinct from continuous sounds (curvature).
- Efficiency of simultaneous presentation of sound information (e.g. curvature plus inflexion/discontinuity points) should be further evaluated

B. Usability Questionnaire

- Participants felt the system would be uncomfortable to use for long periods
- Participants responded generally positively towards the use of sound, and results indicate the SATIN approach is effective in helping designers appreciate the curve, discontinuities, and so on. Note that subjective responses are inconsistent with task performance.
- Participants felt the sound distracted them from the visual and haptic information
- Participants responded neutral when asked if the sound was pleasant, and indicate it may not be comfortable to listen to for long periods.

C. Engagement Questionnaire

- Responses to questions tended towards neutral, suggesting participants felt engaged with the system but not overwhelmingly so.
- Participants rated the ability to control events within the system and manoeuvre objects within the environment as the factors they felt most contributed to realism

D. Overall User Experience

- Participants commented that they disliked the system in terms of the postural issues
- Participants' responses when asked what contributes to a sense of realism were mixed, with three senses identified – sound, vision, haptics.
- Two participants rated the system low in comparison to their initial expectations, explaining that the sound was distracting and the system didn't meet their expectations of virtual reality. Another participant rated the system 4 out of 7 due to the postural concerns, whilst the highest rating (5 out of 7) commented positively on the use of sound.
- Participants tended towards positive responses when asked about the SATIN approach. Two identified the multi-modal aspects of the system as evidence for their rating.

The evaluations revealed a number of issues that had not been apparent in the interface mock-ups evaluated previously. This was the first time that participants had dealt with the full system, which meant they now had to attend to three modalities. It was observed that participants were at the high-end of their workload capacity, and it was considered that this may be due to the multimodal interaction. There was also some difficulty in perception of the sounds and additionally some users felt that the sound was distracting. This led directly to the focus of Study 2 on the appropriateness of sound, and the introduction of workload assessment for subsequent studies.

The results for system performance (see Table 1-3) also presented an issue, in that it was difficult to judge whether these levels of response time and accuracy were adequate. In some respects the overall accuracy of 67% could be considered poor, however without comparative data this was a difficult judgement to make. Similarly, the mean response time of 31 seconds poses interpretational problems; is this fast or slow? Again these concerns provided a focus for the direction of studies undertaken as part of this research.

1.2 Definition of the Problem

1.2.1 Perception

The focus of this thesis evolved from the context in which the SATIN research described above was undertaken. The discussion above identified two key concerns: the need for perceptible feedback within the 'modification-evaluation' loop, and the centrality of curvature in the evaluation of shape quality. This led to a focus of this thesis being human perceptual ability in discriminating differences in curvature. Specifically, it seeks to understand the consequences for perceptual acuity when haptic and sound modalities were combined with vision in order to discriminate differences in curvature. In addition, through understanding the mechanism by which a just noticeable difference is perceived, a means was provided for predicting the level of feedback required to distinguish changes in curvature.

1.2.2 Performance

Through undertaking some preliminary evaluations with the early SATIN prototypes it was realised that the measures of participants' performance (response time and accuracy) was difficult to contextualise. This was because without comparative data or an effective understanding of task difficulty the results proved difficult to interpret. For this reason another focus of this thesis was understanding the level of performance (response time, accuracy, and confidence) that can be achieved when discriminating a range of curvature differences. In this way the research results act as comparative data (or can be used to develop metrics) against which future evaluations could be compared.

1.2.3 Interaction

A further theme of this thesis has been in understanding the effect of combining modalities on interaction, and how sound may affect this. The development of this theme stemmed from observations made during the initial SATIN evaluations. These were not formalised, but were more an impression that interaction was erratic and that participants were adapting behaviour to suit the characteristics of the interface or feedback. It was therefore felt that some understanding was needed of how interaction was undertaken in relation to different modalities and how this might change dependent upon task difficulty or with the use of sound feedback.

1.2.4 Sound

During the course of the site visits it became apparent that one aspect of the SATIN concept that was difficult to envisage by end-users was the use of sound to provide feedback. This was probably because it was such a novel concept, and that end-users were mainly familiar with perceiving sound in a more naturalistic way. It was initially anticipated that in operation the use of sound to perceive object properties such as curvature would be fairly intuitive. It was discovered during preliminary evaluations with the SATIN prototypes that this was not the case. However, the reasons for this were not obvious and it was by no means certain that the difficulty was directly caused by the design of the sound used in exploration. Because of this a further focus of this thesis has been in assessing the appropriateness of sound to convey curve shape and curvature information.

1.3 Research Aims and Objectives

The overall aim of this thesis was to explore the value of haptic and sound feedback in the real-world perception of curve shape differences, and through the knowledge gained provide an evaluative framework for the assessment of virtual interfaces for the exploration of curve shape differences.

In order to meet this aim the following objectives were identified, to:

1. Identify the effect of combining haptic, visual, and sound feedback on perception, performance, and interaction.
2. Inform theory and develop a predictive model based on measurement of just noticeable differences
3. Assess the appropriateness of sound for conveying curve shape and curvature
4. Develop appropriate performance metrics for evaluation of curve shape using unimodal and multimodal feedback
5. Provide a framework for the evaluation of virtual interfaces for curve shape exploration

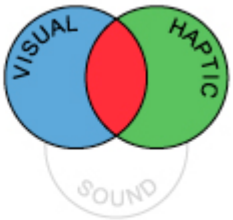
By isolating and combing modalities of real curved shapes in a systematic way, it was hoped to provide a greater understanding of the possibilities and limitations of unimodal and multimodal interaction. Through understanding real-world perception, a baseline for such interaction could be established. This should then provide valuable insights, and guidance, for the development and evaluation of virtual systems for curve shape exploration in the future.

1.4 Structure of the Thesis

The structure of this thesis is illustrated in Figure 1-9. Background material and the context of this study are presented within this introductory chapter. This also describes the structure and contribution of the thesis. Chapter 2 starts by discussing the nature of haptic perception in general, and then how this applies to objects in particular. In line with the context of this research there is then an exploration of what is known in relation to perception of curvature. Following this there is a broadening of focus to consider the perceptual aspects of multimodal interaction. This continues with a discussion of sound and sonification, which is important in the consideration of the effect of sound. Finally, this chapter concludes with a discussion of evaluation of, and guidelines for, multimodal interfaces.

The experimental work undertaken for this thesis is structured in three parts. The first part, Study 1 (Chapters 3 and 4), examines the effect of modality and curvature on just noticeable differences (JNDs). It further examines the effect of these factors plus confidence on performance. The second part, Study 2 (Chapters 5 and 6), explores the appropriateness of using sound to convey curve shape and curvature information. The third part, Study 3 (Chapters 7 and 8), evaluates the effect of sound when used with other modalities in order to explore the curve shape of an object. It also seeks to validate the model of prediction described in the first part of this research. These data collection and analysis chapters will now be described in more detail.

Study 1



Chapter 3 The research here aims to investigate the effect of unimodal and bimodal interaction on perception and performance. This was investigated through a psychophysical experiment where participants were asked to judge differences in curvature between two stimuli. Two types of stimuli were used in the study, one with a low curvature and one with high curvature. These were chosen to represent curvatures from different sized 'everyday' objects. These data were used to generate JNDs, and analysis was undertaken to assess the effects of curvature and modality on these. Performance data such as response time, accuracy, and confidence were also collected, and an observation was made of the amount of interaction undertaken in order to make a judgement. Analyses of these data were made in order to determine the effects of curvature, modality, and magnitude difference on these measures. The outcomes of this study were used to inform the development of performance metrics and assess if prediction of JNDs and performance was feasible.

Chapter 4 The psychophysical data generated through the experiment described in Chapter 3 were analysed to establish if stimulus gradient was a predictor of threshold gradient. Having discovered this to be the case it was analysed with additional data taken from the literature, in order to verify its generalisability. This further analysis confirmed the original findings and led to the development of a number of models for the prediction of unimodal and bimodal JNDs.

Study 2

Chapter 5 This chapter looks at the appropriateness of sound for conveying curve shape and curvature information. This was investigated through two experiments. The first explores the sound's suitability to convey curve orientation, and the second its suitability to convey curve magnitude. In addition, the experiments are structured so that they are undertaken twice; one before and one after a third experiment (described in Chapter 6). In this way some assessment of the effect of practice is also made. There was also investigation of user experience issues such as workload, preference, and helpfulness.

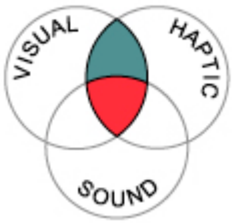
The author was aided in the running of these experiments with help from colleagues at the University of Nottingham. The sounds used in these experiments were designed by colleagues at the Technische Universiteit Eindhoven. All other work undertaken in connection with the two experiments outlined within this chapter was undertaken solely by the author.

Chapter 6 As well as the two fundamental experiments described in Chapter 5, a third experiment also examined the appropriateness of sound. This time the experiment was designed to reflect the more complex curves that would be found within an applied context. Participants had to match auditory information about curve shape or curvature to one of four curve shapes displayed visually. They were also able to interact with these curves in order explore them and so control the auditory feedback. Performance was judged in terms of response time, accuracy, and confidence. There was also

investigation of user experience issues such as workload, preference, and helpfulness.

This experiment was designed and run by the author assisted by colleagues at the University of Nottingham. The analysis of experimental data, discussion of results, and the conclusions drawn that appear within this chapter are the work of the author and were undertaken independently. The sounds used in these experiments were designed by colleagues at the Technische Universiteit Eindhoven.

Study 3



Chapter 7

The research here investigates multimodal interaction. In particular it seeks to understand the effect of sound on perception, performance, and interaction. This was investigated through a psychophysical experiment where participants were asked to judge differences in curve between two stimuli. Three types of stimuli were used in this study; they were designed to assess the effects of different stimuli properties (Curvature, Gradient, and Width) on performance. Additionally, the data generated could be used for the further analysis outlined in Chapter 8.

The data resulting from this experiment were used to generate JNDs and analysis was undertaken to assess the effects of stimuli and modality on them. Performance data such as response time, accuracy, and confidence were also collected. Analyses of these data were made in order to determine the effects of stimuli, sound, and magnitude difference on these measures. Additionally, data had been collected on the duration of exploration at various points on the stimuli. The effects of location and

	sound on exploration duration were analysed in order to understand whether participants interacted equally along the whole length of the curve and whether this was affected by sound. There was also investigation of user experience issues such as workload, sense use, and sense conflict.
Chapter 8	The psychophysical data generated through the experiment described in Chapter 7 were further analysed to evaluate if the predicted JNDs generated by the model (Chapter 4) were accurate and thereby validate the predictive model and the theory upon which it was based. Additionally, multiple regression analysis was undertaken upon visual-haptic performance data in order to assess stimulus gradient, magnitude difference, and width as predictors of response time, accuracy, and confidence. The models produced through this analysis could then be used to provide comparative data and performance metrics for evaluations. In addition, there was a serendipitous finding that JNDs conformed to Weber's Law, that is, the difference was a constant of the stimulus (12%).

Following on from the data and analysis chapters, Chapter 9 provides a synthesis of the experience gained whilst conducting this research in order to suggest a framework for the evaluation of multimodal interfaces for shape exploration. Chapter 10 provides a general discussion of the findings of the research conducted across all three studies and relates this to the main themes of perception and performance. Finally, Chapter 11 discusses the conclusions of this work and how these relate to the research aim and objectives. There are also some suggestions of how this research might be taken forward and ideas for future work.

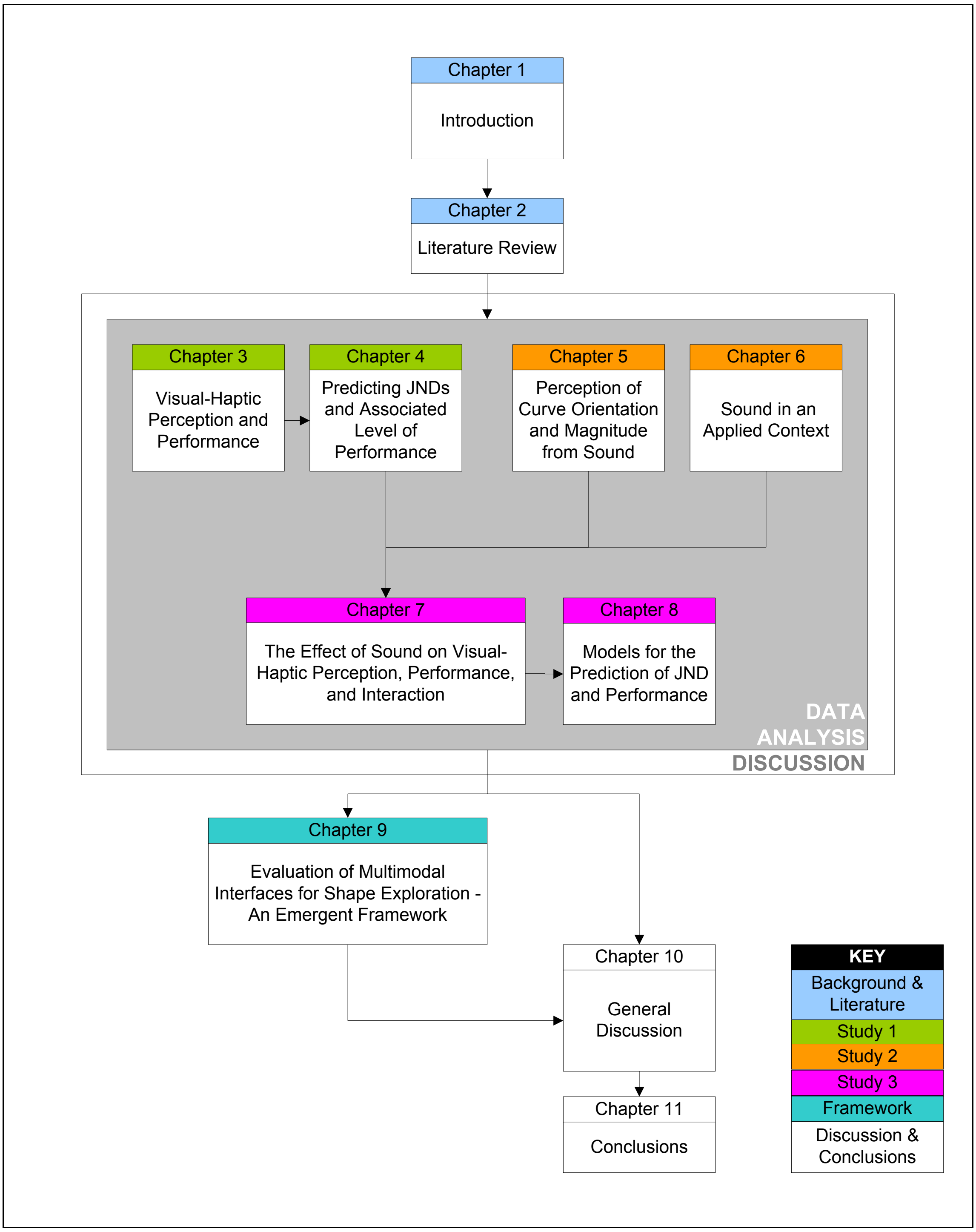


Figure 1-9: Structure of Thesis

Throughout this thesis the key themes were perception, performance, and interaction. Through examining these it was hoped to provide understanding about multimodal interaction for curve shape assessment, and provide a means of predicting JNDs, response times, accuracy, and confidence. This knowledge could then be used for the specification and evaluation of interfaces.

Chapter 2: Literature Review

2.1 Scope of Review

This chapter undertakes a review of available literature that is relevant to the aim of exploring haptic and sound feedback. It starts by examining the nature of haptic perception in general, and moves on to the perception of objects in particular. It then focuses on the perception of curvature and what is known about the acuity of this. Consideration is given to the combination of visual, haptic, and auditory modalities. This aims to understand some of the issues involved in multimodal interaction. Attention is then turned to the perception, classification, and use of sound. This provides necessary background to how sound may be used to convey information, and the human factors issues arising from this. Finally, the chapter concludes with a brief discussion of the evaluation of multimodal interfaces and the available guidance for their development. This sets the context for the development of guidance, which is one of the aims of this thesis.

2.2 The use of Multimodal Technologies in Product Design

2.2.1 Design and Technology

Having reviewed in detail the issues relating to multimodal perception, it would be of interest to look at how designers currently use technology to perceive and judge the shape of objects. A review of design approaches was undertaken as part of the SATIN project (SATIN Consortium 2007). It was found that whilst the detail of approach varied from company to company, there were similarities in designers' approach to the task of creating and modelling objects. In the initial stages this could involve the production of free hand sketches using paper and pencil, or the modelling of objects in clay with various tools. This led to the digitization of these artefacts and the production of Computer Aided Design (CAD) models. In some instances the initial 'hand' crafted phase was skipped and models produced directly with CAD applications. It was felt by those taking part in this project that virtual models were the most efficient way to design a product. However, it was also felt that there were tangible benefits to being able to touch designs. These types of process, and the desire of designers to touch objects were also reported in the Touch and Design project (Bordegoni and Cugini 2005). Whilst both these studies focused on product designers, another study which involved a range of skilled artists concluded that whilst vision had an important role in monitoring progress, it was the haptic sense that provided the significant perceptual information for object creation (Prytherch and Jerrard 2003). It would therefore seem that whilst visual interfaces (such as CAD systems and Graphics applications) are currently important, the addition of a haptic interface would allow for a more naturalist and complete interaction.

There are currently a number of haptic technologies that could prove useful to product designers. These range from commercially available products to more experimental devices (see SATIN Consortium, 2008) for a detailed discussion of these). At the commercial end are technologies such as the FreeForm System from Sensable (2007). This provides a visual virtual representation of a clay-like object which can be shaped via a haptic pen device (Phantom Desktop or Phantom Omni). This enables designers to ‘feel’ the shape that they are producing. A number of studies have evaluated this system with product designers (Cheshire, Evans et al. 2001; Sener 2002; Sener, Pedgley et al. 2003). The system was generally well received because of its haptic feedback capabilities. It was found to be useful in the rapid generation of 3D objects, and provided more flexibility in producing organic shapes and textured surfaces than traditional CAD systems. However, it was also noted that the lack of constrained modelling lead to imprecision and inappropriate model data for production purposes.

There are a number of more experimental haptic devices. These are typically force-feedback displays which give a sense of a grasped object (Harwin and Melder 2002; Maciel, Sarni et al. 2004); pin arrays which use actuators to deform a flat surface to provide object shape (Wagner, Lederman et al. 2002; Pasquero and Hayward 2003); and local surface displays which use haptic devices to mimic the sensation of touching an object across a particular plane (Hayward 2004; Provancher, Cutkosky et al. 2005). Unfortunately, many of these are little more than proof of concept prototypes and as such have not been evaluated with designers. However, an experimental interface developed during the Touch and Design project was evaluated by a number of product designers (Bordegoni, Espinach et al. 2005). The device comprised a visual display of the 3D object (at eye level) with a haptic ‘strip’ used to model the shape. The ‘strip’ was a solid bar supported between two haptic masters. This allowed users to manipulate the shape by scraping away material from the surface of the object (rather like actions observed in clay model making). The evaluation revealed a high level of user satisfaction with the device, and in particular the ability to interact freely with the virtual shape. This form of interaction was reported as being more naturalistic in terms of movement and evoked the feel of clay-modelling.

2.2.2 The Perception of Digital Models in the Virtual Environment

Reviews have previously been conducted into human factors of virtual reality (VR) that have primarily focused on the design of a visual interface and the way that users will interact with a virtual environment (VE) or virtual object (Stanney, Mourant et al. 1998; Nichols and Patel 2002). The issues associated with manipulation and navigation of virtual objects relates to the design of the input device and any associated menus or modality selection techniques (see Patel et al (2006) for a detailed examination of menus and manipulation designs).

Many elements of the visual display will influence user performance and use, including: resolution (Tao, Doug et al. 2006), colours (Billger, Heldal et al. 2004), lag (Liu, Tharp et al. 1993), frame rate (Chen and Thropp 2007), and detail (Dinh, Walker et al. 1999). The influence of these elements will depend on the task being performed. For example, if a design task includes fabric, it is important that the detailed material properties of the fabric are clearly displayed, to show the hang and folds in the fabric. If a task involves design using reflective materials, such as metal or glass, the surface qualities should be represented in some manner. It is important to note that expert designers may not need an accurate display of reflective properties in order to produce effective designs. For example, using zebra stripes on an image may be effective at conveying curved surfaces (SATIN Consortium 2007).

Haptics research comes largely from the robotics community involving teleoperation in VEs. In these types of systems, force-feedback is utilised for object manipulation. This is usually provided at the wrist via kinaesthesis rather than via cutaneous sensation at the fingers (tactile feedback). Richard et al (1996) examined the effect of direct force-feedback, pseudo-force feedback and redundant force feedback on regulation of grasping force. The task required users to put their hand inside a box to reach and pick up a virtual object (ball) then move it through various locations. The results showed that when graphics, haptics, and redundant audio were present, fewer errors were made (-69%) and shortest average completion times occurred. However, with redundant visual information the task was significantly impeded. Hale and Stanney (2004) particularly considered the consequences of combining different models (principally haptic and vision) in interaction. They suggest that as vision frequently dominates the integrated visual-haptic percept caution should be used when vision and haptics are combined for tasks involving size, shape, or position judgements. They also consider the cognitive load of combining visual and haptic information. They emphasise the need to avoid time lags in multiple modalities, but suggest that if the visual system is overloaded (as may be the case with complex 3D CAD models) then object identification information can be provided haptically without adding to cognitive load. There is also the suggestion by Lederman and Klatzky (2004) that multiple haptic contact points will aid object recognition.

2.3 The Nature of Haptic Perception

The term ‘haptic perception’ refers to our sense of the world built up through cutaneous and kinaesthetic sensations (Gibson 1966; Loomis and Lederman 1986). These are derived from receptors within the skin (cutaneous) or beneath the skin within muscles, tendons, and joints (kinaesthetic). The first of these provides tactile sensations such as roughness, coldness, and hardness, whilst the latter provides information on the position of our joints and muscular effort (proprioception). There are two basic modes of haptic perception; perception for action, and action for perception (Wolfe, Kluender et al. 2006). The first of these is perception used for control of objects during action. This relies on feedback from cutaneous mechanoreceptors to ensure appropriate grasp and to avoid slippage. The second, action for perception, involves active interaction with the world in order to seek out haptic information. The active nature of haptic perception differentiates it from other senses as we usually have little haptic sense without engagement whereas we can hear things, see things, and smell things without seeking them out.

Through actively exploring the environment it is possible to perceive an object’s shape, texture, hardness, temperature, size, and weight. However, unlike vision where the object can be perceived almost instantly, touch necessitates exploration over a period of time (Loomis, Klatzky et al. 1991). A drawback of the tactile sense is that there is significant blurring of the sensations as deformations of the skin affect more than the immediate area of contact, thereby inhibiting resolution of detail (Loomis 1981). There are also attentional limitations on what can be perceived. Whilst it is possible to attend to a pattern presented across two fingers on the same hand, there is less of a deficit in the information perceived if presentation is to fingers on different hands (Craig 1985).

There are three facets of tactile perception that are relevant to exploration of objects; pressure sensitivity, spatial acuity, and temporal acuity.

Pressure sensitivity is a measure of the amount of pressure exerted on the skin in order for it to be perceived. Sensitivity varies across the body, with the most sensitive areas being the lips and fingertips (Sekuler, Nash et al. 1973). Pressure sensitivity of the fingers is within the range of 2.1-2.5mg, although factors such as gender, age, and temperature can decrease sensitivity (Weinstein 1968).

Spatial acuity is a measure of how well the skin is able to resolve separate stimuli i.e. the minimum distance between two points such that they are distinguishable. Again sensitivity is dependent upon body location with the hands being most sensitive. Resolution for finger tips is between 1mm (Loomis 1981) and 2mm (Weinstein 1968), with acuity declining with age (Stevens and Patterson 1995). However, depending upon the type of discrimination required, spatial acuity could be even finer. Loomis (1981) investigated the ability to discriminate the relative positions of two stimuli and found that lateral displacements as small as 0.17mm were detectable.

Temporal (vibration) acuity is a measure of the ability to resolve a stimulus over a period of time. The minimum lapse between two stimuli that is resolvable (i.e. the highest frequency) is approximately 1.4 milliseconds or 700Hz (700 cycles per second). However, the highest sensitivity is achieved in the range of 250-300 Hz (Wolfe, Kluender et al. 2006). There is debate about the lower end of this range with some reporting 200 Hz as the optimum frequency (Van Doren, Pelli et al. 1987). Sensitivity of the fingers varies due to hardening with use, so the ring finger is more sensitive than the frequently used index finger (Lederman 1976). Sensitivity to vibration is an important factor in perception of surface texture.

2.4 Haptic Perception of Objects

Haptic perception can either be passive or active, and is a combination of cutaneous and kinaesthetic cues (Loomis and Lederman 1986). The classification of perception as either passive or active is dependent upon the will of the observer in seeking to control the interaction (active) or not (passive). Our interaction with objects tends to be active as we explore them to ascertain their properties. Our ability to distinguish different shapes and objects haptically is very good. When given a range of everyday objects, for example comb, boot, carrot, book, and plate, people are able to make accurate (94%) and fast (less than 5 seconds) identifications (Klatzky, Lederman et al. 1985). The way that we interact with these types of objects is largely dependent upon what we are seeking, and so different types of haptic ‘exploratory procedures’ are adopted (Lederman and Klatzky 1987). As can be seen in Figure 2-1, various hand movements, postures, and combinations of fingers are chosen to perceive such things as size, shape, texture, temperature, and weight. These characteristic movements are classified into eight different exploratory procedures; lateral motion, pressure, static contact, unsupported holding, enclosure, contour following, function test, and part motion test (these latter two are not illustrated, but involve the movement or exploration of the whole or part of the object to discern function).

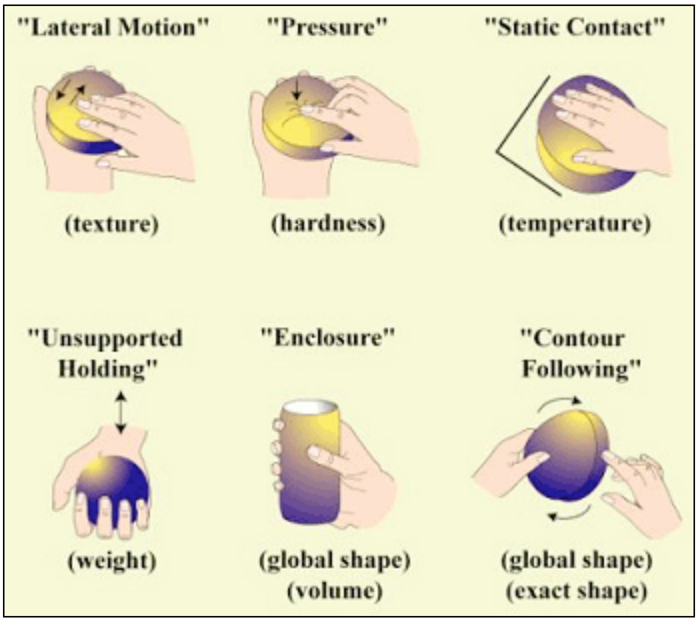


Figure 2-1: Exploratory Procedures (Lederman and Klatzky 1987)

Each procedure seems to be optimal for discovery of certain object characteristics. For instance, lateral motion is the best way to detect surface texture, whilst static contact provides the best temperature feedback. For shape recognition we see that the procedure adopted is one of contour following. The types of object properties that can be explored haptically can be divided into geometric properties (shape and size) and material properties (texture, stiffness, temperature).

2.4.1 Geometric Properties

In some ways, haptic shape recognition is similar to visual recognition where the bounds of an object are identified and then the more detailed features explored. However, where visual exploration is usually a split second process, haptic exploration can range from a few seconds to many minutes. This is due to the limited field of 'view' offered by the hands. An object cannot be taken in instantly but must be explored edge by edge and surface by surface until an overall 'picture' is gained. This is equivalent to a person establishing their bearings in a strange city. This is generally easy with a map since it is possible to form an instant picture of the surrounding space, but is more difficult without as the person has to construct their understanding street by street, forming and holding the relationships in their head. Therefore ascertaining global properties such as shape is less efficient haptically than visually.

This local to global appreciation of shape is evidenced in the work of Lakatos and Marks (1999). They conducted two experiments where participants were required to judge the similarity of shapes whose features varied at a local and global level. In the first experiment comparisons were made between pairs of shapes either visual or haptically. It was found that objects with comparable global shape but different local aspects were considered to be similar when judged visually but different when judged haptically. A second experiment, with the same design but restricting exploration time to 1, 4, 8, and 16 seconds, found that this limitation effected judgements of similarity for local but not global shape. It was concluded that the haptic system first weights local features more heavily than global shape, but that over time this differential decreases.

For vision, it is thought that the perception of objects is achieved through the recognition and synthesis of their component shapes (Bierderman 1987). These components or geons are primitive conical structures that are distinguished through the contrasting of five easily detectable edge properties; curvature, collinearity, symmetry, parallelism and cotermination. The number of geons required to encode the multiplicity of everyday shapes is as little as 36. A number of researchers have hypothesized that a similar mechanism may apply to the haptic perception of shapes (Soechting, Song et al. 2006; Ehrich, Flanders et al. 2008). Their research suggests that haptic attention is focused upon areas of high spatial contrast, and that the synthesis of sensed shapes is biased toward simple geometric objects (such as circles or elliptical arcs).

Given the perceptual dependence upon simplified shapes, it is of interest to know the range of this sensitivity. Louw (2002) explored the extent of haptic sensitivity in relation to the variation in amplitude and width of a curved object. It was found that humans are able to distinguish around 300 different shape stimuli when amplitude and width were extrapolated over a range of 1µm to 1m. It was also found that within this range people were much better at differentiating sharp (large amplitude/small width) from smooth (small amplitude/large width) shapes than distinguishing small (small amplitude and width) from large (large amplitude and width). He therefore concluded that perception of shape was at its poorest when the proportion between amplitude and width was similar for each shape.

2.4.2 Material Properties

Unlike geometric properties which require temporal and spatial exploration, material properties can be instantly accessed. It is argued that people are pre-attuned to these properties so they ‘pop-out’ (Lederman and Klatzky 1997). Because of this it is relatively easy to assess the type of material properties that an object has, whether it is rough or smooth, hard or soft, cold or warm. Material properties therefore provide an instant cue as to the nature of the object explored and are central to haptic recognition (Klatzky, Lederman et al. 1985).

Perception of material properties can also be informed through auditory stimuli. Different objects have a different sound or timbre due to their composition. They also exhibit differences in the way acoustic energy builds (attacks) and then dissipates (decays). Both of these auditory aspects help inform the nature of an object and can be successfully utilised to demonstrate object properties (Klatzky, Pai et al. 2000).

2.4.3 Temperature

'Touch temperature' as opposed to ambient temperature (Gibson 1966) is dependent upon the thermal conductivity of the material touched. Every material has particular conductive qualities so that metal feels cold as it has high-thermal conductivities whereas plastic feels warm because it has low-thermal conductivities. What is actually perceived when touching an object is not its 'temperature' but the gradient between skin and object - hence highly conductive materials dissipate heat rapidly and so they 'feel' cold. Skin temperature is normally around 33^{o}C and when touching an object this shifts either upwards or downwards depending on the temperature at the surface of that object. What is experienced as the thermal 'shift' in the skin’s temperature also affects our perception of surface temperature. A well-known illusion is that of touching a surface at room temperature with two fingers, one that has been immersed in hot water and one that has been in cold water, the differentials in temperature between the surface and the different fingers gives the experience of the surface as having two different 'temperatures' (Egeth, Kamlet et al. 1970).

2.4.4 Texture

As the hand moves across the surface of an object the skin is displaced. It is this displacement and its frequency that provides information about the texture of a particular object. Blake and Sekuler (2006) give the example of glass and sandpaper. Glass gives minimal displacement of the skin at a fairly regular frequency, whereas sandpaper gives erratic displacements at a constant frequency. Each material has its own unique frequency and displacement properties, and so enables discrimination between various surfaces with a high degree of accuracy, and even within sub-types of different materials e.g. coarse or fine sandpaper.

The minimum peak-to-peak threshold of detection is about 1mm, although the direction of these will affect how well fine detail is detected. The best detection occurs when ridges run perpendicular to the long axis of the finger (Essock, Krebs et al. 1992). In terms of comparative difference between different spatial frequencies, a difference of about 3% is the minimum detectable (Lederman and Taylor 1972). Determining the roughness of a surface is not governed by the speed at which the textured surface is felt. Regardless of whether the finger is slowly moved (minimum modulation) or quickly moved (high modulation) the texture is judged the same (Blake and Sekuler 2006).

Haptic simulation of texture has been explored through tactile and force-feedback devices, however neither of these solutions has been entirely successful to date. A naturalistic sensation of texture results from a combination of cutaneous and kinaesthetic feedback and the aforementioned devices provide only one or the other of these. It has been suggested that a multimodal approach, utilising haptic and auditory feedback, may overcome the limitations of current devices and algorithms (McGee, Gray et al. 2001).

2.5 Perception of Curvature

2.5.1 What is Curvature?

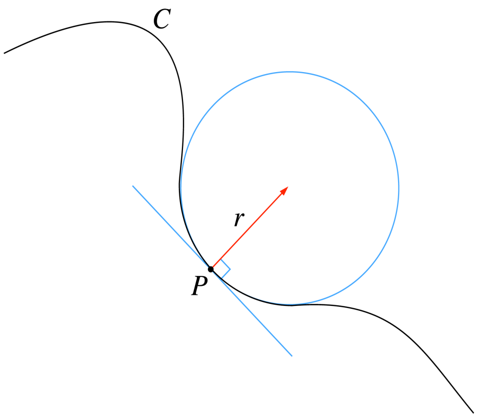


Figure 2-2: Curvature as Radius of Oscillating Circle

Before exploring the research literature, it is useful to define what is meant by curvature and how it is measured. Intuitively we know when a surface is curved and whether this is slight or extreme. However these subjective measures are difficult to quantify. In order to explore our sense of curvature we need a more objective mathematical formulation of what is understood by 'curved'. The degree of curvature at a given point is expressed as the reciprocal radius of the circle tangent at that point. So curvature (k) is calculated as: k= 1/r. For a flat line, curvature is equal to zero and is constant. Curvature is also constant for a circle, with each point being equal to the inverse of its radius. For all other curves, curvature varies and is the inverse of the radius of an 'oscillating circle' at a given point 'P' as illustrated in Figure 2-2.

Another important property of curvature is that as scale increases curvature decreases. This is most easily understood in relation to a circle. A small circle of around 0.05m radius has a curvature of 20/m. If this is scaled to three times the size, so giving a circle with radius 0.15m, then the curvature decreases to 6.66/m; that is it becomes flatter.

2.5.2 Relationship of Shape and Curvature

One important aspect of shape is whether an object's surface is flat or curved; is it a cube or is it a ball? Figure 2-3 shows a series of two-dimensional objects; a square, an oval, and a circle (left to right). In part the difference between them is the extent to which we perceive the lines that form the object as either flat or curved. So the degree to which something is either flat or curved informs our perception of the global shape. In the case of the oval and the circle, the degree to which they are curved is different. The line that forms the oval varies in its level of 'curvedness' whilst the line that forms the circle has the same amount of curve throughout. It can be seen that the degree to which the line curves helps to define the shape. Therefore, in order to perceive shape we need to be able to perceive the difference between a flat and curved surface, and between different degrees of curved surface.

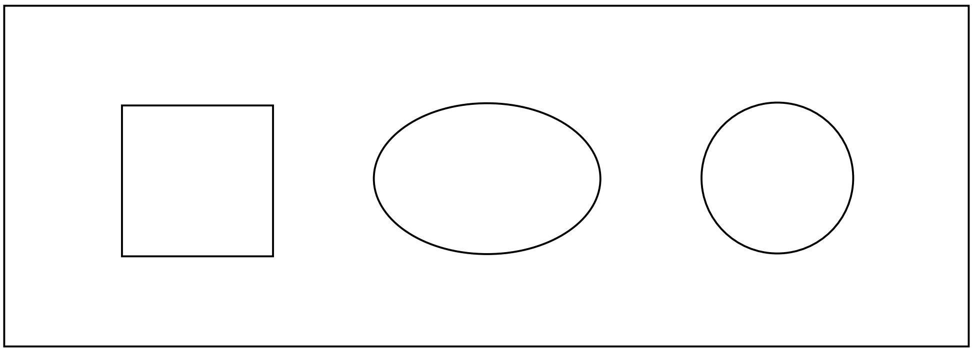


Figure 2-3: Shapes are constituted from flat and curved lines

This differentiation is relatively easy in the example given; however surfaces often display far more subtle differences between what is flat and what is curved, and between what is curved and what is more curved. So discerning these differences in curve shape is dependent upon the acuteness of human perceptual ability. However, to explore this it is necessary to have a more precise understanding of curve shape. This issue is discussed in the research literature, and has been resolved through defining the 'curvedness' of a shape with reference to its curvature (Koenderink and van Doorn 1992; Kappers, Koenderink et al. 1994). This gives a precise and quantified means of understanding the amount of change, or just noticeable difference (JND), required in order to perceive the difference between a flat and curved surface (absolute threshold) and between two curved surfaces (difference threshold). This has led to an area of research which focuses upon haptic detection and discrimination of curvature differences. In understanding the acuity of curvature perception we can precisely understand the limits of shape perception.

In relation to the artefacts of product design, it is also important to be able to detect the rate of change of curvature across the surface of an object; more specifically whether this curvature change is continuous or discontinuous. So of interest in relation to this is the identification of absolute and difference thresholds for curvature perception.

2.5.3 Investigating Perception

We are surrounded by a world of physical events, and how we internalise these stimuli through sense perceptions has been investigated through what is known as Psychophysics. The principles of this type of enquiry are relatively simple as it seeks to establish a quantitative measure of the relationship between physical events in the world and our psychological response. There are two key measures which underpin this investigation; Absolute Threshold and Difference Threshold.

Absolute threshold is the level that a physical event must attain before we can perceive it (hence why this is also known as ‘detection’ threshold). Physical events below these thresholds cannot be detected by human sense organs and so we do not perceive anything as happening. An example of this would be a dog whistle, which when blown emits a sound with a frequency higher than that detectable to humans and so we hear nothing, however a dog in the vicinity of this event would perceive a sound.

Once a stimulus has been detected the next measure considers what increase (or decrease) in physical stimulus will be required in order that the stimulus is perceived to have changed. This is termed the Difference Threshold or Just Noticeable Difference (JND). However consideration of this is not as simple as identifying the change in intensity required. This is because the amount of change required will be dependent upon the intensity of the original stimulus, and so multiple difference thresholds can be identified. For example, where the intensity of a stimulus is low then the increase in that intensity to perceive a difference will be small, however where the intensity is high then the increase required to perceive a JND will be greater. This relationship between the physical stimulus and the increase (or decrease) in intensity required for a difference to be perceived is embodied in Weber’s Law. This law states that the degree by which a stimulus needs to change in order to be perceived is a fixed proportion of the initial stimulus, such that $\Delta I/I=k$ (where ΔI is the intensity change or JND, and k is a constant)(Blake and Sekuler 2006). This means that whilst difference thresholds vary dependent upon the level of the stimulus, the intensity increase tends to be a constant proportion. For example, a weight of 100kg needs to increase by 2kg in order for a change in weight to be perceived. However, a weight of 150kg needs an increase of 3kg. Whilst these are different rates of increase, the proportion of the weight to perceive a difference is 0.02 or 2%. This proportion is known as a Weber Fraction and applies to a wide variety of sensory events (see Table 2-1).

Table 2-1: Weber Fractions

WF	Percentage	Sensory Event
0.013	1.3	Electric Shock
0.020	2	Heaviness
0.048	4.8	Loudness
0.079	7.9	Brightness
0.083	8.3	Taste (salt)

Source: (Teghtsoonian 1971)

In order to calculate Absolute and Difference thresholds three psychophysics methods are used; method of limits, method of adjustment, and method of constant stimuli

(Gescheider 1985). Whilst it is possible to use any of these, the method adopted generally for investigation of haptic thresholds, and that chosen for the research reported here, has been that of constant stimuli. This is considered to be the most accurate of the methods, although it is also the most time consuming (Rose 2006). It is not proposed to outline this method in detail here as this is described in depth as part of the procedure for Study 1 and Study 3 (see Chapters 3 and 7). However, in brief, these methods largely involve the presentation of stimuli to the experimental subject, who is then required to state whether or not they perceive a given sensation, in the case of detection, or if they sense a change in the case of difference. These responses are then analysed to identify the respective thresholds.

Psychophysical methods have been used widely within human factors research. They have been adopted where there has been a need to assess human perception. The focus of research using these methods has been quite broad ranging from practical issues such as work and safety, to more esoteric considerations of aesthetic preference. For example, Ciriello (2010) uses the method of adjustment to determine the maximum acceptable forces for lifting, lowering, pushing, pulling and carrying in female industrial workers. Hsia and Drury (1986) used the method of limits to assess the best handle design in a lifting task, whilst Han et al. (1998) used magnitude estimation to determine passenger aesthetic preference for a variety of train interiors. Psychophysical methods have also been used to gauge perceptual limitations in order to improve the design of interactive devices. Tan et al. (1994) used the method of constant stimuli in order to determine the limitations of haptic force feedback, and so design better controllers for hands and arms. Adelstein (2003) used the method of limits to quantify human sensitivity to latency in virtual environments in order to design countermeasures to this.

The usefulness of psychophysics methods within human factors research is evident by the wide range of subjects to which it has been applied. In particular its use for the assessment of human limitations in order to ensure the optimum design of devices fits well with the objectives of this thesis. Therefore, what follows is a focus upon the psychophysical literature in relation to curvature perception. In examining this focal literature it was hoped to establish what is known in relation to curvature perception and identify areas were this knowledge might be usefully extended in order to meet the objectives of this thesis.

2.5.4 Absolute and Difference Thresholds

There have been a number of studies that have explored haptic curvature perception (Davidson 1972; Gordon and Morison 1982; Goodwin, John et al. 1991; Goodwin and Wheat 1992; Kappers, Koenderink et al. 1994; Kappers and Koenderink 1996; Kappers, Koenderink et al. 1996; Pont, Kappers et al. 1997; Louw, Kappers et al. 2000; Henriques and Soechting 2003; Wijntjes, Sato et al. 2008). The research has considered a range of curvatures and conditions in order to understand perception of curved shape. Fundamental to this understanding has been an investigation of the mechanism of curvature perception and how perception of object shape may be affected by a number of factors (these are discussed in the next two sections). In the course of this research thresholds for a range of curvatures from tightly curved (286/m) to almost flat (0.80/m) have been identified. The types of curvature explored represent an object diameter range of 0.35cm to 20cm, which covers a good proportion of 'hand-sized' objects i.e. those things that we are likely to pick up and feel.

Unfortunately there has not been a standard way in which thresholds have been reported, and whilst authors have made comparisons between findings these have not been comprehensively reported in one place. This has been remedied here by converting thresholds originally given as base-to-peak heights to the equivalent curvature (see discussion 2.4.5, and Figure 2-4 for dimensions used within the literature). However some research, because of its experimental concerns, does not report in detail the curvatures involved and so thresholds could not be reported (Davidson 1972; Kappers and Koenderink 1996). The absolute and difference thresholds reported within the literature, with sufficient detail, are given in Table 2-2 and Table 2-3.

Table 2-2: Absolute Thresholds

Author	Stimulus	Explore	Reported Threshold	Stimulus Width (cm)	Stimulus Curvature (/m)	Threshold Curvature (/m)	JND (/m)
Gordon & Morrison (1982)	Real	Active	Base-to-Peak	2.00	0	1.80	1.80
Gordon & Morrison (1982)	Real	Active	Base-to-Peak	3.00	0	1.16	1.16
Gordon & Morrison (1982)	Real	Active	Base-to-Peak	4.00	0	0.90	0.90
Henriques and Soechting (2003)	Virtual	Active	Curvature	12.00	0	1.25	1.25
Wijntjes et al. (2008)	Virtual	Active	Curvature	16.00	0	0.44	0.44
Wijntjes et al. (2008)	Real	Active	Curvature	18.00	0	0.45	0.45
Pont et al. (1997)	Real	Static	Curvature	20.00	0	0.49	0.49

For absolute threshold there was found to be little difference between perception of real and virtual stimuli (Wijntjes, Sato et al. 2008). However this may depend on the type of virtual stimuli used. The threshold produced by Henriques and Soechting (2003) is about double that of Wijntjes et al. (2008) even though it is within a similar range. If the result of Henriques and Soechting (2003) is disregarded, what should be noticed is that JND decreases with increasing width. This means that as the stimulus gets wider, the amount by which a flat surface needs to increase in curvature in order to be perceived as curved decreases. Gordon and Morrison (1982) expressed this in a slightly different way. Their research identified that when the threshold (base-to-peak height) was divided by the width, a constant ratio of 0.009 was found. This equates to a constant elevation of about half a degree (see Figure 2-4 for an illustration of these dimensions).

This trend within the data lead to further investigation by Louw (2000) across a much wider range of spatial scales from 150µm to 240mm. The focus of this research was to investigate the relationship of stimulus width to threshold (in this case base-to-peak height). Louw (2000) found that width and threshold varied in a systematic way across the whole spatial range. When plotted on double logarithmic scales threshold was a linear function of width such that: $A_{threshold} = C_0 width^{1.3}$. This agrees with the earlier work of Gordon and Morrison (1982), but more importantly establishes that this relationship holds true across a wide spatial range. It is therefore possible to conclude that curvature detection is reliant on the width of the stimulus.

Table 2-3: Difference Thresholds

Author	Stimulus	Explore	Threshold	Stimulus Width (cm)	Stimulus Curvature (/m)	Threshold Curvature (/m)	JND (/m)	WF
Goodwin, et al. (1991)	Real	Static	Curvature	0.35	144.00	158.00	14.00	0.10
Goodwin, et al. (1991)	Real	Static	Curvature	0.35	287.00	319.00	32.00	0.11
Goodwin & Wheat (1992)	Real	Static	Curvature	0.50	286.00	322.85	36.85	0.13
Goodwin & Wheat (1992)	Real	Static	Curvature	0.50	153.85	181.54	27.69	0.18
Gordon & Morrison (1982)	Real	Active	Base-to-Peak	2.00	2.40	4.40	2.00	0.83
Gordon & Morrison (1982)	Real	Active	Base-to-Peak	2.00	2.80	5.20	2.40	0.86
Henriques and Soechting (2003)	Virtual	Active	Curvature	12.00	2.50	4.00	1.50	0.60
Kappers & Koenderink (1996)	Real	Active	Curvature	20.00	-	-	-	0.41
Pont et al. (1997)	Real	Static	Curvature	20.00	0.80	1.43	0.63	0.79

For difference threshold the constant of interest is the Weber Fraction as this indicates the presence of a consistent relationship between the stimuli and thresholds. Table 2-3 shows that for curvature perception there are no consistent Weber Fractions as these ranged from 0.10 to 0.86. Given this, it is unlikely that curvature perception conforms to

Weber's Law. A similar conclusion was drawn by Kappers and Koenderink (1996) when reflecting on their findings.

The width relationship seen for absolute threshold does not seem to apply here. In examining Table 2-3 it will be noticed that thresholds and Weber Fractions vary for similar widths. For example, the 20cm stimuli used by Kappers and Koenderink (1996) and Pont et al. (1997) give rise to Weber Fractions of 0.41 and 0.79. This level of discrepancy indicates that width may not play the same role in defining threshold.

These findings suggest that whilst it would be possible to predict the level of absolute threshold, it is currently unknown what the difference threshold is likely to be for any given curvature (other than it is some factor again higher than that of detection).

2.5.5 Mechanism of Curvature Perception

In order to understand the possible mechanisms of curvature perception it would be helpful to consider the various stimulus dimensions that have been reported. Figure 2-4 illustrates the dimensions explored within the literature; curvature, base-to-peak height, local attitude, gradient, and width (Width A usually describes the extent of a stimulus, Width B is used in the calculation of gradient and in Louw (2000)).

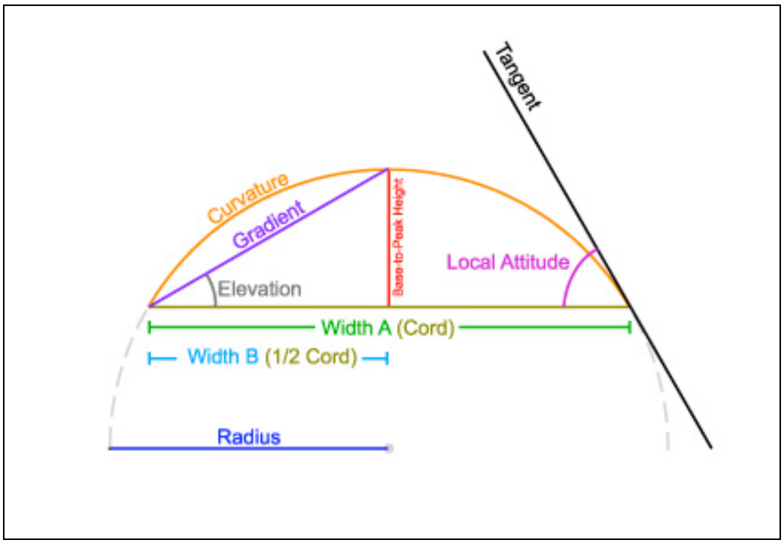


Figure 2-4: Dimensions used within the Haptic Literature

Davidson (1972) explored the relationship between active haptic interaction and successful curvature perception. For this study he used both sighted and blind participants. The focus of the study was to see if the exploration adopted by blind participants was different to that adopted by sighted participants, and whether this was more successful. In the first experiment it was found that a total of five exploratory techniques were adopted; pinch, grip, span, top, front, and trace (see Figure 2-5 for illustration). The pinch, grip, and top techniques were found to be used the most. However, the grip technique was used twice as frequently by the blind participants than the sighted. Davidson (1972) suggests that the increased acuity of blind participants' thresholds was due to this technique. This was tested in a second experiment where sighted participants were restricted to using grip, pinch, and span. It was found that significantly fewer errors were made with the grip technique. Davidson (1972) concluded that this technique was good for focusing on the distinguishing features of the curve since it focused upon the ends-to-middle relationship. In this he suggests that perception is 'kinaesthetically triangulated'. In other words he seems to believe that perception is due to detecting height differences across the stimulus.

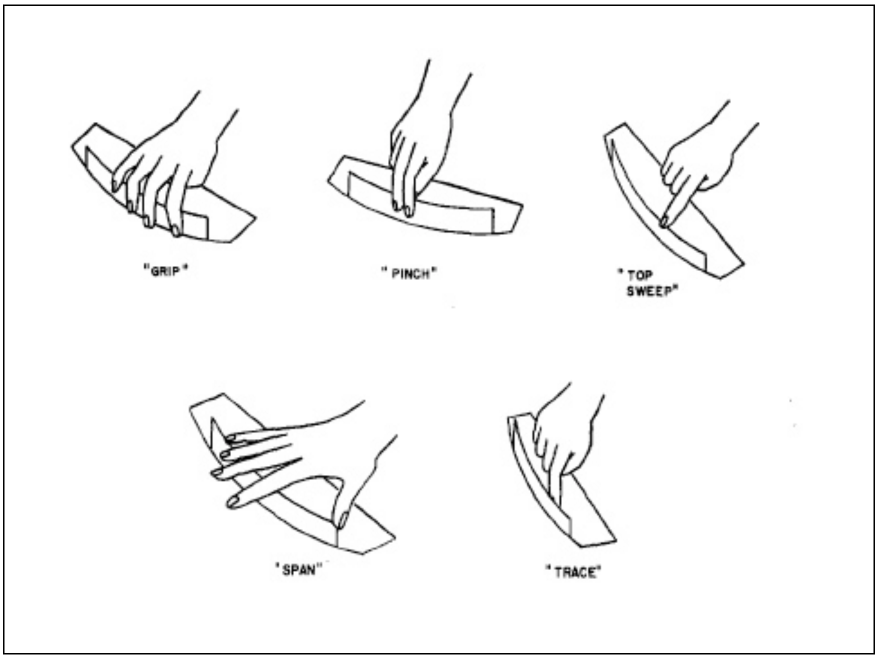


Figure 2-5: Davidson's Exploration Techniques (1972)

Gordon and Morrison (1982) sought to identify the effective stimulus for curvature detection. They proposed that there were three likely effective stimuli for curvature perception; base-to-peak height, the radius of curvature, and the gradient. They measured the absolute thresholds (base-to-peak height) for three different lengths of curvature; 2cm, 3cm and 4cm. The thresholds for each were compared and it was concluded that base-to-peak height could not be the effective stimulus as this varied for each condition. The radius of curvature was dismissed as this also varied for each condition. The final measure gradient was calculated by dividing the base-to-peak height threshold by half of the stimulus width. This resulted in a consistent figure of 0.009 across all three conditions. Gordon and Morrison (1982) therefore concluded that the effective stimulus of curvature detection was gradient. The figure of 0.009 equates to an elevation angle of 0.50. This agrees with the conclusions of Davidson (1972) in terms of an end-to-middle relationship, although they show that this is not dependent upon height alone but the ratio between width and height.

Effective stimulus of curvature detection is also explored by Pont et al. (1997). The study undertook a number of experiments which examined the cutaneous and kinaesthetic sensitivity of the hand. The purpose of this was to ascertain if detection threshold was correlated to structural properties of the hand and/or geometric properties of the stimulus. It was found that thresholds were similar regardless of presentation to the fingers or the palm, and so it was concluded that there was a possible correlation between threshold and contact length. In order to determine if this was the case then a further analysis was undertaken through comparison of three limit cases which would result from variation of length. The first of these was constant curvature. This predicted that as length increased thresholds based on curvature would remain constant, whilst those based on height would increase. The second limit case was constant attitude difference. This time it was predicted that both curvature and height would increase with length. Thirdly there was constant height difference. With this limit case it was predicted that height would remain constant but that curvature would increase with length. Dependent upon which of these limit cases held true it would be possible to identify the effective stimulus for static perception as either curvature difference, attitude difference, or height difference (see Figure 2-6). The analysis revealed that, as length over which the stimulus was touched increased, there was a concomitant increase in both height and curvature thresholds. This led Pont et al (1997) to propose that the effective stimulus for static curvature detection was attitude difference. They further explain that when the total attitude difference over the touched part of the stimulus is in excess of 2^{o} it can be distinguished from flat. A further study confirmed these findings and that this mechanism was also the effective stimulus for active curvature detection (Pont, Kappers et al. 1999).

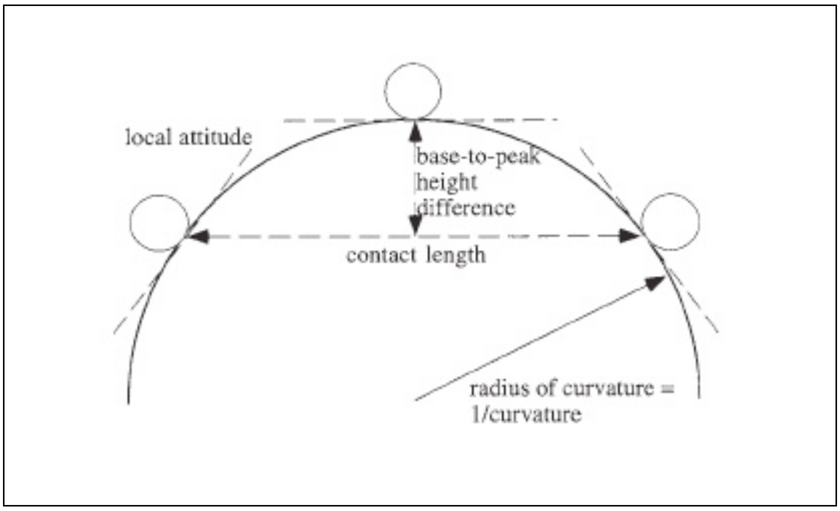


Figure 2-6: Proposed Effective Mechanisms of Curvature Perception (Pont, Kappers et al. 1997)

This is different to the mechanism identified by Gordon and Morrison (1982), which is gradient. However, whilst these are different they are trigonometrically related (as are curvature, width, and base-to-peak height[1]). This is illustrated in Figure 2-7 where it can be seen that the elevation angle (ΔDBE) is half of that formed between the tangent and the chord (local attitude). It is also worth noting that the angle formed between the tangent and the chord (ΔPAB) is half the central angle (θ) which means the central angle equals the total attitude.

[1]In 2008 when these issues were being considered, the geometric relationships were not commented upon within the literature. This meant calculating these independently and using available geometric formulae in order to derive and convert thresholds from the information available within the literature (see Appendix A). However, there has since been an extremely useful paper by Wijntjes et al (2009) which provides geometric formulae which state each of the threshold measures in relation to the others.

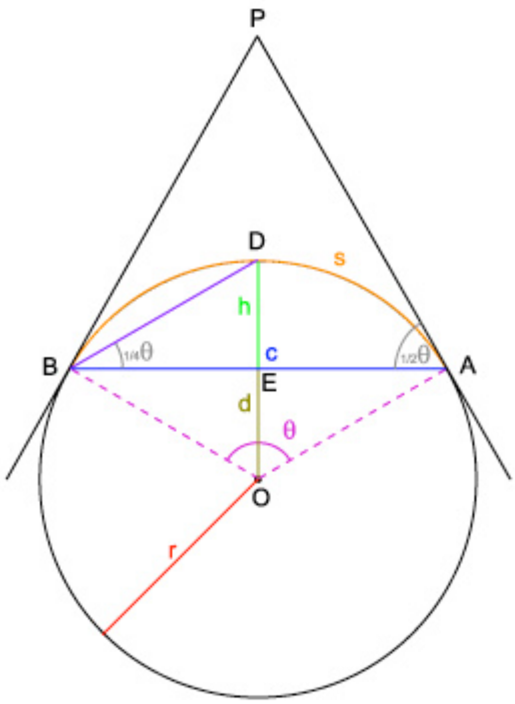


Figure 2-7: Geometric Relationship of Stimulus Dimensions

These types of geometric relationship were used to calculate the equivalent thresholds given in Table 2-4 and Table 2-5 (see Appendix A for formulae). It can be seen from this that the absolute threshold indicated by Gordon and Morrison (1982) as a gradient of 0.009 is effectively the same as the 2^{o} attitude difference indicated by Pont et al (1997). There is little evidence at this time to suggest which of these is the more likely; however for practical purposes they may both be considered the effective stimulus.

Table 2-4: Absolute Threshold Equivalent Measures

Author	Stimulus	Threshold Curvature (/m)	Base-to-Peak Height (cm)	Local Attitude (deg.)	Total Attitude (deg.)	Gradient (deg.)	Elevation (deg.)
Gordon & Morrison (1982)	Real	1.80	0.009	1.03	2.06	0.01	0.52
Gordon & Morrison (1982)	Real	1.16	0.013	0.99	1.99	0.01	0.50
Gordon & Morrison (1982)	Real	0.90	0.018	1.03	2.06	0.01	0.52
Pont et al. (1997)	Real	0.49	0.245	2.81	5.62	0.02	1.40
Henriques and Soechting (2003)	Virtual	1.25	0.225	4.30	8.60	0.04	2.15
Wijntjes et al. (2008)	Virtual	0.44	0.141	2.02	4.03	0.02	1.01
Wijntjes et al. (2008)	Real	0.45	0.182	2.32	4.64	0.02	1.16

Table 2-5: Difference Threshold Equivalent Measures

Author	Stimulus	Threshold Curvature (/m)	Base-to-Peak Height (cm)	Local Attitude (deg.)	Total Attitude (deg.)	Gradient (deg.)	Elevation (deg.)
Gordon & Morrison (1982)	Real	4.40	0.02	2.52	5.04	0.02	1.26
Gordon & Morrison (1982)	Real	5.20	0.03	2.98	5.96	0.03	1.49
Goodwin, et al. (1991)	Real	158.00	0.02	16.05	32.10	0.14	8.02
Goodwin, et al. (1991)	Real	319.00	0.05	33.93	67.86	0.31	16.99
Goodwin & Wheat (1992)	Real	322.85	0.13	53.81	107.62	0.51	26.90
Goodwin & Wheat (1992)	Real	181.54	0.06	26.99	53.97	0.24	13.49
Pont et al. (1997)	Real	1.43	0.72	8.22	16.44	0.07	4.11
Henriques and Soechting (2003)	Virtual	4.00	0.73	13.88	27.77	0.12	6.93

2.5.6 Factors affecting Curvature Perception

There are a number of factors which have been shown to influence the perception of curvature, as follows:

- Static curvature discrimination thresholds have been found to be 1.6 times smaller along the finger than across the finger (Vogels, Kappers et al. 1999). It has also been found that sensitivity to curvature is greater along the hand (palm to fingers) than across the hand (Pont, Kappers et al. 1997).
- With active exploration the technique used to scan a surface can affect the acuity of perception. Blind participants used a griping technique (see Figure 2-5) twice as much as sighted users. The use of this technique was considered to have improved the level of curvature perception (Davidson 1972).
- When exploring curvature dynamically, an edge is judged to be straight when it curves away from the observer (Davidson 1972).
- Static curvature judgements are affected by a previously touched surface curvature. If having first touched a convex surface, a flat surface is touched, this will be perceived as convex. This was also true when touching a concave surface (Vogels, Kappers et al. 1996).
- Surface friction has an effect on curvature judgement. The curvature of a high-friction surface is often over-estimated, where the curvature of a low-friction surfaces is often underestimated (Christou and Wing 2001).
- Discriminating curvature difference is better when surfaces are touched successively with one hand than if touched simultaneously with both hands (Kappers, Koenderink et al. 1994).

- For fingers it was found that touching each stimulus successively with one finger produced lower difference thresholds than touching each stimulus successively with two fingers. However, successively touching each stimulus with two fingers from one hand produced lower thresholds than simultaneously touching each stimuli with one finger from each hand (Horst and Kappers 2007).
- In the range of 'hand-sized' objects active touch produced slightly lower thresholds than static touch. However, for very small objects (0.35cm width) the situation was reversed (Kappers and Koenderink 1996). However, it has also been found that there is little difference between thresholds for static and dynamic touch (Pont, Kappers et al. 1999).

These factors are not definitive but are the key issues that were considered when designing the experimental studies reported in this thesis. They are also issues which need to be considered when developing haptic interfaces, and the consequences for perceptual acuity for different types of interaction. They have particular relevance to interfaces such as the SATIN prototype, where users are expected to interact freely with their hands or fingers. This may have consequences for the fidelity of their perceptual experience.

2.6 Multimodality

This section briefly describes some effects of combining various modality conditions. This is not intended to be a definitive exploration of the issues related to multimodality, but a consideration of the ways in which modalities typically combine and some of the issues resulting from this.

2.6.1 Visual-Haptic

The haptic sense is not used in isolation for exploration of shape but is usually used in conjunction with vision. The combination of cues from these inputs constructs the percept experienced. However, whilst a unified percept is experienced the information received from the visual and haptic senses may not provide equivalent representations of the three-dimensional shape (Norman, Clayton et al. 2004). At a more localised level curvature magnitude is experienced differently in the visual and haptic senses (Kappers, Koenderink et al. 1996). Haptically perceived curvature is overestimated in relation to the visually presented curvature. The difference between the two has a factor range of 0.3 to

3. So when a curvature is felt and then seen, it is often experienced visually as being more flat than expected.

Given that visual and haptic senses produce congruent yet different perceptions then this suggests that one sense dominates the other. In most cases there is visual dominance over haptic perception. Yet in other cases, where visual information is blurred, haptic perceptions take precedence over visual (Heller 1983). It is thought that the visual and haptic senses are combined in the most optimum fashion, so where visual cues are poor touch dominates, however where they are equal vision tends to be dominant (Ernst and Banks 2002).

2.6.2 Visual-Auditory

In a naturalistic environment the role of sound tends to be one of supplementing our visual sense. It can impart important information as to the location of an event or provide confirmatory information about the nature of what is happening. This is illustrated in an experiment undertaken by Sekuler, Sekuler and Lau (1997). Participants were asked to view a short animation in which two discs moved across the screen. The visual information was ambiguous and so could be interpreted in one of two ways; firstly that the discs moved towards each other, passed each other and continued on their trajectory across the screen; or secondly that the discs moved towards each other, collided, and bounced off each other reversing their direction of movement. Either interpretation was plausible. In order to induce the latter interpretation, the animation was accompanied by a brief click (lasting 2.5ms at 75dB) presented before, after, or at the point of coincidence of the two discs. The results showed that the presence of sound increased the perception that the discs collided and bounced off each other, although the effect was strongest when the sound occurred before or at the point of coincidence.

2.6.3 Haptic-Auditory

There is evidence to suggest that auditory information plays a key role in the perception of surface properties. In particular, sound affects the perception of surface roughness. This was demonstrated by Jousmaki and Hari (1998) where participants rubbed their hands together whilst receiving sound feedback through headphones. The sounds were manipulated such that participants felt their hands to be dry and parchment-like at high frequency, and wetter and rougher at lower frequencies. This has been replicated by using other abrasive surfaces (Guest, Catmur et al. 2002). This work illustrates that in addition to the tactile perception of surface roughness through the hands, the act of rubbing them also produces distinctive sounds which are associated with a given level of roughness.

The ability of sound to create illusory effects has been utilised in the simulation of texture. Because sound informs our knowledge about the material properties of an object it can be taken as a signal of our interaction with it. The addition of sound to haptic feedback has been found to enhance the sense of 'roughness' of a given surface and so give the illusion of texture (McGee, Gray et al. 2001; Guest, Catmur et al. 2002). The use of bimodal (haptic and auditory) cues was found to give rise to higher confidence in judgments of surface texture than for unimodal cues (Lederman, Morgan et al. 2002). However, Lederman et al. (2002) also argued that the integration of auditory cues was highly dependent upon amplitude, and at low amplitudes could easily be masked by environmental sounds (and so ignored).

2.6.4 Issues of Multimodality

Whilst modalities can integrate in complementary ways as reported above, there is also the possibility that dominance of one modality over another can lead to erroneous perceptions. A particularly relevant example is that demonstrated by Rock and Victor (1964) where participants were asked to handle an object viewed through a distorting lens. The object was a square, however when viewed through the lens it appeared rectangular and participants reported it as such. Even though the haptic information was correct it was dominated by the visual perception of the object as being rectangular.

Whilst vision is usually the dominant modality, this is not always the case. Shams, Kamitani, and Shimjo (2000) demonstrated a visual illusion provoked by sound. Observers viewed a single flashed light stimulus accompanied by a double click sound. As a result, they reported seeing two successive flashes instead of the actual one.

The success of perception can be affected by attentional demands. This can be found unimodally, for example when expecting a haptic stimulus in one location it takes time to notice stimulation in another location (Spence, Pavani et al. 2000). It can also be found cross-modally where there is competition between modalities for attentive resource. Spence, et al. (2001) showed that where attention was focused for tactile stimulation it took much longer to respond to either auditory or visual stimuli instead.

In light of the cases given above it is therefore not certain that modalities will combine in a beneficial way, or what the outcome of combinations will be. In addition, the need to attend to more than one modality can in itself lead to increased response times.

2.7 Sound and Sonification

2.7.1 Properties and Perception of Sound

The perception of sound is brought about by fluctuations in air pressure caused by the physical properties or interaction of objects. These variations in pressure and frequency produce sensations within the human auditory system (outer and inner ear) that are perceived as loudness and pitch. Optimally, humans are able to hear frequencies in the range of 20Hz to 20 kHz. There is a diminishment of this faculty with age, with a reduction in the upper limit to about 8 kHz by the age of 80. The best absolute audibility thresholds for human hearing are between 2-6 kHz; in this range a sound with a level of 0dB can just be heard. Either side of this range, that is to say in the high and low frequency ranges, the dB level must rise in order for frequencies to be perceived. A consequence of this is that frequencies at the same decibel level are not necessarily perceived as being the same loudness. For example, a frequency of 100Hz at 60dB has the same loudness as 1000Hz at 50 dB (Wolfe, Kluender et al. 2006).

Humans can perceive small changes in intensity of a sound, experienced as increases in loudness, of 1-2dB. Within the environment these judgements need to be made with complex sounds that exhibit a range of frequencies and changes of intensity. In order to detect change in these circumstances the auditory system uses what Green (1982) calls profile analysis. This is the utilisation of neural information about the relative activity of neurons across different frequencies. In practical terms this means it is possible to detect even a small change of intensity in one frequency against a background of different frequencies with constant intensities (Green, Kidd et al. 1983).

As well as being sensitive to changes in intensity (loudness), humans are also sensitive to changes in frequency (or pitch as it is experienced). A sound with a frequency of 500Hz will sound lower in pitch than a sound of 1000Hz. To put this into context, the scale of a piano runs from low A at 27.5 Hz to High C at 4180 Hz, which is a range of 7 octaves; pitch then allows sounds to be ordered from low to high. The smallest difference in frequency used within western music is a semi-tone which is a change of about 6%. Human ability to discriminate frequencies is generally much better than this with a minimum perceptible difference of 0.2-0.3% (Moore 2004). However, it is at its best for mid-range frequencies where discriminations of 0.1% can be made e.g. the difference between 999Hz and 1000Hz (Wolfe, Kluender et al. 2006).

2.7.2 Naturalistic and Abstract Sounds

Sound can be characterised in a number of ways. Here they are divided into two types; naturalistic and abstract sounds. This seems to be the most intuitive and straightforward division, drawing a line between those sounds that emanate from within nature through the physical properties of objects and those sounds that do not and are therefore some form of abstraction.

Naturalistic sounds are those created by the vibration of objects induced through collisions, movement, or flow of external forces (e.g. wind) and relate directly to the material properties and physical characteristics of the object. These are the types of sound that inform us about facets of the environment around us, and as such have been used within, for example virtual environments (VE), in an effort to increase the sense of reality. Whilst these are termed naturalistic sounds for the purpose of this discussion, they too may have an abstract quality to them. This is in the sense that for the most part sounds used within a VE are generated sounds and as such do not necessarily have exactly the same physical characteristics as those naturally occurring sounds for which they stand in. Even so the perceptual qualities of the sound may be identical, equivalent, or plausible and so without auditory comparators can be taken as real (SATIN Consortium 2008). As we use sounds to inform us of our environment and actions within it, then the correct use and design of auditory events can make actions within a VE more intuitive and so may improve performance (Díaz, Hernantes et al. 2006).

Abstract sounds on the other hand have no relationship to natural properties of an object or occurrence. Whilst classifications abound for these types of sound two broad instances will be discussed here; sonification and earcons (for a detailed taxonomy of these types of sound see Hermann (2008)). Sonification is the mapping of a sound to a property or parameter of an object in order to convey information about it (Kramer, Bargar et al. 1999). For example, the fall and rise in temperature of an object may be sonified through an increase and decrease in sound frequency (where low temperature is mapped to low frequency, and high temperature is mapped to high frequency). So as the temperature of an object increased an observer would perceive an increase in pitch, as the temperature fell they would perceive a concomitant fall in pitch. Where an observer can interact with the object and so influence the data sonification the term 'interactive sonification' is used (Hermann and Hunt 2005). An example of this might be where the observer points or touches a specific place on a graph and the data at that point is conveyed through sound.

A further type of abstract sound is the earcon which acts an auditory icon (Blattner, Sumikawa et al. 1989; Brewster, Wright et al. 1993). Unlike a sonified sound there is usually no intrinsic link between the sound produced and an event. They tend to be used for alarms or attention drawing signals; for example the sounds produced by door bells, telephones or fire alarms would fit into this category. They are typified by well-defined repetitive tones in specified rhythms or tempos. Whilst they have advantages in being distinctive and simple to synthesize, they can become annoying and distract from the task if not designed with care (Stanton and Edworthy 1999).

2.7.3 Sonification, Applications, and Human Factors Issues

If we ask the question 'What does a curve sound like?' we quickly realise that there is no naturalistic relationship between either the haptic or visual sense of curvature and sound. So to utilise sound to convey curve shape and curvature characteristics requires that its symbolic possibilities are understood. The field of sonification explores how data can be represented and understood through the medium of sound.

Sonification is the translation of object parameters into sound display dimensions such as loudness, pitch, and tempo (Carlile 2002). However to be considered sonification, Hermann (2008) suggests that this translation should conform to four characteristics; Firstly, that the sound reflects objective properties or relations in the data; secondly, the transformation is systematic; thirdly, the sonification should be reproducible, that is interaction with the same data should produce the same sound; and lastly, the system can intentionally be used with different data.

The success of sonification is largely dependent upon consideration of three factors; mapping, polarity, and scaling (Walker and Nees 2010). Mapping considers the way in which acoustic dimensions such as frequency are assigned to data parameters such as rising temperature. It has been found that not all acoustic characteristics are equally well suited for auditory display; for example pitch and loudness were found to be applicable to a wide variety of representations, whereas attack (speed of sound onset) and tempo were limited in what they could convey (Walker and Kramer 1993). Another consideration is polarity, which is how the increase or decrease in a dimension is handled. This should not be assumed to be intuitive, or even consistent, and relies largely on the mental models constructed by listeners in relation to the data sonified (Walker 2002). Finally, there is consideration of appropriate scaling which is related to the degree by which an increase in, for example, the frequency corresponds to an increase (assuming a positive polarity) in the data. The level of this varies dependent upon the listener, although there has been some success in producing scaling factors for a number of mappings (Walker and Nees 2005). Sonification of data therefore requires careful consideration and is likely to require investigation with the intended end-users.

Sonification has been successfully used in a number of real-world applications in order to convey various types of information. Perhaps one of the simplest are car parking sensors which use sound tempo to convey distance; the nearer the car gets to another object the quicker the sound becomes. This is a fairly straight forward use, but it clearly illustrates that it is possible to understand data dimensions through sound. More complex uses include: precise positioning of medical devices with ‘tactile audio’ (Jovanov, Starcevic et al. 1998), real-time auditory feedback of limb movements to aid neuropathy patients (Ghez, Rikakis et al. 2000), auditory display of gas and oil well data (Barras and Zehner 2000), sonified interactive spreadsheets (Stockman 2005), and ‘vOICe’ which through auditory substitution allows unsighted users to locate and identify objects (Auvray 2007).

It is clear that the use of sound presents a number of opportunities to convey information. The use of sound may even enable the more efficient use of cognitive resources as indicated by multiple resource theory (Wickens, Lee et al. 2004). This postulates that the workload associated with each modality is limited, therefore by distributing information across modalities workload is reduced and cognitive capacity increased. However, there are limitations to the effectiveness of this. For example, if the sounds are distracting or do not have temporal or spatial coherence with other modalities then they may impact on the ability to undertake a task (Wickens, Dixon et al. 2005). There are a number of other human factors considerations that have been highlighted by Walker and Nees (2010). These include the perceptual and cognitive abilities of the listener along with their musical ability and level of training.

2.8 Evaluation and Guidelines for Multimodal Interfaces

The focus of the review will now shift to a brief consideration of the evaluation of multimodal interfaces and the level of existing guidance. This aims to set the context within which evaluation and guidelines might be situated, and so understand the imperatives for the development of such guidance. The nature of multimodal interfaces, such as the SATIN prototype, are similar to virtual environments but also share similarities with augmented reality, for this reason the literature is drawn from both these domains.

2.8.1 Definition, Methods, and Objective Measurement of Usability

Usability can be defined as 'the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use' (ISO-9241-11 1998). Nielsen (1993) suggests that usability is characterised by five factors; learnability, efficiency, memorability, errors, and satisfaction. Although Preece et al. (2002) suggests the factors to consider are; effectiveness, efficiency, safety, utility, learnability, and memorability. Whilst the exact goals may differ, most researchers would assert that by setting goals and quantified measures based on these factors, the extent of usability can be gauged. In addition, a number of subjective qualities that gauge the user's experience have been identified as; satisfying, enjoyable, fun, entertaining, helpful, motivating, aesthetically pleasing, supportive of creativity, rewarding, and emotionally fulfilling (Preece, Rodgers et al. 2002). Again these can vary according to the research and context of use. However, it would not be expected that all systems should address all of these elements. What is important is that evaluation methods (e.g. questionnaires, interview questions) do address those that are relevant to a particular system.

In addition, a number of specific test batteries have been developed with the aim of quantifiably measuring user performance in virtual environments. For example VRUSE, VRSART (Kalawsky, Bee et al. 1999), VEPAB (Lampton, Knerr et al. 1994), MAUVE (Stanney, Mollaghasemi et al. 2003), VECEET (Whelan 1996), NAÏVE (Griffiths, Sharples et al. 2006). These have the collective aim of identifying measurable elements of performance, such as time taken or accuracy to complete a set task, and using this performance to distinguish between different participant abilities or the ways in which different system or environment designs influence user performance.

Traditional objective measures of performance are primarily effectiveness and efficiency (ANSI/NCITS 354-2001 2001; ISO/IEC 25062 2006). Other measures of performance relate to subjective elements of the user experience, including presence, ease of use, ease of learning, and user comfort (Bowman and Hodges 1999). Livingston (2005) proposes using objective measures based on real-object performance versus virtual object performance. It was therefore thought likely that the data gathered from studies reported within this thesis could be used for the purposes of this type of objective assessment and to form the basis of a test battery or evaluation framework (see the framework described in Chapter 9).

2.8.2 Guidelines for interface development

Wilson and Eastgate (2002) suggested that issues affecting VE may not be unique and that it may be desirable to utilise knowledge already gained from other domains. There are a number of accepted guidelines that can be utilised, the most common of which are Nielsen’s 10 Heuristics (Nielsen 1993) and Schneiderman’s eight golden rules (Shneiderman and Plaisant 2005). Usefully, Sutcliffe and Gault (2004) adapt Nielsen’s heuristics for use in virtual environments, so rather than ‘Feedback’ they suggest ‘Realistic Feedback’. This is a small change, but it acknowledges the importance of realism to the virtual experience because of its contribution to a sense of presence (Witmer and Singer 1998; Nichols, Haldane et al. 2000).

These general rules and adaptations may prove useful, and certainly so in ensuring that more typical usability issues are dealt with. However, it has been suggested that guidance and heuristics developed from graphic user interfaces may be too general, ambiguous, or too high-level to be of practical use (Gabbard 1997; Gabbard, Hix et al. 1999; Bowman, Gabbard et al. 2002). More detailed guidance directly related to development of VEs, ARs, or multimodal environments are also available. This includes, for example, guidance relating to attention (Bearne, Jones et al. 1994), interface objects (Deol 1999), tactile interaction (Challis and Edwards 2001), tactile and kinaesthetic displays (Hale and Stanney 2004), and multimodal interaction (Hale and Stanney 2004).

The difficulty however is that this guidance is limited. In a review of AR literature Swan and Gabbard (2005) found that of the 1104 articles reviewed only 38 addressed HCI and only 21 reported user studies for AR devices. It is suggested that the reason for this is the relative newness of the domain. Whilst Swan’s review was conducted in 2004, some years on this situation seems to have changed little with a similar call for more user-based studies made by the same authors (Gabbard and Swan 2008). Their particular concern is that without these additional studies the development of guidelines and standards will not grow. They point out that user-based experiments are the basis from which informal, then adopted guidelines, and finally standards, develop (see Figure 2-8 for an illustration of this pathway). Given the scarcity of guidelines related to haptic or

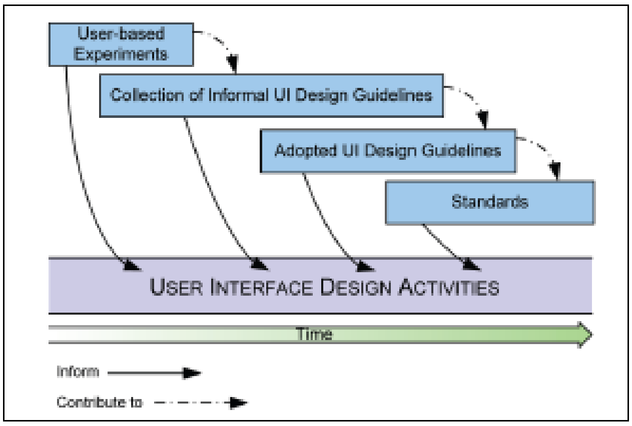


Figure 2-8: Pathway to HCI Standards (Gabbard and Swan 2008)

even multimodal interaction it would seem that as a research community we are presently at the early stages of this process, and that maturity on the form of 'Adopted UI Design Guidelines', let alone standards, is still some way off. This situation is perhaps reflective of the relative newness of haptic technologies and their utilisation within various types of mixed reality environments.

This makes it difficult to find specific guidance for the development of novel haptic interfaces, such as that of SATIN, other than in general terms. This is not unusual, and in many respects reflects the novelty of the device. This should not perhaps be seen as problematic, but more of an opportunity to develop guidance from user studies and so contribute to the generation of knowledge that will eventually solidify into more robust guidance.

2.9 Focus of the Thesis in Relation to the Literature

The literature review has explored the nature of haptic perception of objects. It has found that there are typical ways in which humans interact to discern object shape, and that in doing this a number of properties can be discerned (Davidson 1972; Lederman and Klatzky 1987). It has been shown that curvature is intrinsic to quantifying the 'curvedness' of a shape, and so has helped in understanding human perceptual acuity to changes in curvature and by extension shape (Koenderink and van Doorn 1992; Kappers, Koenderink et al. 1994). This has enabled researchers to identify the absolute threshold for detection of a curved from a flat surface as being a gradient of 0.009

(Gordon and Morison 1982) or a total attitude difference of 2^o (Pont, Kappers et al. 1997). However, there is less certainty about the difference threshold, which has been found to be as little as a 10% (Goodwin, John et al. 1991) or as great as an 86% (Gordon and Morison 1982) change in curvature. This lack of conformance to Weber's Law makes it difficult to predict the threshold for any given curvature. There was also found to be little research into combined visual-haptic thresholds. These two aspects lead to some uncertainty as to the relationship between difference threshold and curvature, and as to the effect of combining modalities on this. This has led to a concern within this thesis to explore the relationship between threshold and stimulus, and to observe the effect on combining modalities on this (Objective 2, see Chapter 1 section 1.3).

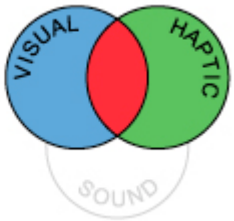
Further difficulties arise when considering multimodality. It has been shown that there are differences between the visual and haptic perceptual spaces (Kappers and Koenderink 1996; Norman, Clayton et al. 2004). Whilst this sets the possibility of conflict, it has also been found that perceptual information is integrated in an optimal fashion, and leads to the dominance of the clearer sense over the other (Heller 1983; Ernst and Banks 2002). The addition of auditory feedback has been found to be confirmatory of visual information, or aided in disambiguating confusing visual information (Sekuler, Sekuler et al. 1997). However, it can also cause erroneous perceptions (Shams, Kamitani et al. 2000). In relation to haptics, the addition of sound has produced compelling illusions that have dominated the haptic sense (Jousmaki and Hari 1998; Guest, Catmur et al. 2002). There is also evidence that the need to attend to more than one modality has a negative impact on performance (Spence, Pavani et al. 2000; Spence, Nicholls et al. 2001). It is therefore unclear as to the effect of auditory feedback on perception of curve shape differences and the level of performance, and so this has been a further focus of this thesis (Objective 1, see Chapter 1 section 1.3). In addition, the field of sonification research is relatively new. The research agenda outlined by Walker (Walker and Nees 2005) makes it clear that more work is required; therefore the research focus here seeks to contribute by ascertaining the appropriateness of sound for this particular type of task (Objective 3, see Chapter 1 section 1.3)

Finally, whilst it was found that there were universally accepted measures of usability (ISO-9241-11 1998; ISO/IEC 25062 2006), and the use of test batteries are common within VEs, these were considered too broad to be of use in contextualising performance in relation to, for example, the SATIN interface. It was also found that the relative newness of AR, means that many more user studies are required before recommendations emerge and coalesce into guidelines (Gabbard and Swan 2008).

Therefore, the research within this thesis seeks to contribute to this evolution (Objectives 4 and 5, see Chapter 1 section 1.3).

Chapter 3: Visual-Haptic Perception and Performance

3.1 Introduction and Rationale



This chapter starts by discussing issues with the assessment, modification, and perception of curvature differences in relation to product design. It identifies the need to know the acuity of modalities in order to provide effective feedback, and to know whether these conform to Weber's Law. It then goes on to discuss the development of metrics for the assessment of multimodal interfaces within this domain. There follows a presentation of the experimental work that was undertaken to explore these issues, and finally the results of this study are discussed in relation to the aims identified in section 3.1.3.

3.1.1 Shape Assessment, Modification, and the Perception of Curvature Differences

As outlined in Chapter 1, a problem facing product designers is assessing the geometric quality of a shape in terms of its adherence to 'Class A' surface tolerances (see Chapter 1 Section 1.1.3). One aspect of these is that they require continuity of curvature across the object surface. Therefore for various points of the model a designer needs to assess if there are significant differences or "discontinuities" in the curvature. Specification for the design of such multimodal interfaces for virtual shape exploration requires an understanding of human perceptual performance of curvature differentiation in each of the proposed interaction modes. There is therefore a requirement to know how well humans can perceive differences in curvature, and how this perception is affected when interaction is multimodal (visual and haptic) as opposed to unimodal (visual or haptic). The specification of the required difference threshold allows interface developers to determine the accuracy and detail required in haptic feedback. For example, should the perceptible haptic difference be in the order of 60% (Henriques and Soechting 2003), there would be little benefit in a physical interface that could accurately convey differences of 1%. Additionally, by identifying the point at which users are less able to immediately determine the differences between curvature profiles (discontinuities), consideration could be given to augmenting visual-haptic information with an additional mode such as sound.

Having assessed an object for discontinuities, as described above, the next step would be to modify the object to eradicate such anomalies. This would require the designer to make an adjustment to the object. The problem here is that the amount of adjustment required in order to rectify the discontinuity may be too small to be perceptible. So a situation may arise where, in making a perceivable adjustment, the discontinuity may be maintained or even increased, thus the problem may not be resolved or could even be made worse. To avoid this situation, or to provide support, it is necessary to know the acuity of human perception in relation to curvature change. More widely this is a problem for any modification in shape, as there is always the question of "has the shape changed?" By examining the acuity of perception for curvature change it is also possible to measure more objectively the notion of shape change. This is because curvature is a precise way of mathematically describing the flatness or curvedness of a given shape. Therefore, to know how acutely curvature change can be perceived, is to know how sensitive human perception is to shape change.

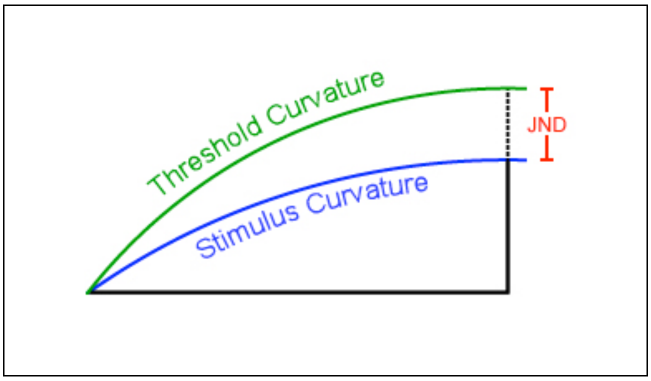


Figure 3-1: Difference Threshold – Stimulus must be increase by a just noticeable difference to point of threshold curvature before change can be perceived.

In terms of the interface, what needs to be gauged is whether the level of feedback given in the visual-haptic modality (for any modifications made) is sufficient to be perceived (and whether this is better or worse than the unimodal condition). For this it would be necessary to know for any given curvature the amount of change needed to give a just noticeable difference (JND). In other words we need to be able to predict the level of acceptable feedback in order to develop a useable interface. Within the literature it is suggested that curvature perception does not conform to Weber's Law for haptic stimuli (Kappers and Koenderink 1996). This means that the JND is not a constant percentage of the stimulus. This finding means that for each curvature a different amount of adjustment may be needed in order that a change is perceived. However, there is also evidence within the literature that for some visual stimuli there was a conformance with Weber's Law (Johnston and Passmore 1994). It is therefore difficult to predict whether the combination of the visual and haptic modalities would or would not conform to Weber's Law.

Given the diversity of results within the haptic literature, there is the possibility that there may be conformance to Weber's Law within the desired range, or that the Weber Fractions are sufficiently close that for practical purposes an averaged figure would be acceptable. If not, then alternative measures could be used to make predictions, for example, the results for high and low curvatures may be used to set a useful range. If we know that the required level of change is not within these bounds (because adjustment for discontinuity correction is ultra-fine) then consideration can be given as to how visual-haptic perception could be augmented in order to improve feedback.

3.1.2 Performance Measures and the Development of Metrics

The previous two points have discussed issues of acuity and how these affect assessment and modification of shapes. A third area of concern in relation to the development of multimodal interfaces is that of performance. In particular the concern is what constitutes an acceptable level of performance? Whilst there is some indication of the levels of acuity that might be expected within the literature, to the knowledge of the author, there is an absence of information concerning the level of performance that might be expected in the judgement of differences between curvatures (or shapes). This is an important factor when evaluating and comparing interfaces as their efficiency and effectiveness are intrinsic to any consideration of their usability (Nielsen 1993; ISO-9241-11 1998). For this reason it was considered essential that any experimental work should also consider performance as indicated by measures of effectiveness and efficiency.

Whilst these two measures are commonly used indicators of usability, it was thought that a further measure, confidence, was also required. Since it is vital that a user feels secure in their interpretation of the information presented by a system, particularly in a high-cost context such as design evaluation, the level of confidence achieved with a particular system is of significant concern. As with the previous two measures it is presently difficult to know what an acceptable level of confidence might be for particular judgements of curvature or shape difference.

It was considered that gathering data on these three measures would indicate the levels of performance (in terms of response time, accuracy, and confidence) that might be expected for judgements involving high and low curvatures. In this way, examination of these measures could be used to define general metrics for the assessment of interface performance. These could be used to indicate performance within a given curvature range for predetermined curvature differences. However, it could not be used more precisely to give metrics for different curvatures within that range or at various curvature differences. In order to provide this level of flexibility it would be necessary to customise the parameters for the evaluation to be undertaken, say for a specific difference in curvature. In order to facilitate this it would be necessary for there to be some form of correlation between the performance measure and curvature difference. So that, for any given curvature the performance could be judged against a specific rather than a general metric. It is known from psychophysics methods that accuracy has a strong relationship to differences between stimuli; so that as the comparison stimuli increases detection becomes easier and so accuracy increases. What is of interest here is to establish if response time and confidence also behave in a similar way.

3.1.3 Aims

Five research questions were posed at the start of this study:

1. What is the level of perceptual acuity for haptic, visual, and combined modalities in recognising changes in low and high curvatures? Is any one modality better than the other, and is there an effect on acuity by combining modalities?
2. Do measured difference thresholds conform to Weber's Law and is this the case for all modalities?
3. Is performance (response time, accuracy, confidence) dependent upon the level of curvature, modality used, or extent of magnitude difference judged? Is any one modality better than the other, and is there an effect on performance of combining modalities? Are there characteristic traits of performance i.e. are response times for some types of judgement quicker than others?
4. For each modality, would it be possible to predict the level of performance at the point of JND for any curvature?
5. Is the amount of interaction (number of comparisons) affected by different curvatures, modalities, or extent of magnitude difference judged? Is there a relationship between the level of interaction and other performance measures?

3.2 Method (Study 1)

3.2.1 Participants

Eight participants, four female and four male, completed the study. Seven of these had participated in a pilot study (haptic only) and so were non-naïve to the study procedure and haptic exploration of curvature. The participants were from a broad range of backgrounds, although none were product designers. In total each participant undertook 3 x 3 hour sessions and for which they were given payment of £10 per hour (£90 in total).

3.2.2 Equipment and Set up

The equipment was set up as shown in Figure 3-2. Participants were seated on an adjustable chair throughout the experiment. At the beginning of the first session this was adapted in order to ensure that they were comfortable throughout the session. The chair offered back and arm support (although they were asked not to support their arms during presentations). The stimuli blocks were presented in a clear holder that was fixed against a black surface. This was arranged so that the participant could easily reach the blocks to be examined.

Figure 3-2: Experimental set-up as seen from participant view point.

3.2.3 Procedure

The procedure for each of the modality conditions was similar, except that:

- For the haptic condition participants were blindfolded and explored the curve using touch alone. They were restricted to using the index finger of the dominant hand and were asked not to rest their hand or arm on the table or chair.

- For the Visual condition participants explored the curves using vision from an upright position and at touching distance from the blocks (they were asked not to touch the blocks)
- For the Visual-haptic condition participants used both visual and haptic exploration to judge the curves as described above.

For each of these conditions there was one three hour session. This was broken down into two separate trials (corresponding to the two types of curvature (high and low) – see 2.1 Stimuli). Each trial lasted approximately 80 minutes and was broken down into three 20 minute blocks. Between each block participants had a three minute break and between each trial there was a ten minute break. In addition to this, participants were advised that they could stop at any point to take a break or withdraw from the trial altogether. Each session was held on different days in order to ensure that participants had sufficient break between sessions. No two sessions were on the same or consecutive days. All documentation used to support this study can be found on the accompanying CD, as detailed in Appendix B.

For each curvature type, the stimuli were presented in pairs; the standard stimulus and a comparison stimulus. The order of the presentations was randomized, and each comparison stimulus was presented twelve times (six on the right, and six on the left). Participants were asked to touch/view the left hand block first, and not give a judgement until they had touched/viewed both blocks. They could examine the stimuli as many times as they wished, however a time limit of 30 seconds was applied. For each presentation participants were asked to judge if the stimulus on the right was ‘more’ or ‘less’ curved than the stimulus on the left. Having given their answer they were then asked to state how confident they were in this answer, responding either ‘high’ or ‘low’ confidence. For each presentation a record was kept of their answers, the response time, and the number of times the participant compared the blocks. In all 96 presentations were made for each curvature type in each modality. A total of 576 judgements were made by each participant.

3.2.4 Stimuli

Two types of stimuli were used in the study, one with a low curvature and one with high curvature. These were chosen to represent curvatures from different sized ‘everyday’ objects. For larger objects, a low curvature was chosen which had a standard stimuli of 1.46/m (radius 68.53cm), and eight comparison stimuli of 0.94, 1.07, 1.21, 1.37, 1.55, 1.76, 2, and 2.27/m (radii ranging from 106.95cm to 44.1cm). For smaller objects, a high curvature was chosen which had a standard stimuli of 15.97/m (radius 6.26cm), and eight comparison stimuli of 10.28, 11.66, 13.22, 15, 17.01, 19.29, 21.88/m (radii ranging from 9.73cm to 4.03cm). The comparison stimuli range was four above and four below the standard stimulus for both high and low curvature. The differences between stimuli were an increment of 13.5% of the previous stimulus and so provided a range that encompassed the maximum Weber fraction reported in the literature (Kappers and Koenderink 1996). This also meant that for both low and high curvatures percentage differences from the standard were the same, and so would provide a useful means of comparing results across curvatures.

The stimulus blocks were made from a nylon composite material to ensure that the surface texture of each block was uniform and smooth. The dimensions for high and low curvature were the same in respect of the block's central-height (5cm) and depth (2cm). Block width for each curvature differed; for low curvatures it was 20cm, and for high curvatures it was 5cm (see Figure 3-3). Exact specifications for all the stimulus blocks used in this study can be found in Appendix C.

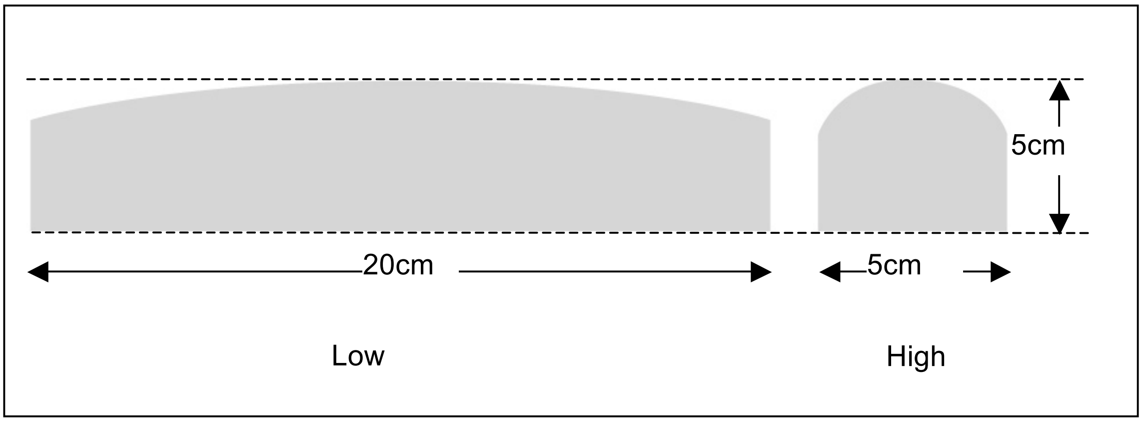


Figure 3-3: Dimensions of stimulus blocks for low and high curvature

3.2.5 Design and Hypotheses

The experiment was a repeated-measures design, and applied the psychophysical method of constant stimuli to establish the difference threshold for curvature in visual, haptic, and visual-haptic conditions. It also sought to examine the effects of three independent variables; curvature (low, high), modality (haptic, visual, visual-haptic), and magnitude difference (small, medium, large)[2] on performance (response time, accuracy, and confidence) and interaction (comparisons made between stimuli). In order to answer the research questions posed (see 3.1.3) the following hypotheses were examined:

3.2.5.1 Perception (H1-H3)

H1: There is an effect of curvature on difference threshold.
H2: There is an effect of modality on difference threshold.
H3: There is an interaction effect between curvature and modality on difference threshold.

3.2.5.2 Performance (H4-H21)

H4-H7: There is an effect of curvature on response time; accuracy; confidence.
H8-H9: There is an effect of modality on response time; accuracy; confidence.
H10-12: There is an effect of magnitude difference on response time; accuracy; confidence.
H13-H15: There is an interaction effect between curvature and modality on response time; accuracy; confidence.
H16-H18: There is an interaction effect between curvature and magnitude difference on response time; accuracy; confidence.
H19-H21: There is an interaction effect between modality and magnitude difference on response time; accuracy; confidence.

3.2.5.3 Predictability (H22-H37)

H22-29: For each modality and all modalities combined, there is a correlation between response time and accuracy.
H30-37: For each modality and all modalities combined, there is a correlation between response time and confidence.

---

[2] The three levels of magnitude difference were used in analysis of performance measures. These were derived by combining the comparison stimuli (see data analysis for details). For determination of thresholds data from each comparison stimuli was used.

3.2.5.4 Performance Characteristics (H38-H85)

H38-H45: For each curvature (high, low), modality (haptic, visual, visual-haptic), magnitude difference (small, medium, large) and all conditions combined, there is a significant difference in response time between correct and incorrect judgements.
H46-H53: For each curvature (high, low), modality (haptic, visual, visual-haptic), magnitude difference (small, medium, large) and all conditions combined, there is a correlation between response time and accuracy.
H54-H61: For each curvature (high, low), modality (haptic, visual, visual-haptic), magnitude difference (small, medium, large) and all conditions combined, there is a significant difference in response time between high and low confidence judgements.
H62-H69: For each curvature (high, low), modality (haptic, visual, visual-haptic), magnitude difference (small, medium, large) and all conditions combined, there is a correlation between response time and confidence.
H70-H77: For each curvature (high, low), modality (haptic, visual, visual-haptic), magnitude difference (small, medium, large) and all conditions combined, there is a significant difference in response time between high confidence correct judgements and low confidence correct; high confidence incorrect; low confidence incorrect.
H78-H85: For each curvature (high, low), modality (haptic, visual, visual-haptic), magnitude difference (small, medium, large) and all conditions combined, there is a correlation between accuracy and confidence.

3.2.5.5 Interaction (H86-H99)

H86: There is an effect of curvature on the number of comparisons made between the standard and comparison stimuli.
H87: There is an effect of modality on the number of comparisons made between the standard and comparison stimuli.
H88: There is an effect of magnitude difference on the number of comparisons made between the standard and comparison stimuli.
H89: There is an interaction effect between curvature and modality on the number of comparisons made between the standard and comparison stimuli.
H90: There is an interaction effect between curvature and magnitude difference on the number of comparisons made between the standard and comparison stimuli.
H91: There is an interaction effect between modality and magnitude difference on the number of comparisons made between the standard and comparison stimuli.
H92-99: For each modality and all modalities combined, there is a correlation between the number of comparisons and curvature difference.

3.2.6 Data Analysis

For determining difference thresholds it was first necessary to calculate from the raw data the number of times each stimulus was identified as being 'more' curved than the standard (expressed as a percentage of the total presentations made of the given stimulus). These percentages were then converted into z-scores. Using the statistics package SPSS, the relationships between z-scores and curvature were analysed using linear regression (see Appendix E for details). The resulting coefficients were used to generate individual difference thresholds for each participant in each modality (see Table 3-2 and Table 3-3). The thresholds generated were found to be normally distributed and so were further examined using analysis of variance (ANOVA) to determine compliance with Weber's Law and similarity of thresholds in different modalities.

For the purpose of analysing performance, the data from individual stimuli were aggregated into groups defined by their difference from the standard stimulus; small, medium, and large. This would enable more meaningful analysis of the performance data, and would enable comparison between the two types of curvature.

Table 3-1: Categorisation of stimulus difference (from standard)

Curvature (/m)		Difference	
Low	High	%	Category
0.94	10.28	-36	Large
1.07	11.66	-27	Medium
1.37	15.00	-6	Small
1.55	17.01	6	Small
1.76	19.29	21	Medium
2.27	24.82	55	Large

The performance data (response time, accuracy, and confidence) were found to be non-normally distributed for some variables; either positive or negative skew which is usual for these types of data. This was considered not to be a problem as ANOVAs are considered to be a robust method of statistical analysis (Davies 1956; Field 2009)(Appendix F for discussion of this). A number of pre-planned contrasts were used in the analysis of these data. For modality, all were contrasted to 'Visual-Haptic' as this was the aggregation of visual and haptic performances so it was of interest to know how each varied from this. For magnitude difference all were contrasted with 'small' which was likely to show the lowest performance and so improvement from this could be gauged. For other statistical tests the appropriate parametric analysis methods were used for non-normal variables (the tests used are noted within the results).

Response time data were excluded for participant seven as there was found to be extreme outliers for several variables (for example, visual low curvature z-score = 14.68). Data for other performance measures for this participant were found to be acceptable (z-score <1.96).

3.3 Results (Study 1)

This section details all results in detail, a summary of the results in relation to the hypotheses can be found in Appendix D.

3.3.1 Perception

For high and low curvature, a mean threshold curvature was calculated from the individual thresholds of each participant as shown in Table 3-2 and Table 3-3 (see Appendix E for calculation of individual thresholds and model statistics). In addition, mean just noticeable difference (JND) and Weber Fractions (WF) were also calculated.

Table 3-2: Threshold Curvature, JND and Weber Fractions for Low Curvature Stimulus

P-ID	Haptic			Visual			Visual-Haptic		
	Threshold	JND	WF	Threshold	JND	WF	Threshold	JND	WF
P1	1.84	0.32	21.79	1.85	0.33	22.72	1.85	0.33	22.79
P2	1.84	0.32	21.94	1.85	0.33	22.56	1.85	0.33	22.41
P3	1.85	0.33	22.77	1.86	0.34	23.17	1.86	0.33	22.84
P4	1.86	0.33	22.91	1.85	0.33	22.62	1.85	0.33	22.81
P5	1.86	0.34	23.31	1.86	0.34	23.24	1.84	0.32	21.80
P7	1.85	0.33	22.64	1.86	0.34	23.11	1.84	0.32	21.86
P8	1.85	0.33	22.48	1.85	0.33	22.32	1.85	0.33	22.81
P9	1.84	0.32	21.64	1.85	0.33	22.62	1.85	0.33	22.39
Mean	1.85	0.33	22.43	1.85	0.33	22.79	1.85	0.33	22.46
SD	0.01	0.01	0.59	0.01	0.00	0.33	0.01	0.01	0.43

Table 3-3: Threshold Curvature, JND and Weber Fractions for High Curvature Stimulus

P-ID	Haptic			Visual			Visual-Haptic		
	Threshold	JND	WF	Threshold	JND	WF	Threshold	JND	WF
P1	20.52	3.85	24.11	20.46	3.83	23.97	20.51	3.85	24.11
P2	20.29	3.62	22.68	20.34	3.70	23.18	20.50	3.85	24.11
P3	20.37	3.74	23.44	20.26	3.66	22.93	20.32	3.72	23.31
P4	20.22	3.60	22.56	20.45	3.76	23.57	20.30	3.64	22.80
P5	20.46	3.79	23.70	20.25	3.58	22.44	20.45	3.81	23.84
P7	20.38	3.72	23.31	20.33	3.70	23.18	20.55	3.85	24.11
P8	20.24	3.56	22.32	20.28	3.66	22.93	20.22	3.58	22.44
P9	20.23	3.62	22.68	20.37	3.74	23.44	20.34	3.70	23.18
Mean	20.34	3.69	23.10	20.34	3.71	23.20	20.40	3.75	23.49
SD	0.11	0.10	0.63	0.08	0.07	0.47	0.12	0.10	0.65

A repeated-measures ANOVA revealed that there was a main effect of curvature on Weber Fractions, $F(1,7)=13.84$, $p<0.01$, partial $\eta^2=0.66$ (see Figure 3-4). Pre-planned contrasts showed that Weber Fractions for low curvature (M=22.56%, SD=0.47) were significantly smaller than for high curvature (M=23.26%, SD=0.59), $F(1,7)=13.84$, $p<0.01$, partial $\eta^2=0.66$. There was no significant main effect of modality on Weber Fractions, $F(2,14)=1.57$, $p=0.24$, partial $\eta^2=0.18$. This means that JNDs for different modalities were not significantly different.

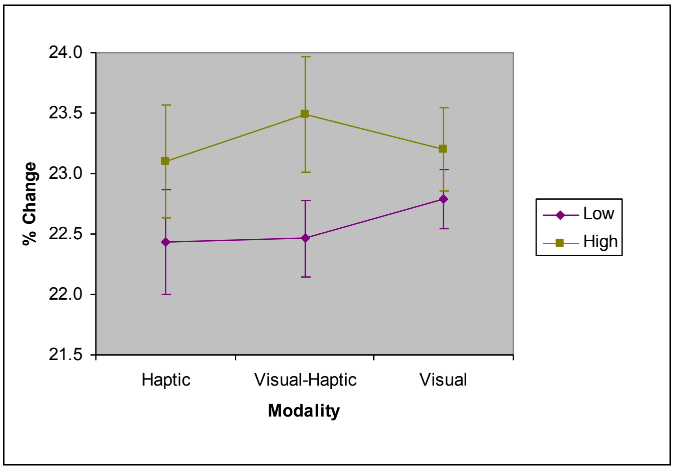


Figure 3-4: Effect of curvature on Weber Fractions

Note: Error bars represent the 95% Confidence Interval of the mean, and do so throughout

3.3.2 Performance

3.3.2.1 Response Time

A repeated-measures ANOVA compared the effect on response time of curvature, modality, and magnitude difference (see Figure 3-5 and Table 3-4) . This revealed that there was a main effect of curvature, $F(1,6)=8.23$, $p=0.03$, partial $\eta^2=0.58$. Pre-planned contrasts showed that response time for ‘High Curvature’ was significantly faster than for ‘Low Curvature’, $F(1,6)=8.23$, $p=0.03$, partial $\eta^2=0.58$. There was a main effect of modality, $F(2,12)=53.95$, $p<0.01$, partial $\eta^2=0.90$. Pre-planned contrasts showed that ‘Visual-Haptic’ had a significantly quicker response time than ‘Haptic’ but was not significantly different to ‘Visual’, $F(1,6)=48.8$, $p<0.01$, partial $\eta^2=0.89$, and $F(1,6)=5.12$, $p=0.06$, partial $\eta^2=0.46$, respectively. There was a main effect of magnitude difference, $F(1.05,6.23)=83.55$, $p<0.01$, partial $\eta^2=0.93$. Pre-planned contrasts showed that ‘Small’ had a significantly slower response time than either ‘Medium’ or ‘Large’ magnitude differences, $F(1,6)=55.19$, $p<0.01$, partial $\eta^2=0.90$, and $F(1,6)=89.87$, $p<0.01$, partial $\eta^2=0.94$, respectively.

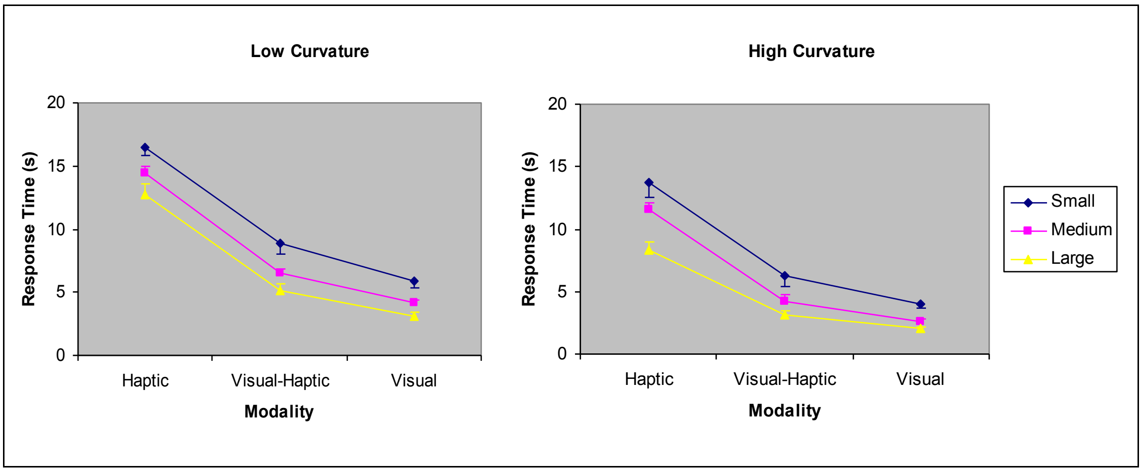


Figure 3-5: Effect of curvature, modality, and magnitude difference on response time

Note: Error bars represent the 95% Confidence Interval of the mean, and do so throughout

Table 3-4: Response times for different levels of curvature, modality, and magnitude difference

Variable	Mean	SD
Low Curvature	8.59	1.88
High Curvature	6.21	1.29
Haptic	12.88	2.30
Visual	3.63	0.44
Visual-Haptic	5.69	2.19
Small Difference	9.20	1.53
Medium Difference	7.25	1.13
Large Difference	5.74	1.02
All	7.40	1.18

There were no significant interactions between curvature, and either modality or magnitude difference, $F(2,12)=0.76$, $p=0.49$, partial $\eta^2=0.11$, and $F(2,12)=0.75$, $p=0.75$, partial $\eta^2=0.05$, respectively. There was a significant interaction between modality and magnitude difference, $F(4,24)=6.55$, $p<0.001$, partial $\eta^2=0.52$. However, there were no significant differences within the pre-planned contrasts. Finally, there was a significant three-way interaction effect on response time between curvature, modality, and magnitude difference, $F(4,24)=3.50$, $p=0.02$, partial $\eta^2=0.37$. Again, pre-planned contrasts did not reveal any significant differences.

A series of scatter plots were used to compare curvature difference against response time for each modality in high and low curvature conditions (see Figure 3-6). These indicated a correlation between the two variables. Correlation analysis (Pearson’s r) confirmed that there was a negative correlation between curvature difference and response time, so that as percentage curvature difference increased, response time decreased. This was found for haptic (r=-0.97, r=-0.96), visual (r=-0.91, r=-0.94), and visual-haptic (r=-0.95, r=-0.91) modalities in low and high curvature conditions, p<0.01 for all. A further scatter plot compared response time for modalities combined against percentage curvature difference for low and high curvatures (see Figure 3-7). This also indicated that there was a correlation between the two variables. Correlation analysis (Pearson's r) confirmed that there was a negative correlation between percentage curvature difference and response time in low and high curvature conditions, r=-0.97, p<0.01 for both.

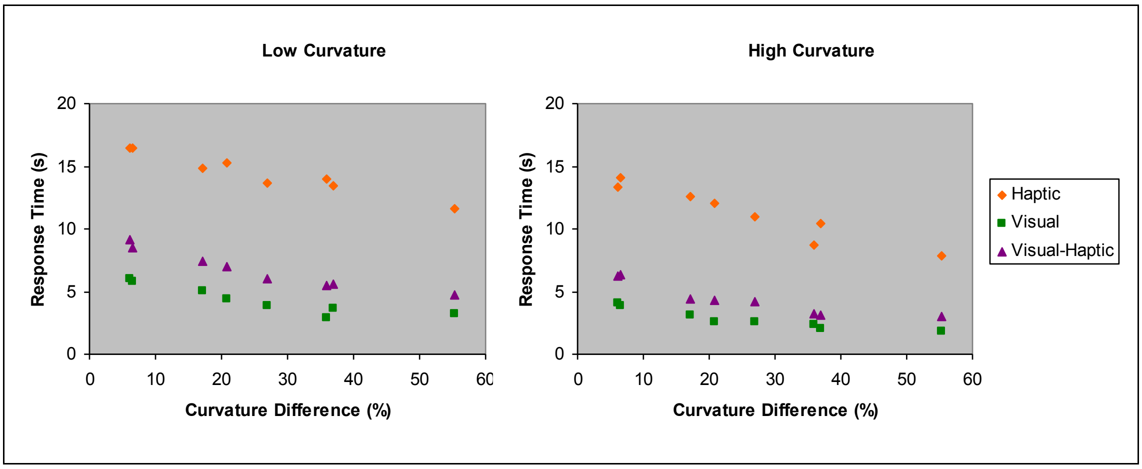


Figure 3-6: Correlation between curvature difference and response time for haptic, visual, and visual-haptic modalities

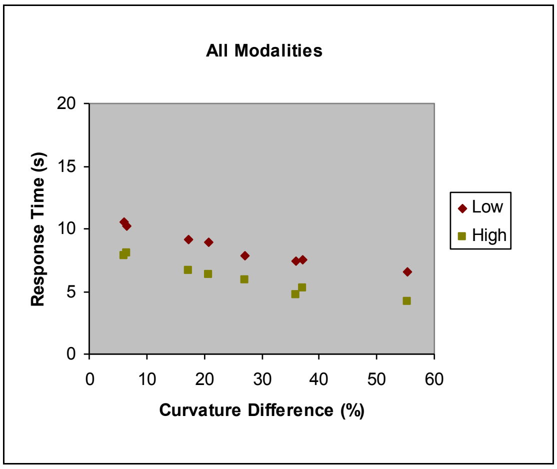


Figure 3-7: Correlation between curvature difference and response time for all modalities with low and high curvatures

3.3.2.2 Accuracy

A repeated-measures ANOVA compared the effect on accuracy of curvature, modality, and magnitude difference (see Figure 3-8 and Table 3-5). This revealed that there was a main effect of curvature, F(1,7)=47.85, p<0.01, partial η^2=0.87. Pre-planned contrasts showed that accuracy for 'High Curvature' was significantly greater than for 'Low Curvature', F(1,7)=47.85, p<0.01, partial η^2=0.87. There was a main effect of modality, F(2,14)=6.27, p=0.01, partial η^2=0.47. Pre-planned contrasts showed that 'Visual-Haptic' had a significantly greater accuracy than 'Haptic' but was not significantly different to 'Visual', F(1,7)=15.71, p<0.01, partial η^2=0.69, and F(1,7)=3.43, p=0.11, partial η^2=0.33, respectively. There was a main effect of magnitude difference, F(2,14)=249.35, p<0.01, partial η^2=0.97. Pre-planned contrasts showed that 'Small' had a significantly lower accuracy than either 'Medium' or 'Large' magnitude differences, F(1,7)=172.36, p<0.01, partial η^2=0.96, and F(1,7)=490.65, p<0.01, partial η^2=0.99, respectively.

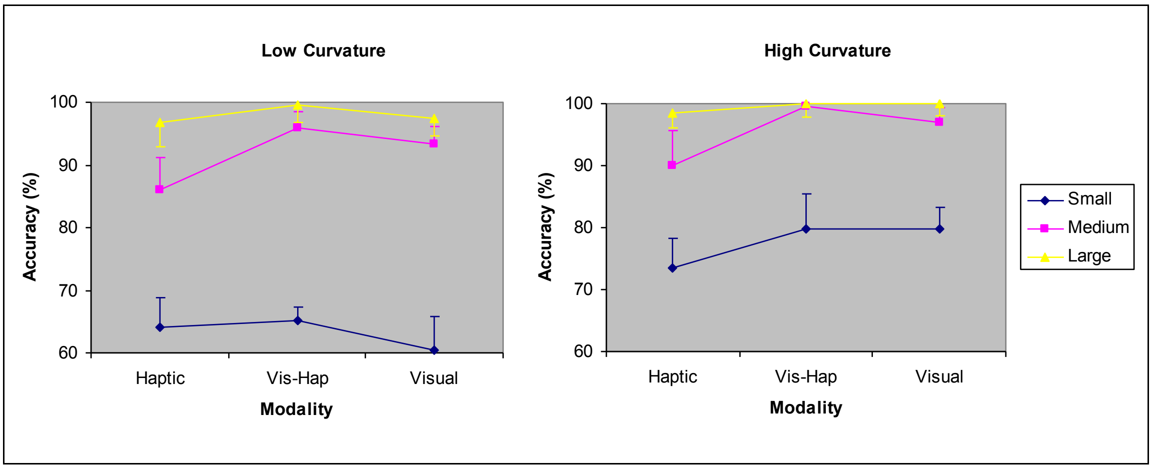


Figure 3-8: Effect of curvature, modality, and magnitude difference on accuracy

Table 3-5: Accuracy (%)for different levels of curvature, modality, and magnitude difference

Variable	Mean	SD
Low Curvature	84.26	2.69
High Curvature	90.86	1.98
Haptic	84.81	3.45
Visual	87.93	3.19
Visual-Haptic	89.93	2.13
Small Difference	70.40	3.58
Medium Difference	93.58	2.99
Large Difference	98.70	1.59
All	87.56	1.80

There were no significant interactions between modality and either curvature or magnitude difference, F(2,14)=0.69, p=0.52, partial η^2=0.09, and F(4,28)=1.55, p=0.21, partial η^2=0.18, respectively. There was a significant interaction between curvature and magnitude difference, $F(2,14)=27.02$, $p<0.01$, partial η^2=0.79 (see Figure 3-9). Pre-planned contrasts revealed that the increase in accuracy for high curvature (compared to low) was greater for 'Small' differences compared to either 'Medium' or 'Large', $F(1,7)=20.89$, $p<0.01$, partial η^2=0.75, and $F(1,7)=49.13$, $p<0.01$, partial η^2=0.87, respectively. Finally, there was no significant three-way interaction effect on accuracy between curvature, modality, and magnitude difference, $F(4,28)=1.25$, $p=0.31$, partial η^2=0.15.

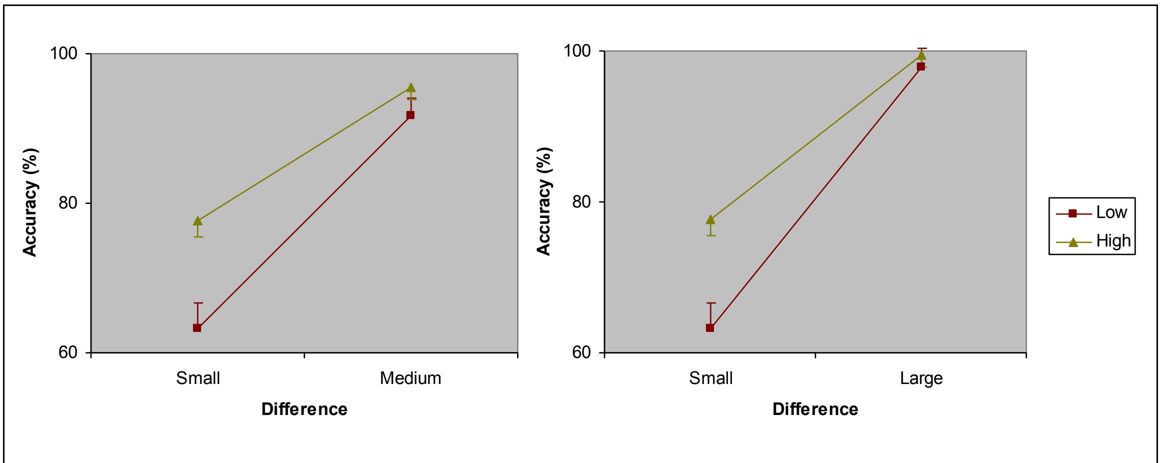


Figure 3-9: Interaction effect of curvature and magnitude difference on accuracy

A series of scatter plots were used to compare curvature difference against accuracy for each modality in high and low curvature conditions (see Figure 3-10). These indicated a correlation between the two variables. Correlation analysis (Pearson's r and Spearman's rho[3]) confirmed that there was a positive correlation between curvature difference and accuracy, so that as percentage curvature difference increased so did accuracy. This was found for haptic (r=0.89, r=0.84), visual (r=0.80, r=0.84), and visual-haptic (r=0.80, r_s=0.87) modalities in low and high curvature conditions, $p<0.05$ for all. A further scatter plot compared accuracy for modalities combined against percentage curvature difference for low and high curvatures (see Figure 3-11). This also indicated that there was a correlation between the two variables. Correlation analysis (Pearson's r and Spearman's rho[4]) confirmed that there was a positive correlation between percentage curvature

[3] Spearman's rho was used for non-normally distributed variables; visual-haptic (high).

[4] Spearman's rho was used for non-normally distributed variables; all modalities (high).

difference and accuracy in low and high curvature conditions, r=0.86, and r_s=0.87 respectively, p<0.01 for both.

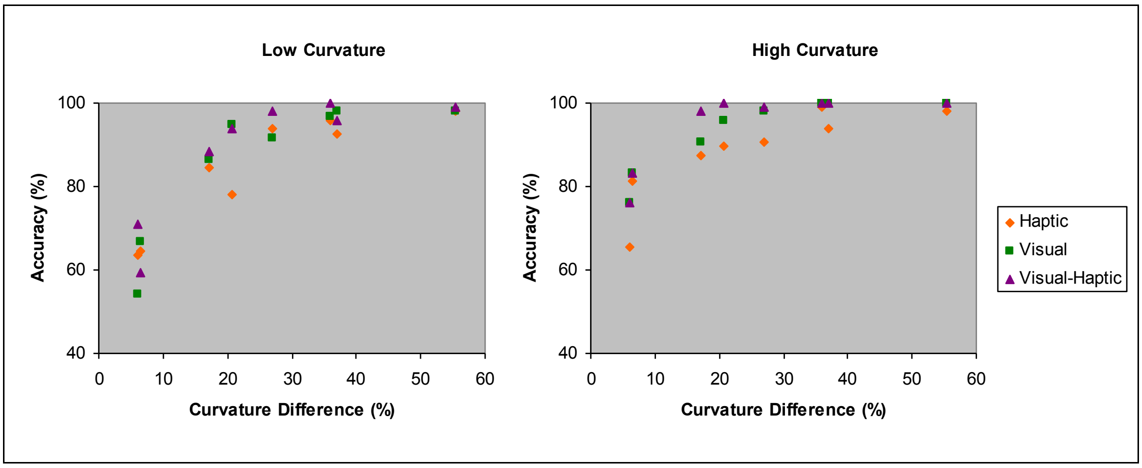


Figure 3-10: Correlation between curvature difference and accuracy for haptic, visual, and visual-haptic modalities

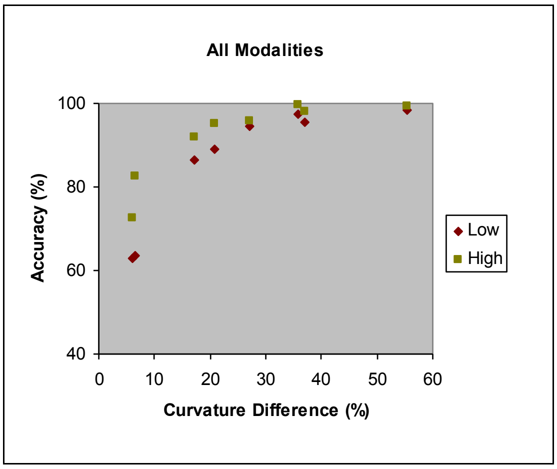


Figure 3-11: Correlation between curvature difference and accuracy for all modalities with low and high curvatures

3.3.2.3 Confidence

A repeated-measures ANOVA compared the effect on confidence of curvature, modality, and magnitude difference (see Figure 3-12 and Table 3-6). This revealed that there was a main effect of curvature, F(1,7)=21.24, p<0.01, partial η^2=0.75. Pre-planned contrasts showed that confidence for 'High Curvature' was significantly higher than for 'Low Curvature', F(1,7)=21.24, p<0.01, partial η^2=0.75. There was a main effect of modality, F(2,14)=16.60, p<0.01, partial η^2=0.70. Pre-planned contrasts showed that 'Visual-Haptic' had significantly higher confidence than 'Haptic' and 'Visual', $F(1,7)=23.91$, $p<0.01$, partial $\eta^2=0.77$, and $F(1,7)=5.67$, $p=0.049$, partial $\eta^2=0.45$, respectively. There was a main effect of magnitude difference, $F(1.08,7.57)=49.91$, $p<0.01$, partial $\eta^2=0.88$. Pre-planned contrasts showed that 'Small' had significantly lower confidence than either 'Medium' or 'Large' magnitude differences, $F(1,7)=39.65$, $p<0.01$, partial $\eta^2=0.85$, and $F(1,7)=54.09$, $p<0.01$, partial $\eta^2=0.88$, respectively.

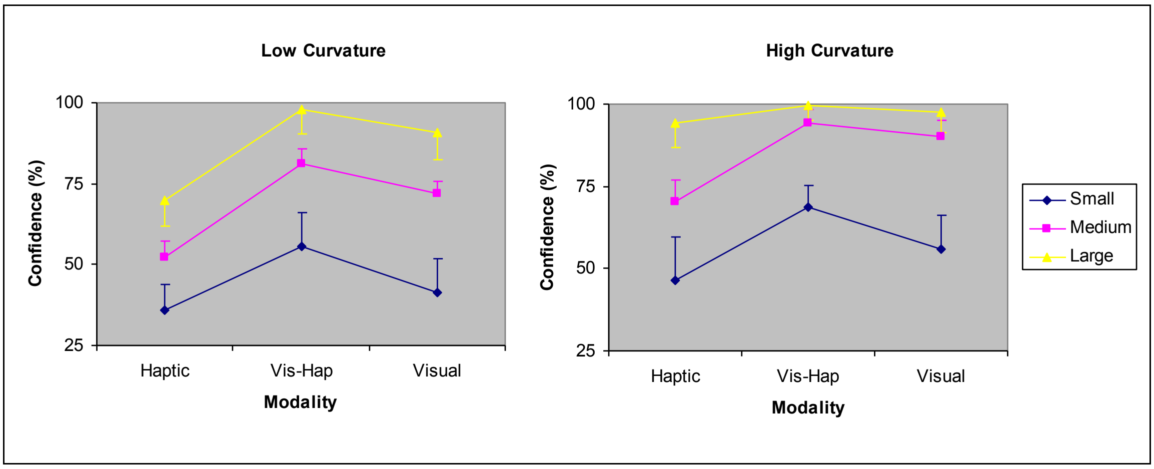


Figure 3-12: Effect of curvature, modality, and magnitude difference on confidence

Table 3-6: Confidence (%) for different levels of curvature, modality, and magnitude difference

Variable	Mean	SD
Low Curvature	66.26	13.75
High Curvature	79.63	9.13
Haptic	61.46	15.06
Visual	74.48	13.66
Visual-Haptic	82.90	7.57
Small Difference	50.61	18.99
Medium Difference	76.65	10.19
Large Difference	91.58	5.61
All	72.95	10.92

There were no significant interactions between curvature, and either modality or magnitude difference, $F(1.11,7.75)=1.55$, $p=0.25$, partial $\eta^2=0.18$, and $F(2,14)=0.71$, $p=0.51$, partial $\eta^2=0.09$, respectively. There was a significant interaction between modality and magnitude difference, $F(4,28)=2.99$, $p=0.036$, partial $\eta^2=0.30$. However, there were no significant differences within the pre-planned contrasts[5]. Finally, there was no significant three-way interaction effect on confidence between curvature, modality, and magnitude difference, $F(4,28)=1.88$, $p=0.14$, partial $\eta^2=0.21$.

A series of scatter plots were used to compare curvature difference against confidence for each modality in high and low curvature conditions (see Figure 3-13). These indicated a correlation between the two variables. Correlation analysis (Pearson's r) confirmed that there was a positive correlation between curvature difference and confidence, so that as percentage curvature difference increased so did confidence. This was found for haptic (r=0.92, r=0.94), visual (r=0.94, r=0.88), and visual-haptic (r=0.93, r=0.86) modalities in low and high curvature conditions, p<0.01 for all. A further scatter plot compared confidence for modalities combined against percentage curvature difference for low and high curvatures (see Figure 3-14). This also indicated that there was a correlation between the two variables. Correlation analysis (Pearson's r) confirmed that there was a positive correlation between percentage curvature difference and confidence in low and high curvature conditions, r=0.97, and r=0.92 respectively, p<0.01 for both.

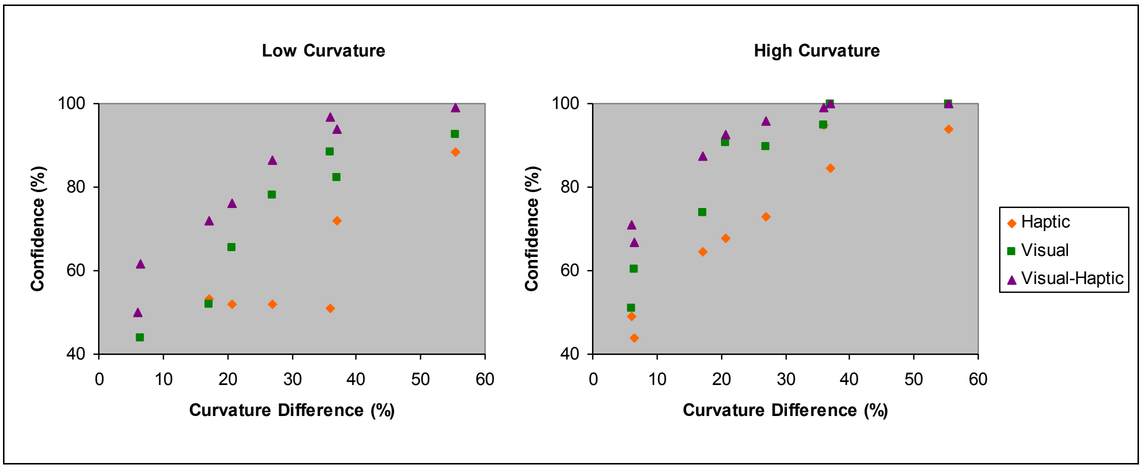


Figure 3-13: Correlation between curvature difference and confidence for haptic, visual, and visual-haptic modalities

[5] The number of contrasts was limited as multiple tests increase the risk of type I error. However on occasion this means that a significant result remains unexplained as it does not occur between the variables of the pre-planned contrasts.

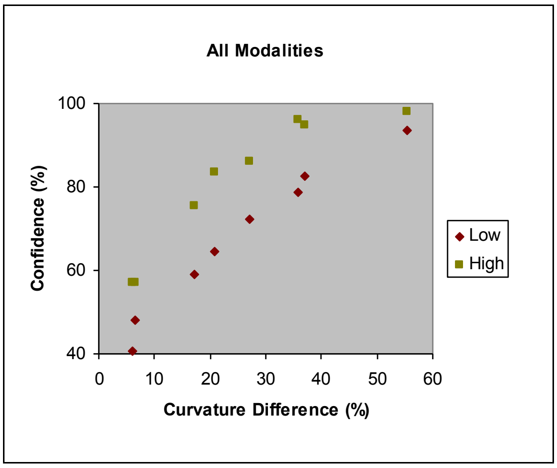


Figure 3-14: Correlation between curvature difference and confidence for all modalities with low and high curvatures

3.3.3 Performance Characteristics

3.3.3.1 Accuracy and Response Time

Dependent t-tests were used to examine the differences between correct and incorrect response times for different levels of curvature, modality, and magnitude difference. These revealed that response times for correct answers were significantly faster than incorrect for low curvature, high curvature, visual[6] ($z=-2.37$, $p=0.01$), visual-haptic, and all conditions (see Figure 3-15 and Table 3-7 for mean response times and test statistics).

A series of scatter plots were used to compare response time against accuracy for each curvature, modality and magnitude difference. These did not appear to indicate any correlation between accuracy and response time. Further analysis was undertaken (Pearson's r and Spearman's rho[7]) which confirmed that there was no correlation between response time and accuracy for any of the variables tested, $p>0.05$.

[6] The differences between correct and incorrect response times for 'visual' were non-normally distributed so the *Wilcoxon signed-rank Test* was used.

[7] Spearman's rho was used for non-normally distributed variables; high curvature, and visual-haptic.

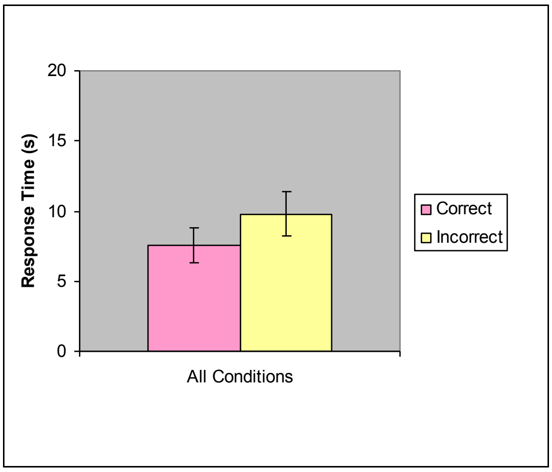


Figure 3-15: Correct and incorrect response times for 'All Conditions'.

Table 3-7: Correct and incorrect response times, and significance of the difference between them

Variables	RT Correct		RT Incorrect		Dependent t-test			
	Mean	SD	Mean	SD	t	df	r	sig. (2-tailed)
Low Curvature	8.61	2.02	11.65	3.37	-4.45	6	0.88	0.00
High Curvature	6.49	1.59	9.73	2.68	-3.11	6	0.79	0.02
Haptic	12.97	2.51	15.04	3.08	-1.78	6	0.59	0.13
Visual	3.73	0.41	5.69	1.64				
Visual-Haptic	5.95	2.34	7.60	3.52	-2.53	6	0.72	0.04
Small Difference	9.02	1.78	9.53	1.41	-1.51	6	0.52	0.18
Medium Difference	7.55	1.35	10.08	2.97	-2.10	6	0.65	0.08
Large Difference	6.07	1.17	9.30	1.01	-2.76	2	0.89	0.11
All	7.55	1.36	9.78	1.72	-3.41	6	0.81	0.01

3.3.3.2 Confidence and Response Time

Repeated-measures t-tests were used to examine the differences between high confidence and low confidence response times for different levels of curvature, modality, and magnitude difference. These revealed that response times for high confidence were significantly faster than low confidence for all variables tested (see Figure 3-16 and Table 3-8 for mean response times and test statistics).

A series of scatter plots were used to compare response time against accuracy for each curvature, modality and magnitude difference. These did not appear to indicate any correlation between accuracy and response time. Correlation analysis (Pearson's r and

Spearman's rho[8]) confirmed that there was no significant correlation between response time and confidence for most of the variables tested, p>0.05. The exception was 'Visual' which had a negative correlation between response time and confidence, r=-0.86, p=0.01.

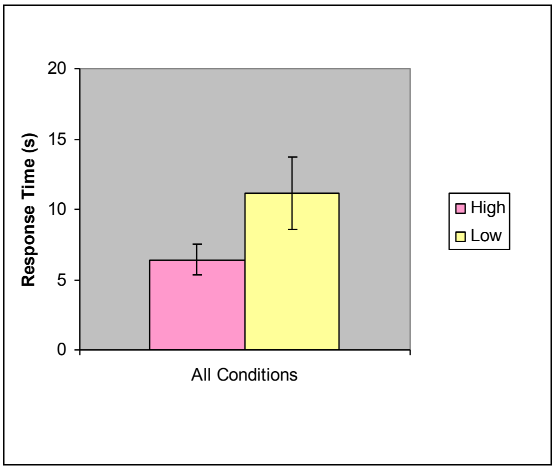


Figure 3-16: High confidence and low confidence response times for 'All Conditions'

Table 3-8: High confidence and low confidence response times, and the significance of the difference between them

Variables	RT Correct		RT Incorrect		Repeated Measures t-test			
	Mean	SD	Mean	SD	t	df	r	sig. (2-tailed)
Low Curvature	7.19	1.93	12.00	3.50	-5.94	6.00	0.92	0.00
High Curvature	5.65	0.77	10.30	2.82	-5.09	6.00	0.90	0.00
Haptic	10.79	1.99	15.92	2.72	-8.19	6.00	0.96	0.00
Visual	3.29	0.72	6.55	1.67	-4.06	6.00	0.86	0.01
Visual-Haptic	5.18	2.22	8.89	3.86	-5.89	6.00	0.92	0.00
Small Difference	7.28	1.60	10.55	1.90	-8.52	6.00	0.96	0.00
Medium Difference	6.54	1.05	10.15	1.85	-7.10	6.00	0.95	0.00
Large Difference	5.43	0.90	12.82	4.86	-4.46	6.00	0.88	0.00
All	6.42	1.16	11.18	2.78	-6.12	6.00	0.93	0.00

3.3.3.3 Accuracy and Confidence

Dependent t-tests were used to examine the differences between response times for high confidence correct (HCC) and low confidence correct (LCC), high confidence incorrect (HCI), low confidence incorrect (LCI) for each curvature, modality, and

[8] Spearman's rho was used for non-normally distributed variables; high curvature, and visual-haptic.

magnitude difference. These revealed that response times for HCC were significantly faster than LCC for all variables tested (see Figure 3-17 and Table 3-9 for mean response times and test statistics), and was also faster than most LCI variables tested. The exception here was ‘Large Difference’ where there was no significant difference in response time (see Table 3-10 for mean response times and test statistics). However, HCC was only significantly faster than HCI for ‘Visual’, for all other variables there was no significant difference (see Table 3-11 for mean response times and test statistics).

A series of scatter plots were used to compare confidence against accuracy for each curvature, modality, magnitude difference, and then all conditions combined. Correlation analysis (Pearson’s r) revealed that there were significant correlations between confidence and accuracy for ‘Haptic’ and ‘Visual-Haptic’, r=0.72, p=0.045, and r=0.77, p=0.03, respectively.

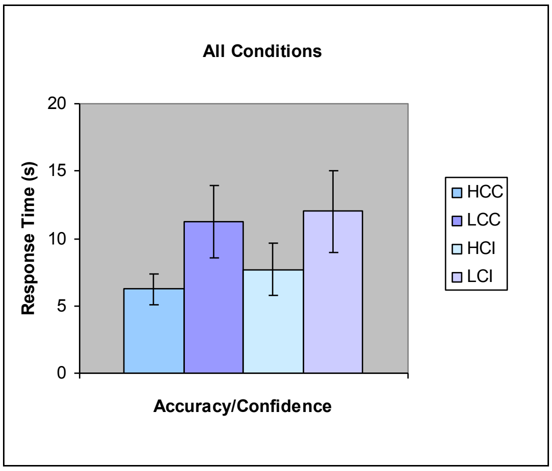


Figure 3-17: High confidence correct (HCC), low confidence correct (LCC), high confidence incorrect (HCI) and low confidence incorrect (LCI) response times for ‘All Conditions’

Table 3-9: Comparison of high confidence correct (HCC) and low confidence correct (LCC) response times

Variables	RT HCC		RT LCC		Dependent t-test			
	Mean	SD	Mean	SD	t	df	r	sig. (2-tailed)
Low Curvature	7.05	1.92	12.00	3.66	-5.61	6	0.92	0.00
High Curvature	5.41	1.15	10.31	3.12	-5.93	6	0.92	0.00
Haptic	10.75	2.11	16.08	3.05	-8.12	6	0.96	0.00
Visual	2.77	0.40	6.42	1.57	-5.56	6	0.92	0.00
Visual-Haptic	5.17	2.25	8.93	4.01	-5.42	6	0.91	0.00
Small Difference	7.23	1.71	10.65	2.19	-7.97	6	0.96	0.00
Medium	6.08	1.08	10.16	2.05	-8.26	6	0.96	0.00
Large Difference	5.38	0.94	12.89	4.85	-4.57	6	0.88	0.00
All	6.23	1.22	11.24	2.94	-6.22	6	0.93	0.00

Table 3-10: Comparison of high confidence correct (HCC) and high confidence incorrect (HCI) response times

Variables	RT HCC		RT HCI		Dependent t-test			
	Mean	SD	Mean	SD	t	df	r	sig. (2-tailed)
Low Curvature	7.05	1.92	8.54	2.95	-1.70	6	0.57	0.14
High Curvature	5.41	1.15	6.47	2.55	-1.29	6	0.47	0.24
Haptic	10.75	2.11	12.65	4.96	-1.15	6	0.43	0.29
Visual	2.77	0.40	3.91	0.44	-6.96	6	0.94	0.00
Visual-Haptic	5.17	2.25	6.04	3.05	-2.44	6	0.71	0.05
Small Difference	7.23	1.71	7.88	2.29	-1.06	6	0.40	0.33
Medium	6.08	1.08	7.59	4.16	-0.85	5	0.35	0.44
Large Difference	5.38	0.94	8.08	2.53	-1.94	2	0.81	0.19
All	6.23	1.22	7.70	2.08	-2.15	6	0.66	0.08

Table 3-11: Comparison of high confidence correct (HCC) and low confidence incorrect (LCI) response times

Variables	RT HCC		RT LCI		Dependent t-test			
	Mean	SD	Mean	SD	t	df	r	sig. (2-tailed)
Low Curvature	7.05	1.92	12.73	3.87	-5.61	6	0.92	0.00
High Curvature	5.41	1.15	10.72	2.39	-5.82	6	0.92	0.00
Haptic	10.75	2.11	16.06	2.72	-5.17	6	0.90	0.00
Visual	2.77	0.40	6.60	1.44	-6.00	6	0.93	0.00
Visual-Haptic	5.17	2.25	9.17	3.83	-5.89	6	0.92	0.00
Small Difference	7.23	1.71	10.42	1.70	-7.62	6	0.95	0.00
Medium	6.08	1.08	13.19	5.22	-3.88	6	0.85	0.01
Large Difference	5.38	0.94	11.62	2.35	-3.43	2	0.92	0.08
All	6.23	1.22	11.99	3.28	-5.38	6	0.91	0.00

3.3.4 Interaction

A repeated-measures ANOVA compared the effect on the number of comparisons of curvature, modality, and magnitude difference (see Table 3-12 and Figure 3-18). This revealed that there was a main effect of curvature, $F(1,7)=17.07$, $p<0.01$, partial $\eta^2=0.71$. Pre-planned contrasts showed that there were significantly less comparisons for 'High Curvature' than for 'Low Curvature', $F(1,7)=17.07$, $p<0.01$, partial $\eta^2=0.71$. There was a main effect of modality, $F(2,14)=62.00$, $p<0.001$, partial $\eta^2=0.90$. Pre-planned contrasts showed that ‘Visual-Haptic’ had significantly fewer comparisons than ‘Haptic’ and ‘Visual’, $F(1,7)=59.92$, $p<0.01$, partial $\eta^2=0.88$, and $F(1,7)=149.65$, $p<0.01$, partial $\eta^2=0.95$, respectively. There was a main effect of magnitude difference, $F(1.10,7.70)=69.25$, $p<0.01$, partial $\eta^2=0.91$. Pre-planned contrasts showed that ‘Small’ had significantly more comparisons than either ‘Medium’ or ‘Large’ magnitude differences, $F(1,7)=56.58$, $p<0.01$, partial $\eta^2=0.89$, and $F(1,7)=73.26$, $p<0.01$, partial $\eta^2=0.91$, respectively.

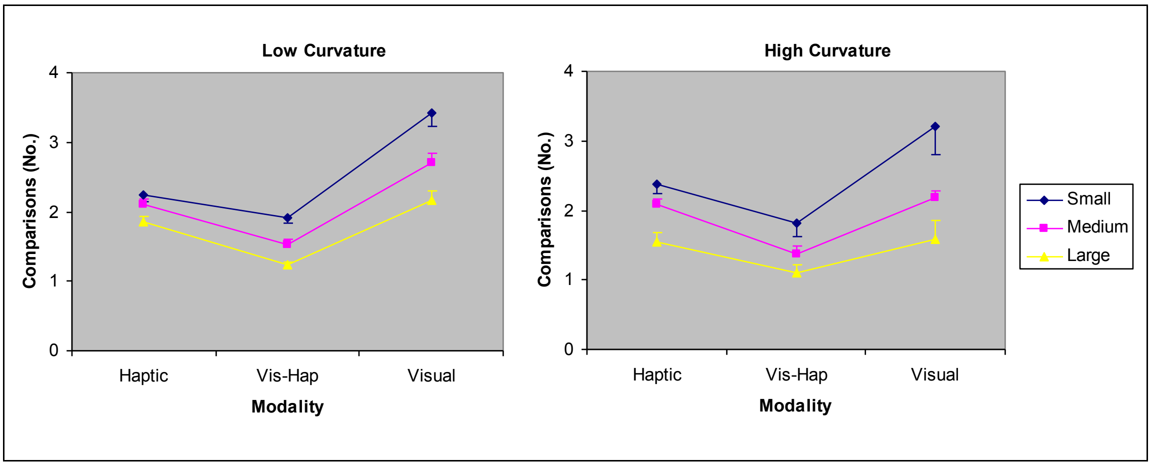


Figure 3-18: Effect of curvature, modality, and magnitude difference on confidence

Table 3-12: Comparisons (count) for main effect variables (curvature, modality, and magnitude difference)

Variable	Mean	SD
Low Curvature	2.13	0.27
High Curvature	1.92	0.27
Haptic	2.03	0.23
Visual	2.55	0.37
Visual-Haptic	1.49	0.30
Small Difference	2.49	0.37
Medium Difference	2.00	0.27
Large Difference	1.58	0.21
All	2.02	0.26

There were no significant interactions between curvature, and either modality or magnitude difference, $F(2,14)=2.40$, $p=0.13$, partial $\eta^2=0.25$, and $F(2,14)=3.64$, $p=0.053$, partial $\eta^2=0.34$, respectively. There was a significant interaction between modality and magnitude difference, $F(4,28)=20.69$, $p<0.01$, partial $\eta^2=0.75$ (see Figure 3-19). Pre-planned contrasts revealed that the decrease in comparisons for “Visual-Haptic” (compared to “Visual”) was less for either “Medium” or “Large” differences (compared to “Small”), $F(1,7)=15.53$, $p=0.01$, partial $\eta^2=0.69$, and $F(1,7)=32.40$, $p<0.01$, partial

η^2=0.82, respectively. Finally, there was no significant three-way interaction effect on confidence between curvature, modality, and magnitude difference, $F(1.54,10.79)=1.18$, $p=0.34$, partial η^2=0.14.

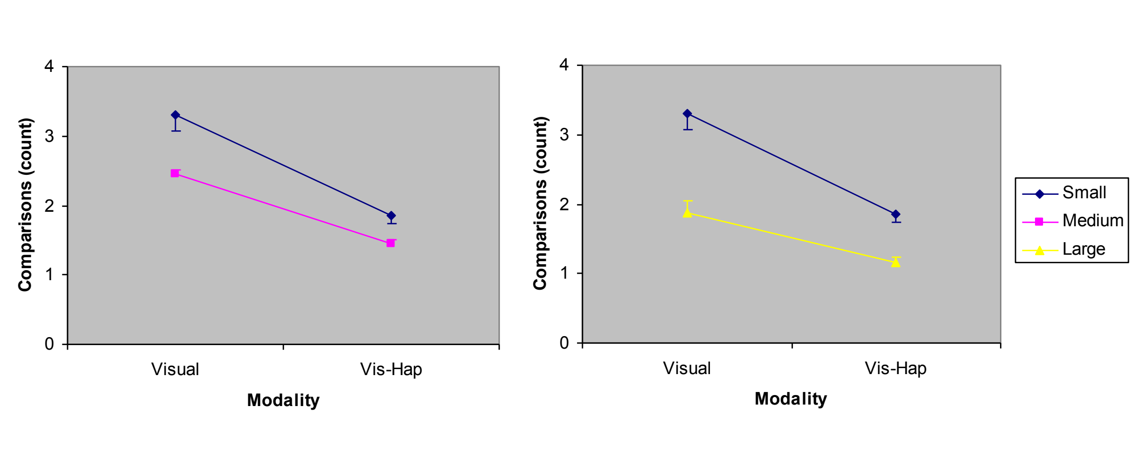


Figure 3-19: Interaction effect of modality and magnitude difference on comparisons

A series of scatter plots were used to compare curvature difference against comparisons for each modality in high and low curvature conditions (see Figure 3-20). These indicated a correlation between the two variables. Correlation analysis (Pearson's r) confirmed that there was a negative correlation between curvature difference and comparisons, so that as percentage curvature difference increased the number of comparisons declined. This was found for haptic (r=-0.96, r=-0.97), visual (r=-0.94, r=-0.96), and visual-haptic (r=-0.98, r=-0.99) modalities in low and high curvature conditions, p<0.01 for all. A further scatter plot compared comparisons for modalities combined against percentage curvature difference for low and high curvatures (see Figure 3-21). This also indicated that there was a correlation between the two variables. Correlation analysis (Pearson's r) confirmed that there was a negative correlation between percentage curvature difference and comparisons in low and high curvature conditions, r=-0.98, p<0.01 for both.

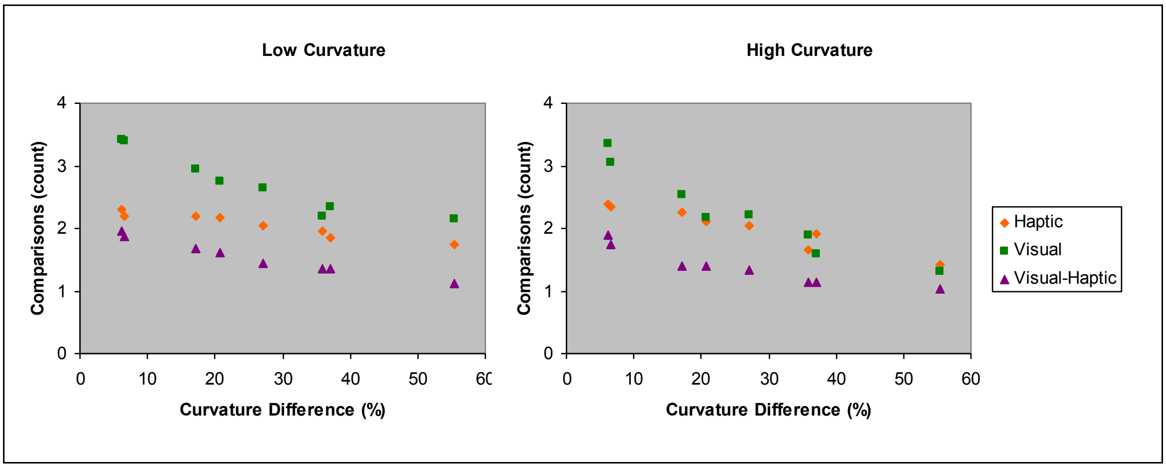


Figure 3-20: Correlation between curvature difference and comparisons for haptic, visual, and visual-haptic modalities

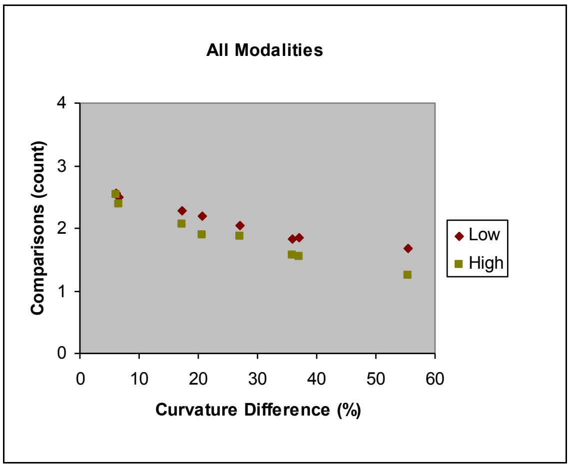


Figure 3-21: Correlation between curvature difference and response time for all modalities with low and high curvatures

3.4 Discussion (Study 1)

3.4.1 Perception

The experiment sought to address the following questions in relation to perception:

- What is the level of perceptual acuity for haptic, visual, and combined modalities in recognising changes in low and high curvatures? Is any one modality better than the other, and is there an effect on acuity by combining modalities?
- Determine whether measured difference thresholds conform to Weber's Law and whether this holds true for all modalities.

The results observed for low and high curvatures were very close. This similarity was found in haptic (22.43%, 23.10%), visual (22.79%, 23.10%), and visual-haptic (22.46%, 23.49%) modalities. However, analysis of variance showed that these values were significantly different. We can conclude from this that differences at these curvature levels did not conform to Weber's Law, that is they were not a constant proportion of the stimuli. The findings here concur with those within the haptic literature (Kappers and Koenderink 1996) that the perception of differences in curvature does not conform to Weber's Law.

For the haptic modality the mean Weber fraction across curvatures was 22.77%. This result is far more acute than those reported by Henriques and Soechting (2003) at 60% and Kappers and Koenderink (1996) at 41%. The Weber fractions here are closer to the 18% reported by Goodwin and Wheat (1992) for far smaller radii. However, a direct comparison with the previous findings within a similar range may be difficult. The 41% reported by Kappers and Koenderink (1996) is at an accuracy level of 84%, whereas this study used a level of 75% to calculate thresholds. The 60% given by Henriques and Soechting (2003) was for a combined mean for curvature differences over a number of orientations in a virtual environment, whereas the results reported here are solely for a horizontal plane in a real environment. The variety amongst Weber fractions not only highlights the problem of differing experimental conditions but also non-conformity to Weber's Law. So that for each curvature tested it is likely that a different JND will result, and the level of this may not be predicted from the stimulus provided.

The level of acuity in perceiving curvature difference was found to be the same for haptic, visual, and visual-haptic modalities. This means that the same differentiation can be achieved regardless of whether the judgement is made by touching or viewing the curved shape. This finding is counter to that previously reported in the literature. Ittyerah and Marks (Ittyerah and Marks 2008) observed a difference in performance across modalities, with vision being the best, then visual-haptic, and finally haptic. However, this discrepancy may be due to differences in the experimental conditions. The concern within this study was to mimic an applied setting; therefore the stimulus was positioned at table-level giving the participant an approximate viewing angle of about 60^{o}. Ittyerah and Marks (2008) used an angle of 90^{o} which would provide optimum viewing conditions for the visual modality. They also used two stimuli, one presented visually and the other touched. For this study the same stimulus block served both modalities within the visual-haptic condition. This presents the possibility that visual performance was hindered by occlusion from the touching hand. This study may therefore not have presented conditions that were optimum for the visual-haptic modality (or even visual modality), but did perhaps reflect the acuity levels attained in a more realistic setting, where sub-optimum conditions often apply. To this extent the modalities can be treated as providing equal sensitivity to curvature change for practical purposes.

The finding that curvature differentiation does not conform to Weber's Law has implications for the SATIN development and similar interfaces. Without a standard factor it may be difficult to scale differences across the whole curvature range. However, whilst it was found that threshold was significantly different for high and low curvatures, within an applied context it may still be possible to adopt more pragmatic solutions. The difference between high and low curvatures was small and so the mean (23%) of these could be used as a guide figure. Given that this is above the reported percentage change for even the smallest radii (Goodwin, John et al. 1991), there should be confidence that an acuity level of 23% is appropriate to a range of curvatures from 144/m (radius 3.55mm) to 1.46/m (radius 68.53cm).

A threshold of 23% (all modalities) or even the acutest threshold of 10% (haptic modality) (Goodwin, John et al. 1991) would not be sufficient to provide an accurate assessment of discontinuities for Class-A surfaces; these need to vary in curvature by less than 0.01/m. From the results presented here and those reported within the literature it is clear that haptic, visual, and visual-haptic modalities do not have sufficient sensitivity to judge discontinuities of curvature without some form of augmentation, for example the use of sound.

3.4.2 Performance

The experiment sought to address the following questions in relation to performance:

- Is performance (response time, accuracy, confidence) dependent upon the level of curvature, modality used, or extent of magnitude difference judged? Is any one modality better than the other, and is there an effect on performance of combining modalities?
- For each modality, would it be possible to predict the level of performance at the point of JND for any curvature?

Judgements made for "High" curvature stimuli were found to have quicker response times (6.2s), more accuracy (90%), and higher levels of participant confidence (80%) than when judgments for "Low" curvature stimuli were made (8.6s, 84%, 66%). This fits with anecdotal evidence that most participants found it easier to judge differences for “High” curvature stimuli as opposed to “Low”. However, given that the “Low” curvature stimuli were four times as long as the “High” curvature stimuli, the response time increases are only a factor of 1.39 different. To look at this another way, when the response times are normalised for length, then the “Low” curvature responses are quicker than the “High”. This may suggest that assessment of difference is not carried out in a linear fashion, i.e. over the whole length of the stimuli, but is made in some other way; perhaps with a concentration of inspection at the ends of the stimuli to detect slope differences.

For ‘Small’ magnitude differences it was found that these had slower response times (9.2s), were less accurate (70%), and had lower levels of participant confidence (51%) than when judging either “Medium” (7.2s, 94%, 77%) or “Large” (5.7s, 99%, 92%) magnitude differences. This means that as the difference between the stimuli become greater the quicker responses become, they increase in accuracy, and the participant feels more confident in their judgement. This is largely as expected as “Small” differences are below the level of the JND and so one would expect accuracy and confidence to be low, where as “Medium” and “Large” are above the JND. However, whilst accuracy is very good for both “Medium” and “Large” differences the level of confidence is only of a comparable level for “Large” differences. This means that only when the JND is well exceeded are belief (confidence) and actuality (accuracy) well aligned. So for judgements near to the JND it may be beneficial to provide additional feedback to supplement that available through the haptic and visual modalities, and so boost confidence to a more realistic level. It is also worth noting that for accuracy there is an interaction between magnitude difference and curvature, such that the increase experienced for “Medium” and “Large” differences (as compared to “Small”) is greater for “High” curvature stimuli (as compared to “Low”). This fits with the finding of the perceptual results that acuity (which is based on accuracy and difference from stimuli) is different for high and low curvatures.

In terms of response time and accuracy there is no significant difference between “Visual-haptic” (5.7s, 90%) and “Visual” (3.6s, 88%) modalities. However there was a significant difference between ‘Visual-Haptic’ and ‘Haptic’ (12.9s, 85%) modalities. These results indicate that in combining modalities there is no detrimental effect on performance. The fact that the multimodal performance is better than the haptic performance, but not better than the visual performance, may suggest that the multimodal performance is predominantly visual. This fits with suggestions in the literature that vision dominates (Ernst and Banks 2002). In relation to confidence, visual-haptic performance is significantly better than both visual and haptic. This suggests that in combining modalities there is an intangible that boosts confidence.

For all the performance measures there is a strong correlation between performance and magnitude difference. This should make it possible to predict performance at various degrees of difficulty; that is from the point of JND to magnitude differences where a 100% performance should be possible. However, performance levels differ depending on the degree of curvature, and so for high curvature the level of performance is slightly raised. This meant that whilst predictions could be made for each modality, the accuracy of these would vary dependent upon the level of curvature. The alternative would be to base predictions on modality for a given curvature. The results here could be used to set two predictive ranges, one for 'High' and one for 'Low' curvature situations.

3.4.3 Performance Characteristics

The experiment sought to address the following question in relation to performance:

- Are there characteristic traits of performance i.e. are response times for some types of judgement quicker than others?

A number of performance characteristics were investigated. Correct answers were faster than incorrect for "High" and "Low" curvature, "Visual", Visual-Haptic" and for all conditions overall. It is interesting to note that there was no difference in response times for "Haptic", although there is for "Visual-Haptic" (and "Visual"). This may be a further indication that the combined modality is dominated by the visual channel. High confidence answers were faster than low confidence in all cases.

These findings are reflected in HCC being faster than both LCI/LCC, but not HCI. This suggests that it was high confidence and not correct answers that produced faster responses. Even so, there was no correlation between overall confidence (%) and response time. Therefore whilst the mean response time was greater for high rather than low confidence this did not translate into a correlation between overall confidence (%) and response time. This lack of significant correlation was likely due to the overall high level of performance and concomitant quick responses. It would therefore seem important not to infer too much meaning to individual slow or fast response times in terms of overall confidence or accuracy[9]. There was however a correlation between confidence and accuracy. It would seem therefore that if someone has a high level of confidence in their performance this is probably a good reflection of their actual performance i.e. they achieved a high level of accuracy.

3.4.4 Interaction

The experiment sought to address the following question in relation to interaction

- Is the amount of interaction (number of comparisons) affected by different curvatures, modalities, or extent of magnitude difference judged? Is there a relationship between the level of interaction and other performance measures?

The number of comparisons made for “High” curvature stimuli (1.9) was significantly less than for “Low” (2.1). This fits with other performance data, where “High” curvature has better response times, accuracy, and levels of confidence. There was found to be a negative correlation between the number of comparisons for “Low” curvature and response time. This seems counter intuitive as it means that as the number of comparison decrease the response time increases. This may reflect that at low curvatures some participants seemed to scan the stimulus very slowly in the haptic condition and therefore made fewer comparisons, but took longer to do them. This was also noticeable behaviour in the visual condition. However, this is rather anecdotal and further study would be needed in order to understand this finding.

For “Small” (2.5) magnitude differences there are significantly more comparisons made between stimuli than for either “Medium” (2) or “Large” (1.6) magnitudes. This suggests that below threshold judgements require more feedback than those that are above the JND. Even so, as seen by the performance results, this does not make them as accurate as those above the JND. It would seem that more interaction does not necessarily equate to greater accuracy, or even confidence. This lack of relationship is supported because there was found to be little correlation between the number of comparisons and any of the performance measures for any magnitude difference. Where there was a

[9] It is suspected that as performance declines slower response times may increase in frequency and this would likely lead to a correlation becoming significant.

correlation (between comparisons and accuracy for “Large” magnitude differences) this indicated that more did not mean better, as it was observed that more comparisons gave less accuracy. This finding may reflect that, at large differences where judgement should be easy, over-analysing led to poorer decisions.

There were significantly fewer comparisons made for “Visual-Haptic” (1.5), than for either “Haptic” (2) or “Visual” (2.5). This indicates that when modalities are combined there is a reduction in the amount of haptic interaction that takes place. Unfortunately it was not possible to know whether visual comparisons remained constant, decreased, or increased. Further work would be needed in order to clarify the situation of visual interaction in a multimodal context. However, the fact that the haptic interaction reduced is interesting. This may be a further indication that the dominant mode is visual; some participants reported that they had only touched the block because they had to, although this was mainly for larger magnitude differences. Even if it is the case that the visual dominates, there is evidence to suggest that the combination of both haptic and visual is important as it seems to increase confidence.

3.5 Summary of Chapter

The study detailed in this chapter set out to understand the effects of combining modalities on perception and performance. It also sought to assess the possibility of predicting perception and performance. The results showed that there was no significant difference between modalities in perceiving curvature differences. It also found that there was no conformance to Weber’s Law for any of the modalities. This means that JNDs were different for high and low curvatures which may make prediction based on curvature problematic. However, it was also found that Weber Fractions across modalities and curvatures were very similar. This meant that a change of approximately 23% could be used to standardise feedback. Unfortunately this is insufficient in order to detect discontinuities in class ‘A’ surfaces, and would mean that augmentation of some form would be required. The relatively large Weber Fraction of 23% would also allow for a wide tolerance in the accuracy of visual-haptic interfaces. That is, they would not need to replicate the curvature of an object exactly.

Performance was found to be quicker, more accurate, and more confident for high curvatures compared to low. This means that it was easier to detect discontinuities for high curvature stimuli compared to low. Similarly, medium or large magnitude differences were found to be quicker, more accurate, and more confident than small magnitude differences. There was found to be no detrimental effect on performance when combining modalities. Visual-haptic performance was equal to visual performance, and was significantly better than haptic performance (for response time and accuracy). An advantage of combining modalities was a significant boost in confidence over unimodal performance.

3.6 Related Chapters

This study indicated that it may be difficult to predict JNDs across curvatures. Further consideration of this issue, and a discussion of an alternative means of prediction, is outlined in Chapter 4. It was found that the JNDs for visual-haptic interaction were too large for the detection of discontinuities in Class A surfaces. It was therefore suggested in this chapter that some means of augmentation was used. It was thought (through work with the SATIN prototype) that sound would potentially provide a good means of conveying curve shape and curvature information. The appropriateness of this medium was investigated in Study 2 and is reported in Chapter 5, and Chapter 6. The effect of augmenting visual-haptic interaction with sound was examined in Study 3 and is reported in Chapter 7. The findings here also contribute to the general discussion (see Chapter 10).

Chapter 4: Prediction JNDs and Associated Level of Performance

4.1 Introduction

This chapter builds on the work undertaken in Chapter 3. It starts by setting the scene as to why we may be interested in developing a predictive model for threshold and its associated level of performance. Having done this there is discussion as to why the results from Study 1 (see Chapter 3) indicate curvature's lack of suitability as the predictor variable. This leads to a re-examination of the literature, and the identification of an alternative theoretical approach that could underpin a predictive model. The data generated in Study 1 were then re-analysed in the light of this theory.

4.2 The Value of a Predictive Model

4.2.1 Ensuring Perceptible Feedback

The problem for any interface is to ensure that there is sufficient feedback to the user in order that they may perform their tasks effectively (Norman 1988). In relation to product design interfaces this means ensuring that the user has sufficient information about object properties that they can make proper assessment of these (i.e. ensure that they have G3 continuity) and make the necessary modifications. For the detection of curvature discontinuity, this means that they should be able to sense differences in curvature. Where human perceptual ability is insufficient for this task then augmentation should be considered.

To facilitate this it would be necessary to know for any given curvature the degree of difference required in order that a person could perceive a change. In other words, it is necessary to predict the JND. By predicting this it is possible to know if assessment of a given curvature is within human capabilities, or if not, that additional feedback could be given through a redundant modality such as sound. It also allows for assistance in modifications that cannot be made because the adjustments are too fine to be recognised by the user.

In addition to assessment and modification, prediction of JND would enable interfaces to be developed to an appropriate level of fidelity. For example, because of the physical limitations of haptic devices it is not always possible to reproduce the curves that exist within a CAD model exactly. However, a close approximation of the shape and curvature can be reproduced. The challenge then is to ensure that this approximation is near enough. By predicting the JND it would be possible to know if an interface's fidelity was at an appropriate level or if further refinement was needed in order to match that which could be resolved by the user.

4.2.2 Objective Evaluation Criteria

One of the difficulties encountered throughout the SATIN evaluations was that of understanding the level of performance achieved (see Chapter 1 section 1.1.5). In relative terms it was possible to compare performance between prototypes, for different sound configurations, or changes in hardware. However, it was not possible to get an absolute sense of whether the interface had performed to an optimum level. The difficulty in particular was assessing whether, given the level of task difficulty, efficiency (response times) and effectiveness (accuracy) met or exceeded what might be expected. To judge this it would have been necessary to be able to first quantify task difficulty, and then know the relative level of performance associated with that. When comparing curvatures, a determinant of task difficulty is the ease with which differences in curvatures can be perceived. For example, differences that are at about the level of JND are much harder to perceive than those that are well above this level. Consequently tasks that are at threshold level will naturally have poorer associated performance than those that are well above this level. Given that there is a strong correlation between performance measures and magnitude difference (see Chapter 3 section 3.3.2), then once JND is known it may be possible for performance to be calculated. Therefore, it can be argued that the key to an objective measure is the prediction of the JND for a given curvature.

The importance of knowing the level of optimum performance is not only relevant to individual evaluations, but also provides a means of comparing different interfaces. At present it would be difficult to assess the fidelity, and performance of different interfaces as there is a lack of information within the literature regarding what might be considered optimum performance within this task domain. By providing a means of predicting the acuity and performance of an optimum interface, it is possible to evaluate the range of interfaces available against this. In this way, the true validity of novel interfaces, such as that developed within SATIN, can be known.

4.2.3 Provision of a Cost-effective Method

At present, in order to know what the difference threshold is for any given curvature it is necessary to undertake a psychophysical experiment. This type of experiment requires a high time commitment as there is often the need to run up to about 100 trials per participant per condition. The perception and performance study (see Chapter 3) used 8 participants each taking 3 hours to complete the necessary trials. Therefore in total 24 hours plus analysis time were required to determine the visual, haptic, and visual-haptic difference thresholds for one high and one low curvature. This level of resource use is not particularly economical, and may be a barrier to a more quantified understanding of multimodal interfaces within the domain of interest. By being able to predict the acuity and performance at a given curvature, devices may be evaluated more easily as there would be no need to pre-test the curvature of interest.

There are also other benefits associated with knowing the threshold for a particular modality. With the SATIN project, one of the key aims was to evaluate the contribution made by augmenting haptic and visual feedback with sound. However, in order to understand whether this augmentation was beneficial, it was necessary to first establish the level of performance without sound. By being able to predict the threshold, and the associated level of performance, the need for generating comparative data is circumvented and so evaluation can be completed more quickly, and more cost-effectively.

4.3 The Limitations of using Curvature to Predict Threshold and Performance

It was found in Study 1 that the two curvatures explored had different Weber Fractions, and so did not conform to Weber's Law. This meant that as curvature changed the proportion of the stimulus needed for a difference to be detected also changed; that is it was not a constant proportion of the stimulus curvature. Therefore it is not possible to predict what the threshold for any given curvature will be except in very general terms. In addition to this, it was found that performance is also affected differently dependent upon the level of curvature within the stimulus; performance is better for high curvature than for low curvature conditions. In practical terms this means that whilst there is a correlation between performance and curvature difference (from the standard stimulus) which may enable prediction, because of the differential effect of curvature separate metrics would be needed for high and low curvature conditions.

4.4 Alternative Theoretical Approach derived from the Literature

When approaching Study 1, attention was focused on the problem of detecting differences in curvature. The concern was to understand how different modalities affected the perception of curvature and what this meant in terms of performance. However, the focus on curvature may have resulted in missing other aspects upon which prediction could have been based. As curvature was not found to be suitable as a predictor, the literature was re-examined in order to identify possible alternatives.

As explained in the literature review, the mechanism for 'curvature' detection is slope difference over the whole stimulus width (Pont, Kappers et al. 1999). This finding of Pont et al. (1999) builds on the understandings already present in the literature on absolute thresholds. Gordon and Morrison (1982) found that the change necessary to detect a curved surface from a flat one was about half a degree. A similar level of change was also identified by Pont et al (1997) although they talk about this in terms of 'attitude difference'. These findings were for isolated ranges of curvature, however Louw et al. (2000) investigated this over the 'whole range of spatial scales' and produced similar findings; detection threshold was found to be a power function of width with an exponent of 1.3. It is therefore well known that detection of curvature is related to a change in slope or angle of elevation (see Figure 4-1)

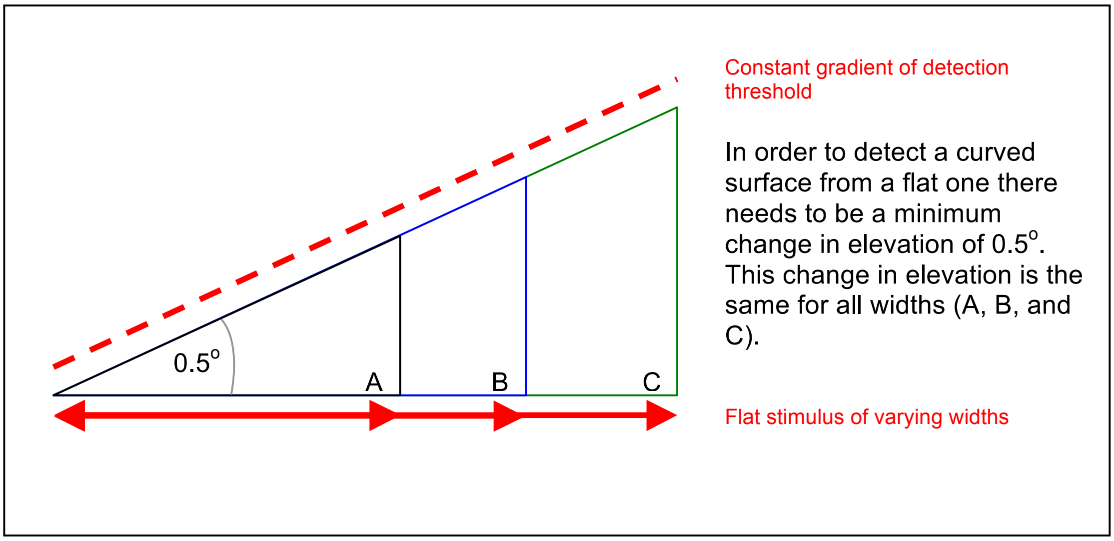


Figure 4-1: Constant Gradient of Detection (absolute) Threshold

It is thought that given the active mechanism of detection is slope then there is reason to assume that the same or similar mechanism operates for the differentiation of changes in curve shape. It therefore follows that detection of difference relies on sensing a noticeable increase in gradient above that already detected (see Figure 4-2). There is some support for this in the work of Louw et al. (2002). In looking at perception of shape differences Louw et al. (2002) found that those shapes with similar proportions (ratio of width to height) were more difficult to distinguish from each other; in other words there was no increase in gradient and so no difference could be sensed.

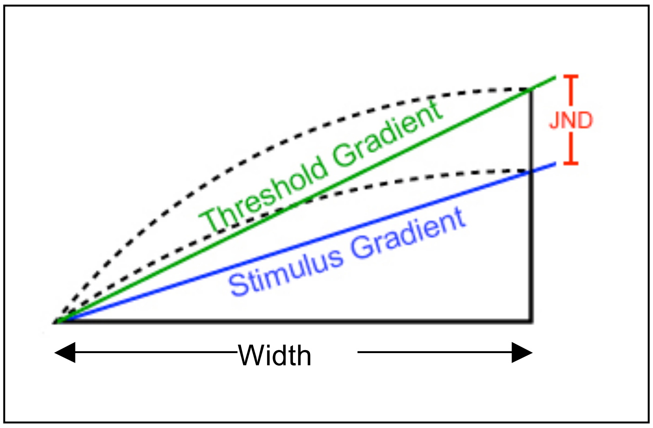


Figure 4-2: Conceptualising difference threshold as perception of gradient change

The curvature (dotted line in Figure 4-2) will vary relative to the threshold gradient. Therefore, by predicting the threshold gradient, the threshold curvature can also be known. Given that detection relies on slope rather than curvature per se, it would seem more appropriate to talk of difference perception in terms of gradient rather than curvature.

The stimulus gradient depends on both the width and height of the stimulus for its calculation, and so too does the threshold gradient. Because of this the finding of Louw (2000) in relation to detection - that threshold is proportional to stimulus width to the power 1.3 - may not be applicable to this situation. For example, the same width may be paired with differing heights resulting in stimuli with different gradients and therefore different thresholds. Louw's formula would give just one threshold based on the width regardless of height (and thereby gradient) and so is not appropriate for the prediction of difference thresholds (without modification). This can be evidenced by looking at results from Study 1. The widths of Stimulus 'A' (used in the pilot) and Stimulus 'B' (used in the main study) were the same at 20cm. If Louw's formula was applied, the predicted value of the threshold base-to-peak height would be the same for both stimuli as it is based on the same width. However the actual thresholds had different base-to-peak heights and so would indicate that this is not applicable to the prediction of difference thresholds. It should also be noted that much of the research undertaken has been in the context of haptic perception, so there is some need to ensure that this is also applicable to multimodal perception.

4.5 Further Analysis of Perception Data

Given the revised theoretical approach to detecting differences in gradient between two curved objects it was felt necessary to reanalyse the data from Study 1 to explore if this new view was supported.

4.5.1 Data Analysis

The data used in the analysis were that generated in Study 1 (see Chapter 3) which looked at difference thresholds for various modalities and curvatures. As gradient is related to curvature, all that was required for the re-analysis was to calculate the gradients for each of the curved blocks and their thresholds. The low curvature block (1.46/m) had a gradient of 0.073, whilst the high curvature block (15.97/m) had a gradient of 0.208 (see Figure 4-3). In addition to this data, supplementary haptic data were converted from the pilot study for Study 1 (stimulus curvature 0.45/m with gradient of 0.022) and the literature. These additional data were used in the regression analysis in order to make the results more robust, and for external validity (by comparison to data found within the literature).

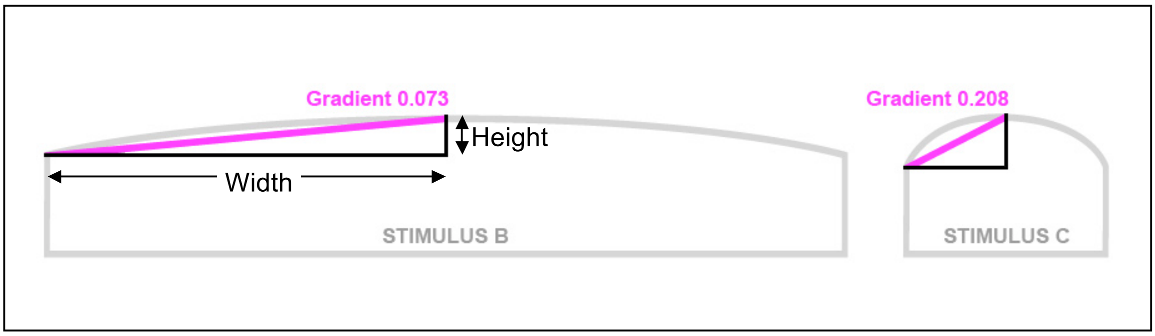


Figure 4-3: Calculation of Gradient for Stimulus Blocks

4.5.2 Results

4.5.2.1 Regression Analysis

For each modality, threshold gradient was plotted against stimulus gradient (see Figure 4-4, the trend line is for all modalities combined). A regression analysis was undertaken on the data in SPSS. This showed that stimulus gradient made a highly significant contribution to predicting threshold gradient for ‘Haptic’, ‘Visual’, and ‘Visual-Haptic’ modalities and for all modalities combined; R^2=1, adjusted R^2=1, and p<0.001 for all variables. The model coefficients for each modality and all modalities are shown in Table 4-1.

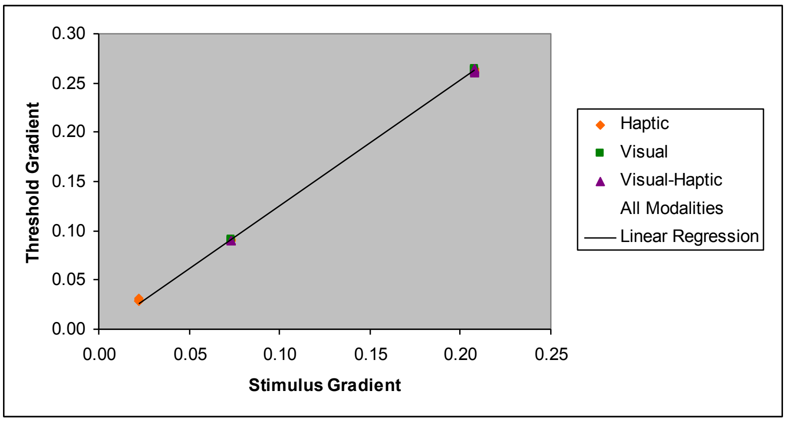


Figure 4-4: Threshold gradient as a function of stimulus gradient (all modalities)

Table 4-1: Model coefficients for each modality and all modalities combined

Model		Unstd. Coe. B	Unstd. Coe. SE	Std. Coe. Beta	t	Sig.	95% CI Lower	95% CI Upper
Haptic	Constant	1.46E-005	0.001		0.008	0.986	-0.001	0.001
	Stimulus Gradient	1.260	0.005	1.000	250.945	0.000	1.250	1.271
Visual	Constant	-0.003	0.000		-6.385	0.000	-0.004	-0.002
	Stimulus Gradient	1.279	0.003	1.000	426.763	0.000	1.272	1.285
Visual-Haptic	Constant	-0.004	0.001		-5.541	0.000	-0.005	-0.002
	Stimulus Gradient	1.285	0.004	1.000	300.467	0.000	1.276	1.295
All Modalities	Constant	-0.001	0.000		-3.219	0.002	-0.002	-0.001
	Stimulus Gradient	1.270	0.003	1.000	424.277	0.000	1.264	1.276

4.5.2.2 Regression Analysis Incorporating Haptic Data from the Literature

Haptic threshold data from Study 1 and the pilot experiment were used with difference threshold data from the literature (Gordon and Morison 1982; Goodwin, John et al. 1991; Goodwin and Wheat 1992) to give the regression plot shown in Figure 4-5. This indicated a strong correlation between stimulus gradient and threshold gradient. A regression analysis was undertaken on the data in SPSS. It was found that stimulus gradient made a highly significant contribution to predicting threshold gradient; R=0.998, R^2=0.996, adjusted R^2=0.995, and p=0.00. The model coefficients were as shown in Table 4-2.

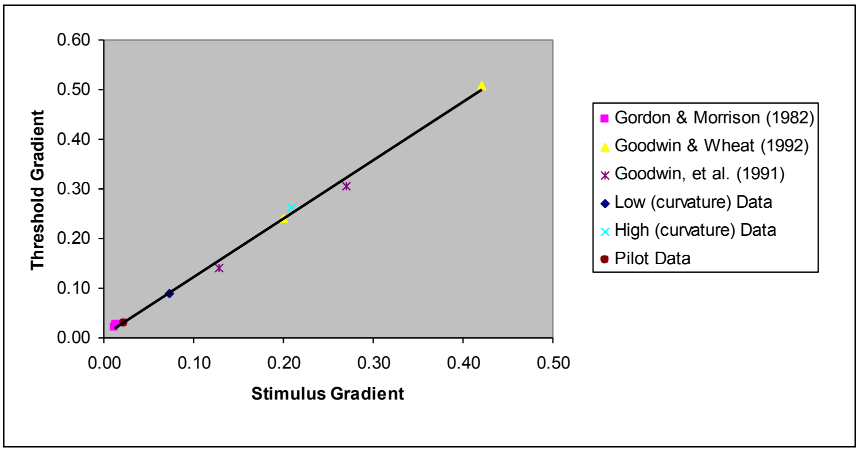


Figure 4-5: Threshold gradient as a function of stimulus gradient (haptic only)

Table 4-2: Model coefficients for experimental haptic data and data from the literature

Model		Unstd. Coe.		Std. Coe.	t	Sig.	95% CI	
		B	SE	Beta			Lower	Upper
Haptic	Constant	0.004	0.006		0.625	0.552	-0.010	0.017
	Stimulus Gradient	1.18	0.029	0.998	41.335	0.000	1.113	1.248

4.5.3 Discussion

The re-examination of the data found a strong relationship between stimulus gradient and threshold gradient[10]. In predictive terms, stimulus gradient was found to account for 100% of the variation in threshold gradient for each and all modalities. However caution should be exercised with this figure, as there was a limited range of data points from which the regression analysis was made; the data clustered around two curvature points for “Visual” and “Visual-Haptic” and therefore made a high level of linearity more likely. Even so, the same figure was achieved for “Haptic” which had data from three curvature points. Further to this the “Haptic” data was analysed in combination with a range of data from the literature, and an extremely high level of model prediction was maintained (99.6%). The adjusted R also indicated that in general terms the model would account for 99.5% of the variation in gradient threshold. Given this there is strong evidence that stimulus gradient is a good predictor of threshold gradient for the haptic modality, and that this is also likely to be true for visual and visual-haptic modalities.

[10] There is the possibility that such a high r-squared is due to the fact that parameters are equivalent. However, given that the parameters regressed are distinct physical entities it is unlikely that this is the case.

These results indicate that the perception of differences between two curved surfaces operates in a similar way to the detection of a curved from flat surface; that is through perceiving a change in gradient. It is therefore possible to propose models to describe this characteristic of difference perception for each modality as given in Figure 4-6.

Visual Haptic Modality:
Threshold Gradient = (Stimulus Gradient x 1.285) -0.004

Visual Modality:
Threshold Gradient = (Stimulus Gradient x 1.279) -0.003

Haptic Modality:
Threshold Gradient = (Stimulus Gradient x 1.260) -0.005

Figure 4-6: Models to describe Threshold Gradient in Visual-Haptic, Visual, and Haptic Modalities

Once threshold gradient has been calculated, the amount of change necessary to give a perceptible difference or JND is the threshold gradient less the stimulus gradient. In this way these models help in the prediction of JNDs.

4.6 Summary of Chapter

This chapter identified that it would be useful to develop a predictive model in order to gauge interface feedback, produce objective evaluation metrics, and as a cost-effective method of producing comparative data. However, because JNDs were found to be different for high and low curvature (in Study 1) this dimension was thought to be limited as a predictor. Through re-evaluation of the literature, gradient was identified as a potential geometric property that may prove to be more useful. Through further analysis of the Study 1 data it was found that stimulus gradient was a very strong predictor of threshold gradient. This was also found to be the case when haptic data from the literature was integrated into the regression analysis. This indicates that perception of differences between two curves operates in a similar fashion to the detection of a curved from flat surface; that is through perceiving a change in gradient.

4.7 Related Chapters

In order to test the theory that stimulus gradient is a predictor of threshold gradient, Study 3 used stimuli defined by their gradients (amongst other dimensions) and is described in Chapter 7. For each stimulus the JND was predicted from the model derived in this chapter prior to the commencement of the study. A comparison was then made between the predicted and actual results in order to assess the validity of the model. A discussion of this and the results of the analysis can be found in Chapter 8. The findings here also contribute to the general discussion (see Chapter 10).

Chapter 5: Perception of Curve Orientation and Magnitude from Sound

5.1 Introduction and Rationale

There are limitations in the ability of vision and touch to convey information about changes in object shape (see Chapter 3 section 3.4.1), and so some form of enhancement would be needed in order to convey this at the required acuity. The auditory modality was considered to be redundant in the assessment of shape and so presented an opportunity to use sound in order to provide additional information to convey changes in shape more effectively. The success of this would depend upon the ability of users to perceive and understand this type of information.

Prior to this study, a number of formative evaluations were carried out on the SATIN prototype (see Chapter 1 section 1.1.5). These revealed that whilst the use of sound was largely successful there was some evidence that the sounds might be confusing or that workload was excessive. It was therefore felt that a more rigorous investigation of the issues was required in order to validate the concept of using sound to convey curve shape and curvature data.

5.1.1 Study Structure

The study was broken down into a series of three experiments (see Figure 5-1). These were conducted with the same participants in two sessions about a week apart. Briefly, Experiment 1 evaluated how well participants were able to discriminate the orientation of a curve using sound. Experiment 2 evaluated how well participants were able to discriminate the magnitude of a curve using sound. Experiment 3 evaluated the appropriateness of using sound to convey curvature and curve shape; this was a more complex activity than the earlier two experiments and reflected the demands of an applied setting. This chapter presents the fundamental experiments assessing communication, whilst Chapter 6 will report upon the applied experiments.

It should be noted that a further aspect of Experiment 1 and 2 was that they sought to understand the effect of practice, so Session 2 is an exact repeat of Session 1. For the purposes of these experiments, Experiment 3 acts as an opportunity to practice listening to sound for curve information.

Finally, it can be seen that Experiment 3 has been split across the two sessions. This is because the duration of the experiment was too long to undertake in one session. Whilst the same in all respects, each half of the experiment evaluated different sonification methods.

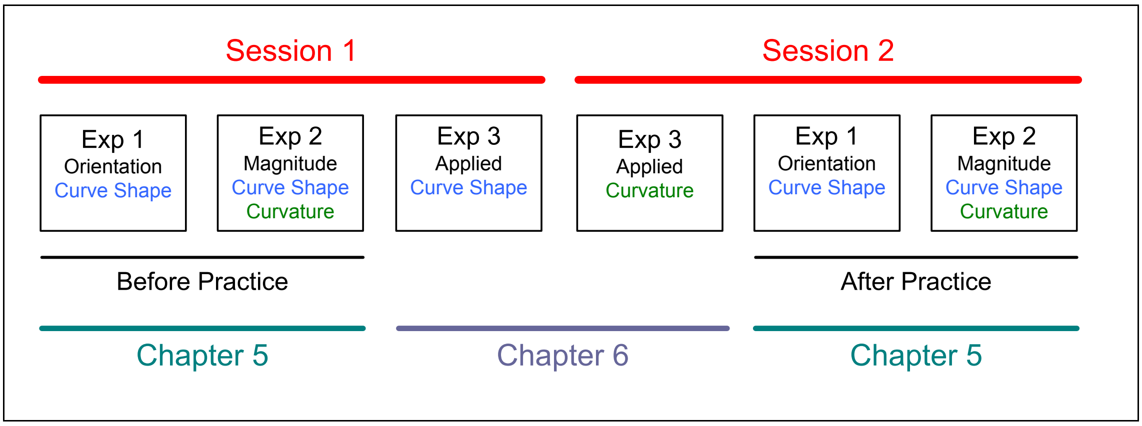


Figure 5-1: Structure of the Sound Study

5.1.2 Exploring shape with sound

In thinking about a shape one of the first things to note is that it occupies a given portion of space and that this space is bounded by the edges of the object. For example, if we view a cup as a two-dimensional outline, as in Figure 5-2, it can be seen that as the eye follows the outline of the curve shape it travels in numerous directions, and in so doing covers most directions in a range that approximates 360°. In order for sound to successfully convey curve shape it must be able to communicate these changes in orientation in a way that is understandable to users. Experiment 1 therefore examined the suitability of sound for conveying curve orientation. This was measured in terms of how well participants could identify the orientation of a circle segment.

The other thing to note is that a shape is principally made up of varying degrees of curved and flat lines. It is this combination of curves and lines that help us to recognise an object. Figure 5-3 illustrates the segments of different curves and lines that make up the outline of the cup. At an extreme, this is how we recognise the difference between a square and a circle; one is completely made of flat lines the other of curved lines. It is therefore important for sound to be able to convey these in such a way that we are able to get a sense of flatness and curvedness. Of importance in this context is sound's ability to convey differences in the curvedness of a shape. Experiment 2 therefore looked at how well sound conveyed different magnitudes of curve. For the purposes of that experiment this was explored by how well participants could identify differences in magnitude between two curves i.e. which was more curved.

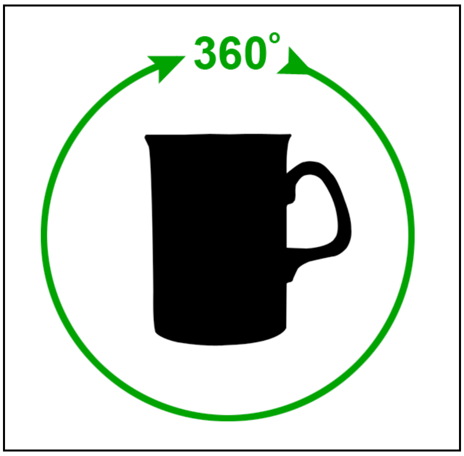


Figure 5-2: Sound needs to convey the orientation of curves through 360 degrees

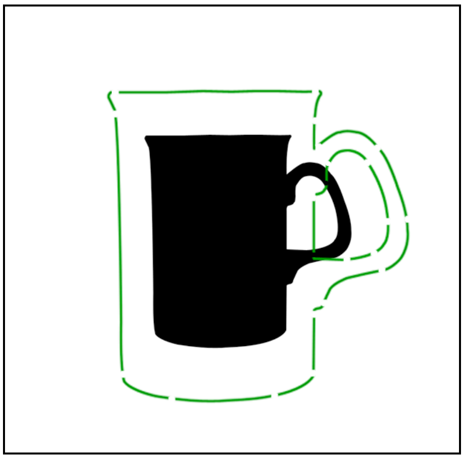
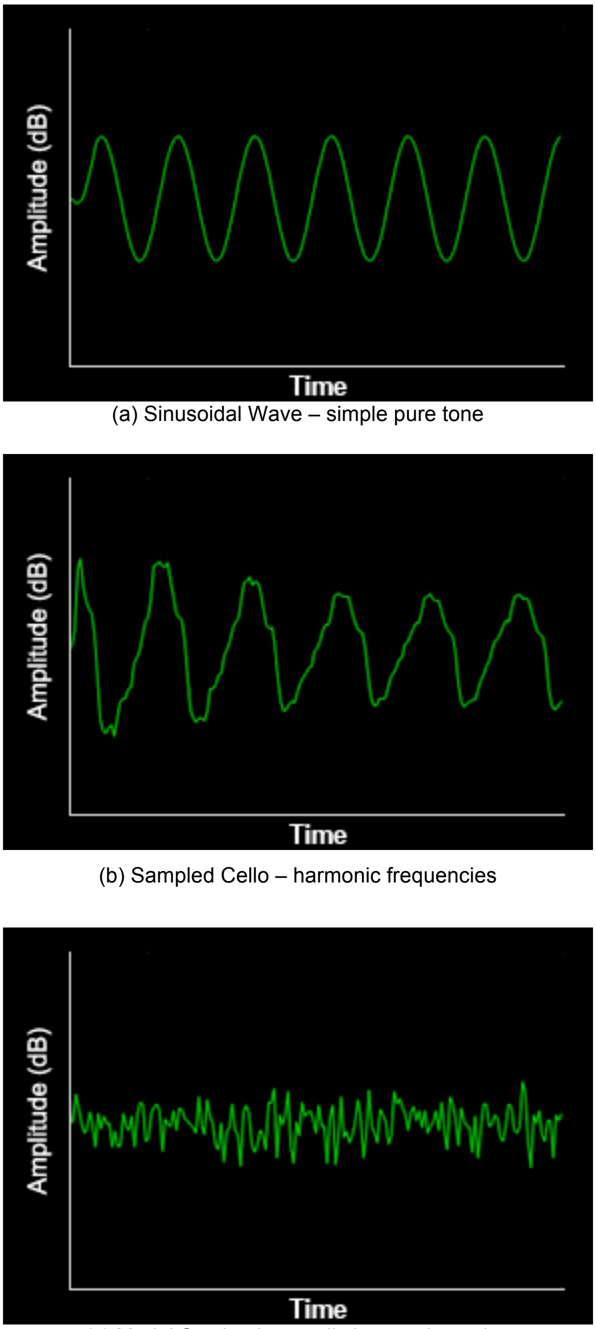


Figure 5-3: Sound needs to convey different magnitudes of curve shape

5.1.3 Types of Sound

During early prototyping of the SATIN interface a number of different types of sound were used. The initial sounds were derived from a workshop with end-users, and these were then implemented within the SATIN prototype. Evaluations found that participants experienced difficulties in using the sound interface. However, it was unclear from these evaluations as to whether the difficulties arose through the technical characteristics of the haptic strip used on the prototype, or perceptual issues with the types of sounds used. The first of these was a technical issue that was more appropriately investigated by the development team. The latter issue was of a more fundamental nature and would need more controlled investigation than could be achieved in situ. For this reason it was decided to pursue this through a fundamental study of the appropriateness of sound for conveying curve shape information.



(a) Sinusoidal Wave – simple pure tone

(b) Sampled Cello – harmonic frequencies

(c) Modal Synthesis – realistic complex noise

Figure 5-4: Types of Sound ranging from simple to complex

The sounds used to convey the curve shape information were of different complexities (see Figure 5-4). These were chosen as they were thought to offer different qualities; one was a simple clear tone which was thought to be easy to perceive (a), a second was a harmonic sound that was considered to be more pleasing than the sine wave (b), and a third was a naturalistic sound that was thought to add realism (c).

However, it was believed that some types of sound, because of their less complex nature, may be easier to perceive than others i.e. they were less 'noisy' and so were able to convey information more accurately. Conversely, it was thought that whilst more complex, realistic sounds may be more intuitive and so increase performance. One purpose of this study was therefore to establish if some sound types were better for conveying information than others; that is are more easily perceived?

5.1.4 Workload

During the course of Study 1 it was observed that participants exhibited different degrees of effort when undertaking the comparison task. This was particularly noticeable in the haptic only condition where participants seemed to have a strong concentration on the task they were undertaking. The increased difficulty of this condition over the visual and visual-haptic conditions was evidenced by slower response times, less accuracy, and lower confidence. There were similar observations made during the SATIN prototype evaluations undertaken in January and September 2008 (see Chapter 1 section1.1.5). In particular it was noticed that some participants closed their eyes in order to focus on the sound feedback. This behaviour suggested that the sounds may have required a high degree of concentration in order to be perceived and understood.

Given these observations it was thought necessary to assess the level of workload experienced in undertaking tasks with sound feedback. A measure of workload is the NASA Task Load Index (NASA-TLX; Hart & Staveland, 1988). This was thought an appropriate tool to use in understanding these issues as it was designed to assess workload under conditions of sustained attention. The NASA-TLX provides an index of global workload and identifies the relative contributions of six workload sources (mental demand, physical demand, temporal demand, performance, effort, and frustration). By using this measure it should be possible to gauge the overall level of workload experienced, and identify the relative contributions of different sources.

5.2 Conveying Orientation with Sound (Study 2 - Experiment 1)

5.2.1 Aims

Five research questions were posed at the start of this study:

1. Is sound an appropriate medium through which to convey information about curve orientation? That is, can participants confidently, quickly, and accurately identify the orientation of a curve? Does performance change with practice?
2. Is there an optimum sound type for conveying this information?
3. Does sound facilitate similar performance across all orientations?
4. What attitude do participants express towards their experience of using sound? Does this change with practice?
5. What level of workload do participants experience? Does this change with practice?

5.2.2 Method

5.2.2.1 Participants

A total of twenty participants were selected to undertake the study. Fifteen were postgraduate students recruited evenly from courses in music, physics/maths, and

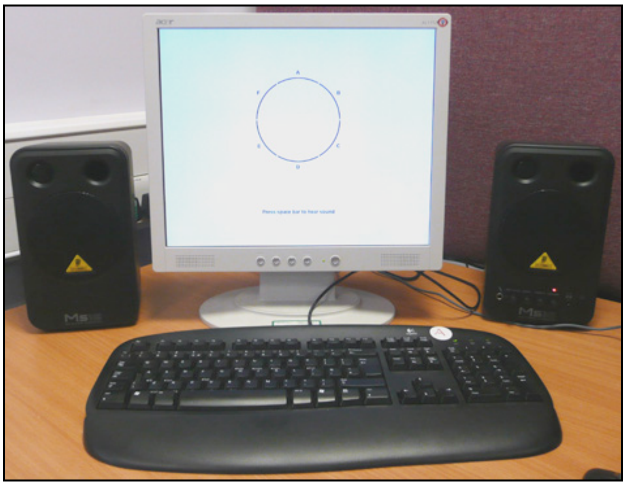


Figure 5-5:Set up for orientation experiment

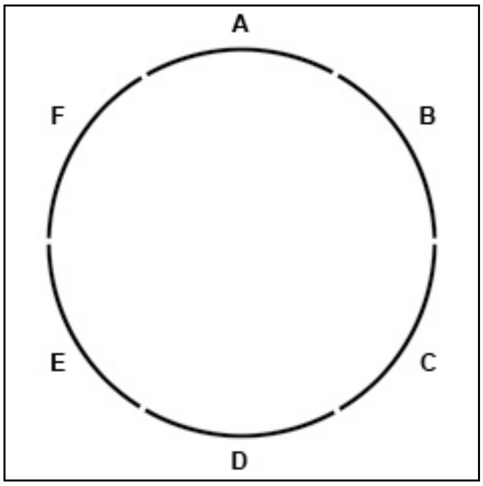


Figure 5-6: The six-segment circle displayed to participants

product design. Five participants were non-students from a broad range of backgrounds and were considered a 'general' category. Recruitment from this mix of backgrounds was designed to ensure a range of abilities for the specified tasks. The rationale for this choice of backgrounds was as follows: It was expected that physics/maths students would have a high conceptual understanding of curvature, and that music students would have a higher than average ability to discriminate sounds. Product designers were selected to represent the target users of the SATIN system, and general participants represented non-specialist potential users (for example, clients or managers). In this way the results would be balanced for background.

5.2.2.2 Equipment and Set up

The experiment took place in a dedicated usability lab. Participants were seated in front of a computer on which the experiment would be run. This was controlled by E-Prime software which enabled the participant to self-administer the experiment, whilst also collecting the data. A facilitator was present in order to introduce proceedings and resolve any problems. The sound was delivered through a pair of Behringer MS15 speakers positioned at either side of the monitor, and input was via the computer keyboard (see Figure 5-5).

5.2.2.3 Procedure

Each participant performed the experiment twice, once in the first session (before practice) and once in the second session (after practice). For the purposes of this experiment, 'practice' was to undertake Experiment 3 which involved listening to curve shape information and matching it to the correct visual curve (see Chapter 6 for details). There was approximately one week between the first and second sessions for this experiment. At the start of the first session participants completed a consent form which covered all experiments in this study. They were paid an inconvenience fee of £50 after completion of the second session.

The participant read through a description of the experiment and was given the opportunity to ask questions before commencing. The computer presented an image of a six segment circle to the participant (see Figure 5-6), and requested that they press the spacebar to hear the sound. On pressing the space bar the participant would hear a sound representing one of the segments. Once the sound finished (2 seconds) they were asked to identify the segment they had just heard by pressing the appropriate key from 'a' to 'f'; the diagram of the segments was still present. There was no opportunity for the participant to hear the sound again, and they could not progress until they had responded. Once they had responded, a further message asked them to rate their confidence in their judgement either 'high' (sure) or 'low' (not sure or don't know). Having responded they were again invited to press the space bar to hear the next sound. This sequence was repeated 18 times; once for each of the six segments, for each of the three sound types. The presentation of each sound was randomised. On completion of the 18th sound they were asked by the experimenter to complete a NASA-TLX workload assessment. Following this they completed a questionnaire about their experience of using sound to judge orientation. All documentation used to support this study can be found on the accompanying CD as detailed in Appendix G.

5.2.2.4 Stimuli

The task was designed to explore the use of sound to communicate the orientation of a particular curved segment (see Figure 5-6). For each sound heard, the participant stated which segment of the circle the sound related to. Therefore it was decided to use a 4.87cm radius circle split into six equal segments with a curvature of 20/m. This meant that each segment was roughly comparable with the E2 block from set C used in Study 1 (see Chapter 3) which had a good success rate (94%). An informal pilot of the task confirmed that this was an appropriate choice.

5.2.2.5 Sonification

5.2.2.5.1 Sound Types

The sounds were derived from those used in the SATIN prototype evaluations (described in Chapter 1 Section 1.1.4) and were designed by colleagues from Technische Universiteit Eindhoven who had expertise in sound design. Three different types of sound were used for this experiment; ‘Cello’, ‘Physical’, and ‘Sine’. The ‘Cello’ sound was a sampled cello with the decay removed. This presented a sound with a constant frequency of 131Hz which is equivalent to C3 (an octave below middle C). The ‘Physical’ sound was produced from the vibrations of an object, for instance when one flicks a glass or rubs a hand across a wooden table. In this case the object was a circular plate which produced a rich sound with inharmonic overtones. The ‘Sine’ sound was simply a sinusoidal tone. For a more technical description of these sounds see the paper co-authored with Shelley et al. (2009). These sounds are included on the accompanying CD within the ‘sounds’ folder.

5.2.2.5.2 Sound Mapping

Each sound was mapped to the highest and lowest points of the overall shape, so that at the midpoint of the top of the circle the frequency was 400Hz dropping to 100Hz as it reached the midpoint of the bottom of the circle. So for segment D which is at the bottom of the circle the sound started low falling to an even lower tone at 100 Hz then rising again to the starting frequency (because the shape here is symmetrical). This change in frequency was mapped to the change in x-position as the sound traced round the curve and so segments B and F, and C and E were inverses of each other; that is they use the same tones but played out in the opposite direction. This meant that segments B and C, and E and F could be easily confused as they had similar patterns of frequency change. For a more technical description of the way the sound was mapped to the curved shape see the paper co-authored with Shelley et al. (2009).

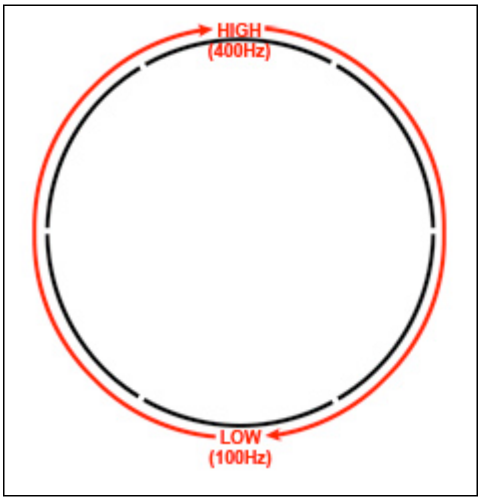


Figure 5-7: Sonification method following shape of curve.

5.2.2.6 Design and Hypotheses

A repeated-measures experimental design was applied comprising three main independent variables; practice (before, after), sound type (Cello, Physical, Sine), and curve orientation (A, B, C, D, E, F, G). The dependent variables were: response time (s), accuracy (%), and confidence (%). In order to answer the research questions posed at the start (see 5.2.1) of this experiment the following hypotheses were examined:

5.2.2.6.1 Performance (H1-H15)

H1-H3: There is an effect of practice on response time; accuracy; confidence.
H4-H6: There is an effect of sound type on response time; accuracy; confidence.
H7-9: There is an effect of curve orientation on response time; accuracy; confidence.
H10-H12: There is an interaction effect between practice and sound type on response time; accuracy; confidence.
H13-H15: There is an interaction effect between practice and orientation on response time; accuracy; confidence.

5.2.2.6.2 Error Rates (H16-17)

H16: There is an association between practice and error rates.
H17: There is an association between sound type and error rates.

5.2.2.6.3 User experience Issues (H18-H23)

H18: There is an effect of practice on attitude to user experience issues.
H19-H23: There a significant difference between expected and observed attitudes towards enjoyment; difficulty; performance; improvement; irritability.

5.2.2.6.4 Workload (H24-H30)

H24: There is an effect of practice on workload score.
H25-H30: There is an effect of practice on mental demand; physical demand; temporal demand; performance; effort; frustration.

5.2.2.7 Data Analysis

The ePrime data were collated for analysis in SPSS. In some cases variables were found to have a non-normal distribution; either positive or negative skew which is usual for these types of data. This was not a problem as ANOVAs are considered to be a robust method of statistical analysis (Davies 1956; Field 2009)(see Appendix F for discussion of this). A number of pre-planned contrasts were used in the analysis of these data. For sound types all were contrasted to ‘Sine’ which was the most pure sound (being a sine wave) and should therefore make a good comparator. For the segments, all were contrasted with ‘a’ as this was considered the optimum orientation (and matched that of the SATIN prototype). For other statistical tests the appropriate parametric analysis methods were used for non-normal variables (the tests used are noted within the results).

5.2.3 Results

This section details all results in detail, a summary of the results in relation to the hypotheses can be found in Appendix D.

5.2.3.1 Response Time

A repeated-measures ANOVA compared the effect on response time of practice, and sound type (see Figure 5-8 for response times before and after practice for each sound type). This revealed that there was a main effect of practice, such that response time before practice (5.52s, SD=1.49) was significantly slower than after practice (4.54s, SD=1.09), $F(1,19)=22.52$, $p<0.01$, partial $\eta^2=0.54$. There was no significant effect of sound type, or interaction effect between practice and sound type, $F(2,38)=1.17$, $p=0.32$, partial $\eta^2=0.06$, and $F(2,38)=0.44$, $p=0.65$, partial $\eta^2=0.02$, respectively.

A further repeated-measures ANOVA compared the effect on response time of practice, and curve orientation (see Figure 5-9 for response times before and after practice for each curve orientation). This revealed that there was a main effect of practice, meaning that response time before practice (5.52s, SD=1.92) was significantly slower than after practice (4.54s, SD=1.34), $F(1,19)=22.52$, $p<0.01$, partial $\eta^2=0.54$. It also revealed that there was no significant effect of orientation, or interaction effect between practice and orientation, $F(5,95)=1.42$, $p=0.22$, partial $\eta^2=0.07$, and $F(5,95)=0.15$, $p=0.98$, partial $\eta^2=0.01$, respectively.

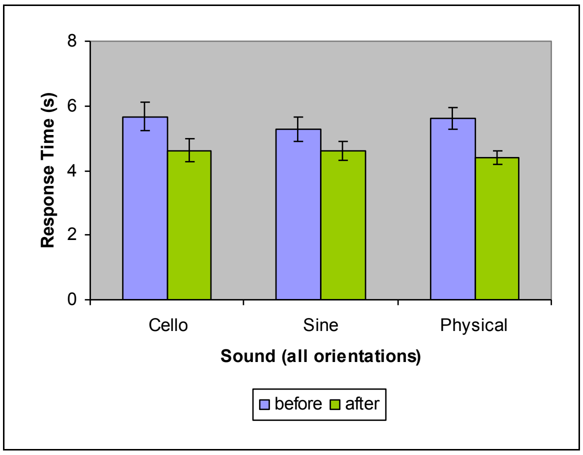


Figure 5-8: Response times for each sound type before and after practice

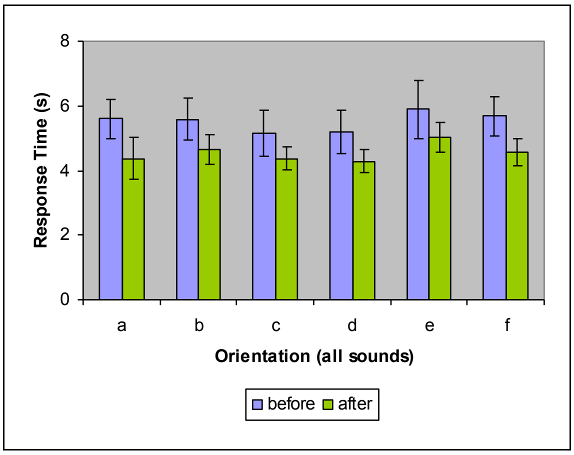


Figure 5-9: Response times for each curve orientation before and after practice

Note: Error bars represent the 95% Confidence Interval of the mean, and do so throughout

5.2.3.2 Accuracy

A repeated-measures ANOVA compared the effect of practice and sound type on accuracy (see Figure 5-10 for accuracy before and after practice for each sound type). This revealed that there was a main effect of practice, such that accuracy before practice (65.28%, SD=27.67) was significantly less than after practice (72.22%, SD=23.30), $F(1,19)=5.72$, $p=0.03$, partial $\eta^2=0.23$. It also revealed that there was a significant effect of sound type, $F(2,38)=29.08$, $p<0.01$, partial $\eta^2=0.60$. Pre-planned contrasts showed that ‘Sine’ (77.92%, SD=25.43) was significantly more accurate than ‘Physical’ (50.83%, SD=17.28), although it was not significantly different to ‘Cello’ (77.50%, 24.03), $F(1,19)=45.74$, $p<0.01$, partial $\eta^2=0.71$, and $F(1,19)=0.01$, $p=0.93$, partial $\eta^2=0.001$, respectively. Finally, there was no interaction effect between practice and sound type $F(2,38)=0.09$, $p=0.91$, partial $\eta^2=0.01$.

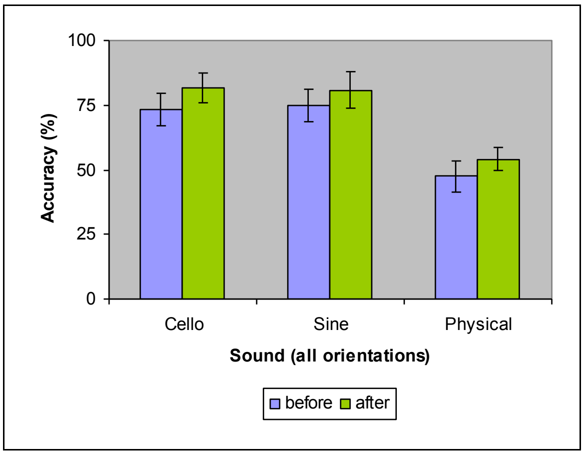


Figure 5-10: Level of accuracy for each sound type before and after practice

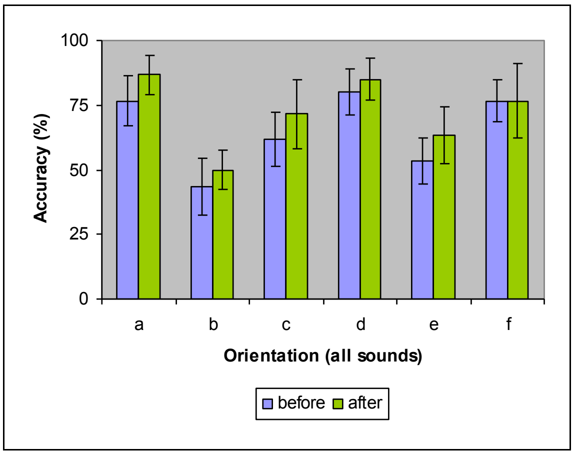


Figure 5-11: Level of accuracy for each curve orientation before and after practice

A further repeated-measures ANOVA compared the effect on accuracy of practice, and curve orientation (see Figure 5-11 for accuracy before and after practice for each curve orientation). This revealed that there was a main effect of practice, such that accuracy before practice (65.28% SD=31.87) was significantly less than after practice (72.22%, SD=29.73), $F(1,19)=5.72$, $p=0.03$, partial $\eta^2=0.23$. It also revealed that there was a significant effect of orientation, $F(3.23,61.43)=10.91$, $p<0.01$, partial $\eta^2=0.36$. Pre-planned contrasts showed that segment ‘a’ (81.67%, SD=29.19) was significantly more accurate than ‘b’ (46.67%, SD=27.01) or ‘e’ (58.33, SD=27.99) ,$F(1,19)=44.33$, $p<0.01$, partial $\eta^2=0.70$, and $F(1,19)=15.91$, $p<0.01$, partial $\eta^2=0.46$, respectively. There was no significant difference between segment ‘a’ and all other segments. Finally, there was no interaction effect between practice and orientation $F(1,19)=0.27$, $p=0.61$, partial $\eta^2=0.01$.

5.2.3.3 Confidence

A repeated-measures ANOVA compared the effect on confidence of practice, and sound type (see Figure 5-12 for confidence before and after practice for each sound type). This revealed that there was a main effect of practice, meaning that confidence before practice (64.44% SD=25.58) was significantly less than after practice (73.33%, SD=23.41), $F(1,19)=6.05$, $p=0.02$, partial $\eta^2=0.24$. It also revealed that there was no significant effect of sound type, or interaction effect between practice and sound type, $F(1,19)=2.78$, $p=0.07$, partial $\eta^2=0.13$, and $F(2,38)=1.31$, $p=0.28$, partial $\eta^2=0.06$, respectively.

A further repeated-measures ANOVA compared the effect on confidence of practice and orientation (see Figure 5-13 for confidence before and after practice for each curve orientation). This revealed that there was a main effect of practice, such that confidence before practice (64.44% SD=33.68) was significantly more than after practice (73.33%, SD=31.04), $F(1,19)=6.05$, $p=0.02$, partial $\eta^2=0.24$. It also revealed that there was no significant effect of orientation, or interaction effect between practice and orientation, $F(5,95)=1.06$, $p=0.39$, partial $\eta^2=0.05$, and $F(5,95)=2.26$, $p=0.06$, partial $\eta^2=0.11$, respectively.

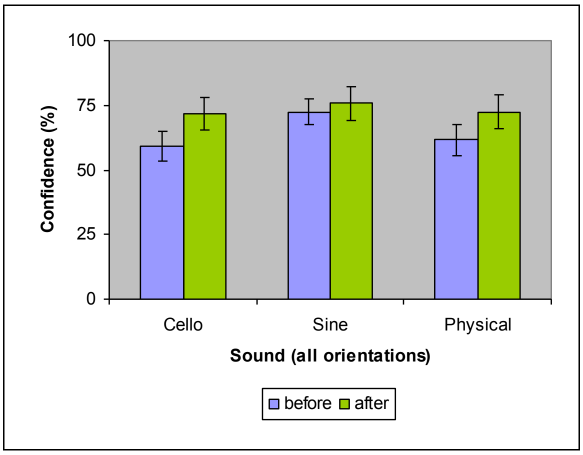


Figure 5-12: Level of confidence before and after practice for different sound types

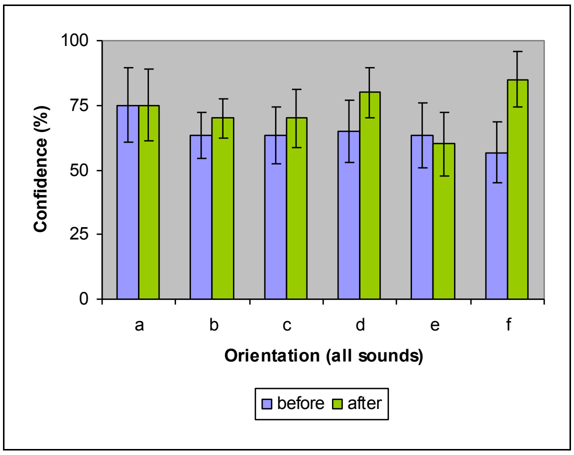


Figure 5-13: Level of confidence before and after practice for different orientations

5.2.3.4 Error Rates

The accuracy data were coded into four types of error (no error, mirror, shift, and total) and the frequency for each calculated. No error occurred when the segment was accurately identified. Mirror was where the segment was incorrectly identified as the vertical or horizontal mirror of the correct segment, for example the answer ‘b’ when the correct answer was ‘f’, or ‘b’ when the correct answer was ‘c’. Shift was when segments ‘a’ or ‘d’ were incorrectly identified as a neighbouring segment, so for a that was ‘f’ or ‘b’, and for ‘d’ that was ‘c’ or ‘e’. A total error was when there was considered to be no relationship (mirror or shift) of the segment identified to the correct segment, for example responding ‘c’ when the answer was ‘a’.

The frequencies of errors before and after practice are shown in Figure 5-14. This indicates that there was a slight increase in accuracy (no error) and decrease in error rate (mirror, shift, and total) after practice. However, a two-variable chi-squared test revealed that there was no association between practice and error type, $\chi^2(3, n=720) = 6.09, p>0.05$.

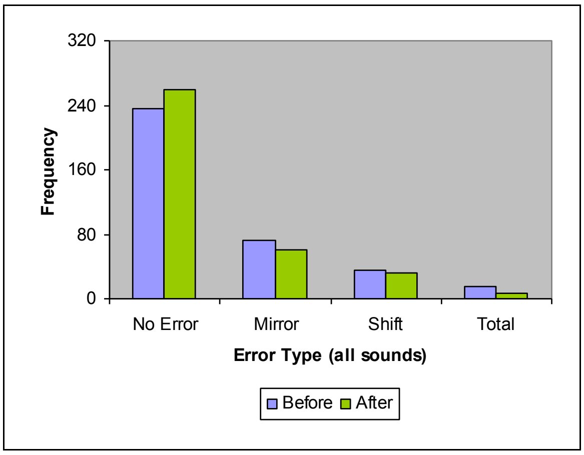


Figure 5-14: Error frequency before and after practice

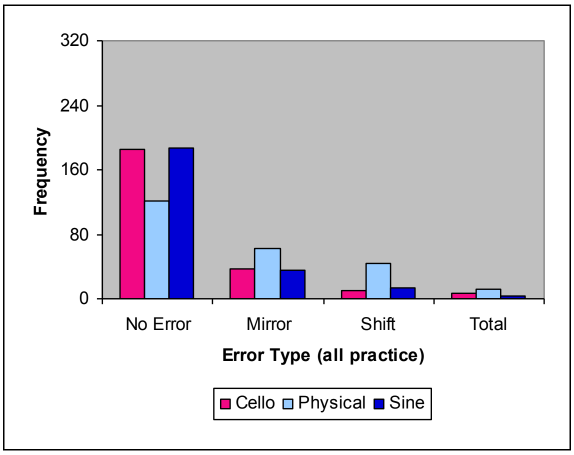


Figure 5-15: Error frequency by sound type

The error frequency in relation to each sound type is shown in Figure 5-15. This indicated that ‘Cello’ and ‘Sine’ had similar levels of error, and that both promoted a more accurate performance than ‘Physical’. A two-variable chi-squared test confirmed that there was a significant association between sound type and error, $\chi^2(6, n=720) = 61.68, p<0.05$. An examination of standardized residuals showed that performance with the ‘Physical’ sound had significantly more than expected mirror and shift errors and significantly less than expected no errors, z=2.6, 4.5, and -3.3 respectively. Residuals also showed that Cello had significantly less than expected shift error, z=-2.7.

5.2.3.5 User Experience

The frequencies of responses for each question in the user experience questionnaire are given in Figure 5-16. The graphs indicate little difference between attitudes before and after practice. Five two-variable (practice, attitude) chi-squared tests compared the before and after practice frequencies for each user experience question. These found no significant ($p>0.05$) difference between attitudes before and after practice.

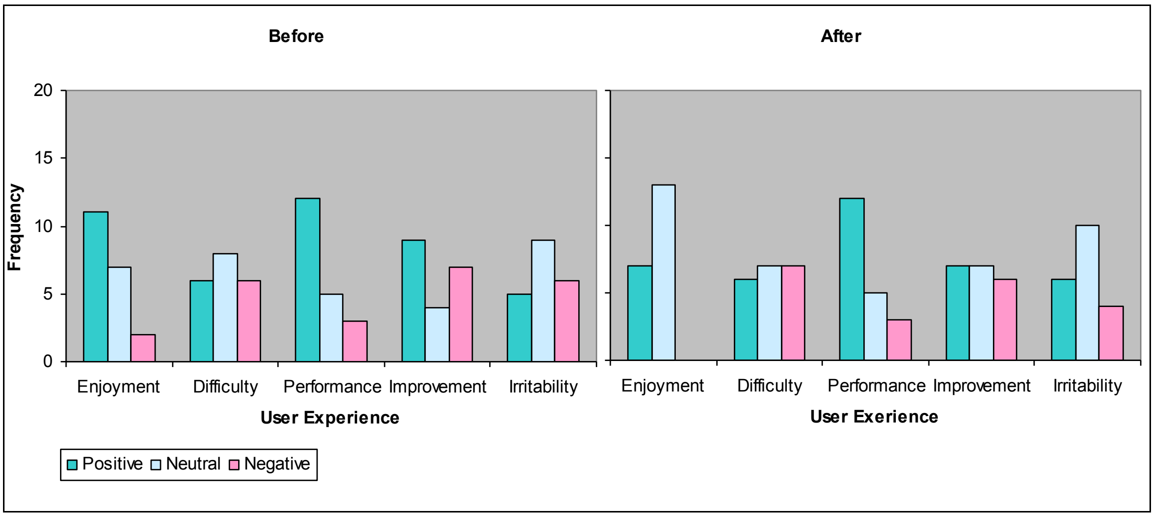


Figure 5-16: User Experience – attitude frequencies for each question before and after practice

Across both practice conditions there appeared to be an even spread of attitude towards all user experience issues except for 'Performance', for which a more positive attitude seemed to be expressed. The data across both practice conditions was collapsed (by averaging the data), and a series of five one-factor chi-squared tests were used to analyse the difference between expected and actual frequencies for each level of attitude (positive, negative, and neutral). The results are shown in Table 5-1, and confirmed that there was an even spread of attitude towards all issues except for 'Performance', which had a significantly positive attitude expressed towards it. There was a significant result for 'Enjoyment' and this was due to the low negative attitude.

Table 5-1: Association of frequency with attitude

Issue	n	χ^2	df	Exact Sig.
Enjoyment	20	7.30	2	0.02
Difficulty	20	0.28	2	0.96
Performance	20	6.70	2	0.04
Improvement	20	0.29	2	0.96
Irritability	20	2.00	2	0.47

5.2.3.6 Workload

NASA-TLX was used to provide an estimation of the subjective workload experienced by each participant. The mean workload score before practice was 46.97 (SD=9.88) and after was 41.9 (SD=10.93). This indicates that, although decreased, there was little difference in subjective workload before and after practice. A dependent t-test confirmed that there was no significant difference between the two scores, t(19)=1.89, p=0.07, r=0.16.

A breakdown of the sub-factors influencing workload are shown in Figure 5-17. This gives an idea of the relative contribution each factor made to the participants' sense of workload. Mental demand, performance, and effort are scored more highly than the other factors. As with the workload score, the sub-factors changed little after practice compared to before. An exception to this was mental demand which showed a noticeable decrease in score from 210 (SD=94.28) before practice to 158.5 (SD=127.02) after. A dependent t-test confirmed that there was a significant difference between scores for this sub-factor, t(19)=2.20, p=0.04, r=0.20. All other sub-factors when tested (dependent t-tests) showed non-significant results (p>0.05).

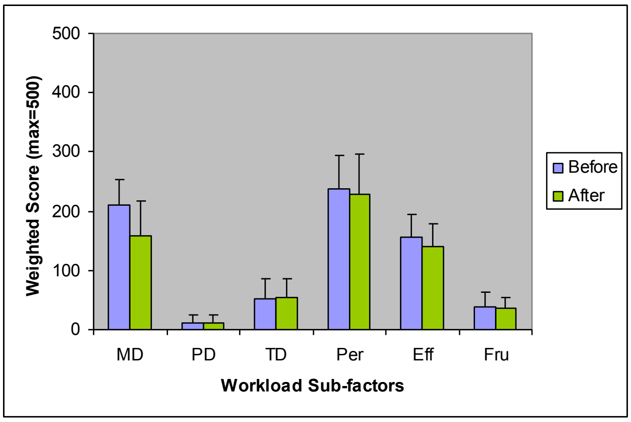


Figure 5-17: Workload Sub-Factors

Note: Mental Demand (MD), Physical Demand (PD), Temporal Demand (TD), Performance (Per), Effort (Eff), and Frustration (Fru)

5.2.3.7 Background

Table 5-2 summarises performance measures according to participant background. The results here are combined for all sounds as the primary aim of this analysis is to examine the impact of participant background on performance in general.

Table 5-2: Performance according to participant background

Background	Response Time (s)		Accuracy (%)		Confidence (%)	
	Mean	SD	Mean	SD	Mean	SD
Music	4.31	0.79	78.89	4.21	78.33	20.54
Physics/Maths	4.65	0.71	80.56	10.39	72.22	20.13
Product Design	5.48	1.12	65.56	15.04	65.00	14.65
General	5.69	0.80	50.00	17.46	60.00	22.10

A separate one-way independent ANOVA was undertaken for each of the performance measures. These revealed that there was no significant effect of background on either response time or confidence, $F(3,16)=2.88$, $p=0.07$, and $F(3,16)=0.85$, $p=0.49$. However, there was a significant effect of background on accuracy, $F(3,16)=6.13$, $p=0.01$. Pre-planned post-hoc tests (Dunnett's t-test) showed that those with a 'General' background were significantly less accurate than those with a 'Music' or 'Physics/Maths' background, $p<0.05$. There was no significant difference between 'General' and 'Product Design', $p>0.05$.

5.2.4 Discussion

5.2.4.1 The Appropriateness of Sound

The experiment sought to address the following questions in relation to the appropriateness of sound:

- Is sound an appropriate medium through which to convey information about curve orientation? That is, can participants confidently, quickly, and accurately identify the orientation of a curve? Does performance change with practice?
- Is there an optimum sound type for conveying this information?
- Does sound facilitate similar performance across all orientations?

The experiment indicated that participants were able to quickly (5s), accurately (69%) and confidently (69%) identify the orientation of a curve. The level of performance across these different measures was high, and there appears to have been little difficulty in achieving this task. Whilst performance was good, intuitively it was significantly better after participants had undertaken a few hours of practice. However, the improvements in response time, accuracy, and confidence were relatively small at, -1s, +7%, and +9% respectively. This shows that there is some benefit in acquainting users with sound exploration, but that the performance improvements are not substantial.

The type of sound used to explore curve orientation was found to have a significant effect on participant's accuracy (there was no effect on response time or confidence).

When compared with 'Sine' (78%) the level of accuracy attained with 'Cello' (77.5%) was similar, but 'Physical' (51%) was significantly less. It is likely that the increased 'noise' associated with this sound type of sound may have made it difficult for participants to correctly identify the orientation of the curve. Further insight into this was gained from examining the error data. Use of the 'Physical' sound led to higher than expected 'shift' and 'mirror' errors, accounting for 18% and 26% of judgments respectively and so contributing 44% of the total drop in accuracy. These errors are illustrated in Figure 5-18. The mirror errors were of two types, horizontal between B-C and F-E, and vertical between B-F and E-C.

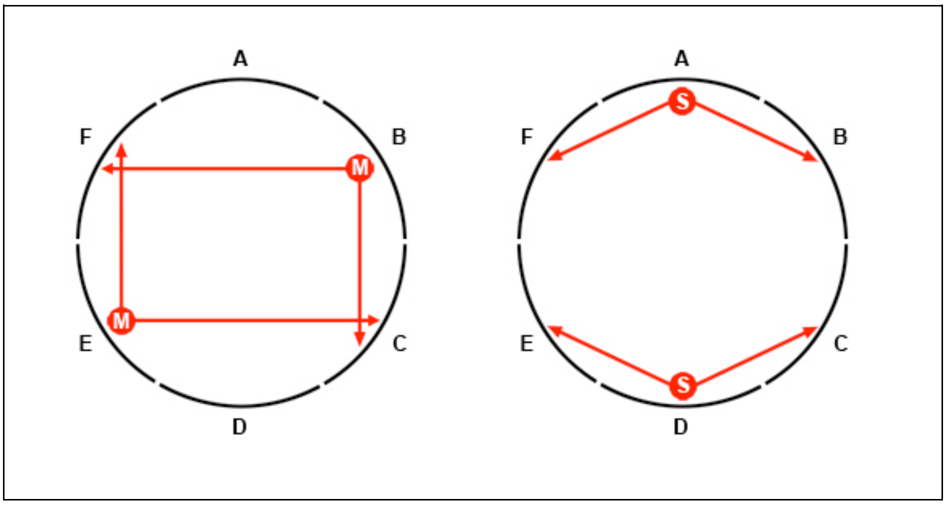


Figure 5-18: Mirror (left) and Shift (right) errors

Practically all the 'mirror' errors (23%) were of the first type. In essence these are the same as shift errors since identification has shifted to a neighbouring segment. So for 'Physical' 51% of judgments were correct and 41% were errors based on a shift in perception to a neighbouring segment. These types of error are likely to occur because the participant cannot accurately judge the relative level of the start frequency and so confuse the segment with an adjoining one. The low rate of other types of error (vertical mirror, vertical-horizontal mirror, and total error) means they had a sense of the direction of frequency change (high to low or low to high) but had insufficient sense of the starting frequency. This type of error was seen in the other two sound types, but was not of a significant level. This would indicate that there is something in the nature of the physical sound that makes it more prone to this type of error. Given the increased 'noise' within this type of sound, as opposed to the 'Sine' or 'Cello', then it is likely that this is the cause of the perceptual confusion observed. The level of error was not something that decreased significantly with practice, and so again indicates a more inherent property of the sound that led to a continued misperception.

The orientation of the curve was observed to have an effect on the participant's performance. Whilst the level of confidence and response time was independent of curve orientation, it was found that curve orientation 'A' was more accurately judged than orientations 'B' or 'E'. For all other curve orientations judgements were not significantly different to 'A'. This prompts the question as to why judgements for 'B' and 'E' should be less accurate. The answer may lie in the sonification method. Firstly we will discount 'A' and 'D' from this discussion as they are fundamentally different to 'B' and 'E' in that they have a pattern of increasing and decreasing frequency that is relatively easy to perceive. To understand what may be happening we need to consider what differentiates 'B' and 'E' from 'C' and 'F'. These share a similar pattern of constant rise or constant fall in frequency. However, they differ in the rate of change of frequency at the beginning and the ends of the curve. This is illustrated in Figure 5-19 by looking at the difference between 'B' and 'C'. It can be seen that 'C' had a pronounced change in frequency at the beginning of the curve whereas 'B' had a more subtle change in frequency (indicated by the '1' arrow in the diagram), this situation is reversed for the latter half of the curve (indicated by the '2' arrow in the diagram). It is likely that the subtle change seen in 'B' and 'E' is more difficult to perceive and so leads to greater error, whereas the more pronounced change in 'C' and 'F' signals these orientations more effectively and so leads to less error. This effect is possibly compounded by the fact that participants were only able to hear the sounds once. It may therefore be the case that when sounds can be heard multiple times this difference in frequency change may have less effect on performance.

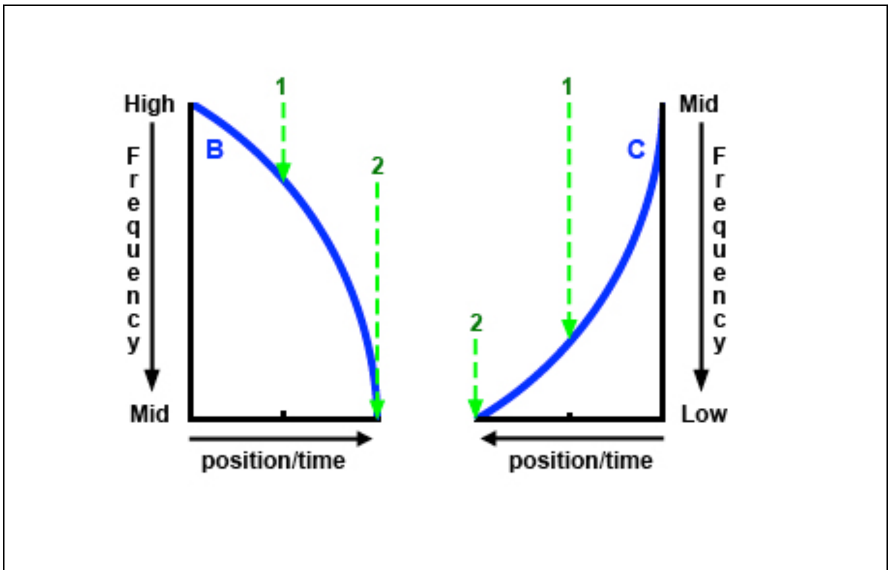


Figure 5-19: Comparison of Rate of Change in Frequency for Curve Orientations B (left) and C (right)

5.2.4.2 User Experience Issues and Workload

The experiment sought to address the following questions in relation to user experience and workload:

- What attitude do participants express towards their experience of using sound? Does this change with practice?
- What level of workload do participants experience? Does this change with practice?

The attitudes towards 'Difficulty', 'Improvement', and 'Irritability' were found to be evenly spread. This meant as many people were positive as negative, and that a similar number expressed a neutral attitude. Whilst this means that around two-thirds of participants did not feel negatively, a substantial proportion of those that undertook the task did. Therefore around a third of participants felt that the task was difficult, that they had no sense of improving at it over time, and that they found the sounds to be irritating. Despite this there was a significant lack of negative attitude towards 'Enjoyment' of the task, with the majority of participants having a neutral to positive attitude to the statement 'I found the task enjoyable'. In addition a significant number of participants had a positive attitude towards their performance.

User experience therefore seems to be somewhat mixed. Whilst there was little disagreement to the view that the task was enjoyable and a belief by many that they performed accurately, there was a proportion of participants who were irritated by the sounds, found the task difficult and had no sense that they had improved. These attitudes did not change with practice, and so greater familiarity with the task did not help overcome some of these negative aspects.

The level of workload experienced by participants was moderate, with an average NASA-TLX score of 45 across the two sessions. The NASA-TLX score ranges from 0 to 100, so the score here would indicate that participants were operating fairly comfortably within the middle of the range. The workload score was not significantly changed by practice. However, on examination of the sub-factors it was found that there was a significant decrease in Mental Demand after practice. This coheres with the performance results which saw an improvement after practice. So not only does performance improve but the mental demands of the task decrease with greater familiarity. Again this suggests that sound is an appropriate medium to use.

5.2.4.3 User Background

Within the experimental design, background had not been considered as a separate factor, and steps were taken to ensure that there was a range of aptitudes in the participants undertaking the task. During the course of the experiment it was observed that the 'General' category seemed less able in performing the task. Participants within this category had no formal musical training and this may have affected their ability to undertake the task. Whereas there was a high-level of training within the 'Music' category as these had been selected from students undertaking a music degree. There was at least a moderate level of musical training in both the 'Maths/Physics' and 'Product Design' categories as these participants had indicated undertaking graded exams (between grades 2 to 6). For this reason it was decided to explore the data to identify any performance differences between the 'General' group and the other three categories.

It was found that there was no significant difference between categories for response time and confidence, but that there was a significant difference in accuracy between the 'General' category (50%) and both the 'Music' and 'Physics/Maths' categories (79% and 81% respectively). This suggests that musical-training can improve performance. It is also interesting to note that those with a musical background had less variability (SD=4%) in their accuracy than those from other types of background. This coheres with other results which showed that practice improved performance, and as musicians develop their skills through practice then this level of consistency is not surprising.

5.3 Conveying Magnitude with Sound (Study 2 - Experiment 2)

5.3.1 Aims

Five research questions were posed at the start of this study:

1. Is sound an appropriate medium through which to convey information about curve magnitude? That is, can participants confidently, quickly, and accurately identify the magnitude of a curve? Did performance change with magnitude difference? Does performance change with practice?
2. Is there an optimum sound type for conveying this information?
3. How do sonification methods compare?
4. What attitude do participants express towards their experience of using sound? Does this change with practice?
5. What level of workload do participants experience? Does this change with practice?

5.3.2 Method

5.3.2.1 Participants

These were the same as in Experiment 1 (see 5.2.2.1).

5.3.2.2 Equipment and Setup

This was the same as for Experiment 1 (see 5.2.2.2).

5.3.2.3 Procedure

Each participant read through a description of the experiment and was given the opportunity to ask questions about this by the facilitator before commencing. On commencing the experiment the computer presented the words ‘Sound A’ and ‘Sound B’ on the left and right of the screen with the instruction to ‘press the spacebar to hear the sounds’ at the bottom. On pressing the spacebar ‘Sound A’ was highlighted and a sonified curved sound was heard for two seconds. The word ‘Sound B’ was then highlighted and a different sonified curve sound played for two seconds. The sounds presented consisted of the standard curve sonification and a randomly chosen comparison curve sonification. The assignment of these as A or B was also randomised. When sound B had finished, the participant was prompted by a message on the screen to indicate if sound B was more or less curved than sound A. There was no opportunity for the participant to hear the sounds again, and they could not progress until they had responded. Once they had responded, a further message asked them to rate their confidence in their judgement either 'high' (sure) or 'low' (not sure or don't know). The sequence then started again and was repeated until all the comparison sounds had been heard in each position (A and B). On completion of the last pair of sounds the participant was asked by the facilitator to complete a NASA-TLX workload assessment. Following this they completed a questionnaire about their experience of using sound to judge magnitude. All documentation used to support this study can be found on the accompanying CD as detailed in Appendix G.

This procedure was undertaken firstly for the curve shape sonification and secondly for the curvature sonification. The presentation of each formed one session. In all two sessions were undertaken for Experiment 2, one before practice and one after practice (see Figure 5-1). For the purposes of this experiment practice was the undertaking of Experiment 3.

5.3.2.4 Curved Shapes

In order for sounds to be produced it was necessary to specify a number of curved shapes upon which the sonifications would be based. This required the specification of a standard curve, the one to which all others were compared, and six comparison curves (three decreasingly flatter, and three increasingly more curved). To enable comparisons with haptic-visual performance the standard curve chosen was based on the stimulus 'C' standard used in Study 1 (see Chapter 3), the comparison curves were also derived from this set (see Table5-3 for specifications). The curve to be sonified could then be derived as shown in Figure 5-20.

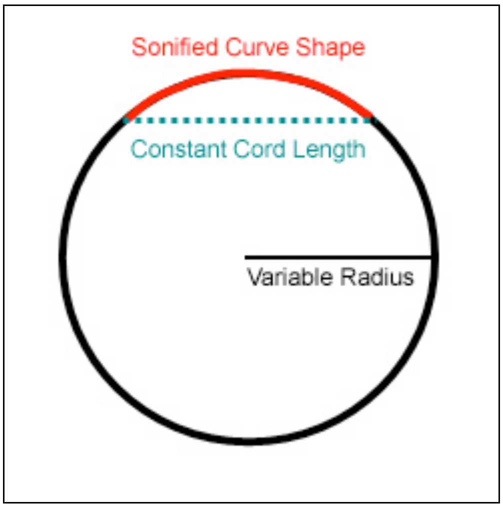


Figure 5-20: Derivation of curve to be sonified

Table 5-3: Specification for curved shapes

ID	Mag.Diff.	Curv.(/m)	Rad.(cm)
G1	Large	10.28	9.73
D1	Medium	12.42	8.05
A1	Small	15.00	6.67
Std	-	15.79	6.26
A2	Small	17.01	5.88
D2	Medium	20.55	4.87
G2	Large	24.82	4.03

5.3.2.5 Sonification

5.3.2.5.1 Sound Types

These are as described in Experiment 1 (see 5.2.2.5.1).

5.3.2.5.2 Sound Mapping

Two mappings were used in this experiment; curved shape and curvature. For both mappings the frequency ranged from 100Hz to 400Hz (about a two octave range). The curved shape sonification follows the shape of the curve from left to right. So the sound begins with a frequency of 100Hz at the left end of the curve, rises in frequency to the midpoint and then decreases in frequency back to 100Hz at the right end of the curve. The midpoint frequency changes depending on the height of the arc sonified with the maximum 400Hz reached with the highest arc; curve G2.

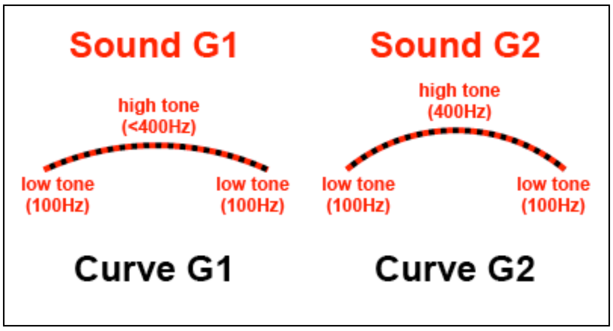


Figure 5-21: Curve shape sonification for curve magnitude

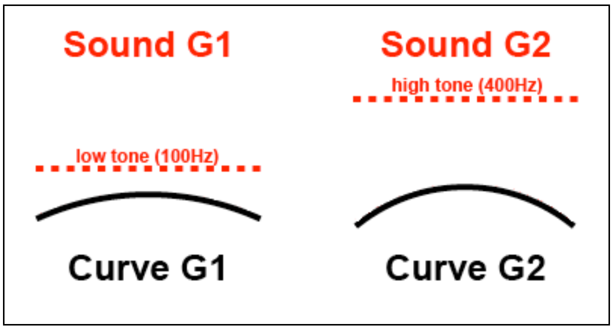


Figure 5-22: Curvature sonification for curve magnitude

The curvature sonification is mapped to the curvature value of the curve. The shape with the lowest curvature is G1 (10.28/m) and this is mapped to a frequency of 100Hz. The shape with the highest curvature is G2 (24.82/m) and is mapped to a frequency of 400Hz. Since the curvature of any point in a circle is the same, when the sound travels from the left to the right of the curved shape (arc of a circle) the tone produced remains constant.

5.3.2.6 Design and Hypotheses

The experiment was a repeated-measures design. There were three main independent variables; practice (before, after), sound type (Cello, Physical, Sine), and magnitude difference (small, medium, large). The effects of these would be measured on the dependent variables response time, accuracy, and confidence.

Whilst there were two sonification methods, curve shape and curvature, it was felt that the addition of another experimental variable would over complicate an already difficult analysis, so the analysis was conducted separately for each. However the aims and the hypotheses considered were the same for both and a comparison of the two will be included in the discussion.

In order to answer the research questions posed at the start (see 5.3.1) of this experiment the following hypotheses were examined:

5.3.2.6.1 Performance (H1-H18)

H1-H3: There is an effect of practice on response time; accuracy; confidence.
H4-H6: There is an effect of sound type on response time; accuracy; confidence.
H7-9: There is an effect of magnitude difference on response time; accuracy; confidence.
H10-H12: There is an interaction effect between practice and sound type on response time; accuracy; confidence.
H13-H15: There is an interaction effect between practice and magnitude difference on response time; accuracy; confidence.
H16-H18: There is an interaction effect between sound type and magnitude on response time; accuracy; confidence.

5.3.2.6.2 User experience Issues (H19-H24)

H19: There is an effect of practice on attitude to user experience issues.
H24: There a significant difference between expected and observed attitudes towards enjoyment; difficulty; performance; improvement; irritability.

5.3.2.6.3 Workload (H25-H31)

H25: There is an effect of practice on workload score.
H26-H31: There is an effect of practice on mental demand; physical demand; temporal demand; performance; effort; frustration.

5.3.2.7 Data Analysis

The ePrime data were collated for analysis in SPSS. The data for each performance measure were grouped into small (A1, A2), medium (D1, D2), and large (G1, G2) magnitude differences and means derived for each sound type. Each magnitude group is therefore based on 4 trials from 20 participants which gave 80 trials in total. The raw scores for accuracy and confidence were converted to percentages.

In some cases variables were found to have a non-normal distribution; either positive or negative skew which is usual for these types of data. This was not a problem as

ANOVAs are considered to be a robust method of statistical analysis (Davies 1956; Field 2009)(see Appendix F for discussion of this).

A number of pre-planned contrasts were used in the analysis of these data. For sound types all were contrasted to ‘Sine’ which was the most pure sound (being a sine wave) and should therefore make a good comparator. For magnitude difference ‘medium’ and ‘large’ were contrasted with ‘small’, as this would give a sense of the level of acuity.

5.3.3 Results

This section details all results in detail, a summary of the results in relation to the hypotheses can be found in Appendix D.

5.3.3.1 Response time

5.3.3.1.1 Curve Shape

A repeated-measures ANOVA compared the effect on response time of practice, sound type, and magnitude difference (see Figure 5-23 and Table 5-4). This revealed that there was a main effect of practice. After practice the response time is significantly faster (4.94s, SD=0.90) than before practice (5.45s, SD=1.16), $F(1,19)=31.46$, $p<0.01$, partial $\eta^2=0.62$. There was no main effect of sound type, $F(2,38)=0.33$, $p=0.72$, partial $\eta^2=0.02$, but there was of magnitude difference, $F(1.24,23.55)=29.46$, $p<0.01$, partial $\eta^2=0.61$. Pre-planned contrasts showed that a small magnitude difference (5.74s, SD=1.25) had a significantly longer response time than a medium (5.00, SD=0.87) or large (4.85s, SD=0.77) difference, $F(1,19)=44.06$, $p<0.01$, partial $\eta^2=0.70$, and $F(1,19)=29.27$, $p<0.01$, partial $\eta^2=0.61$, respectively.

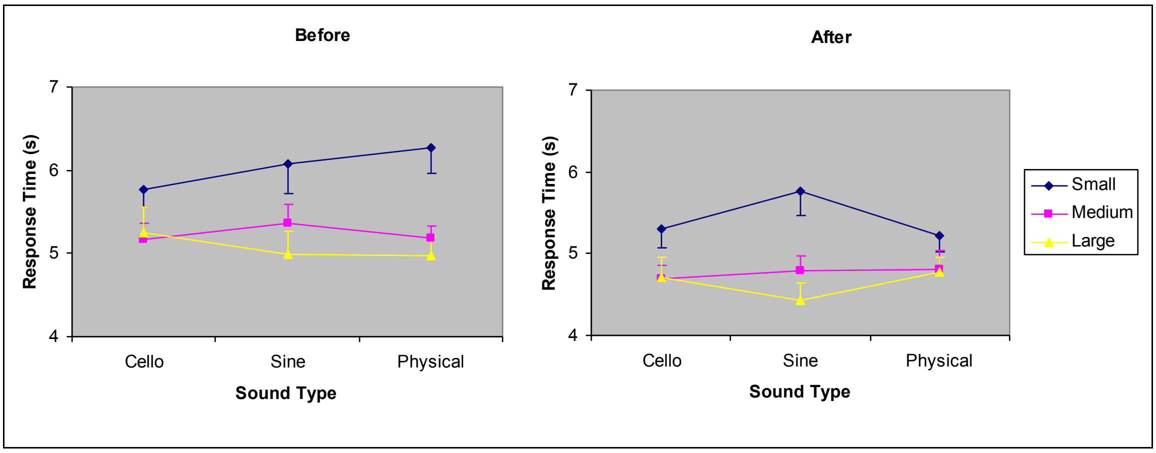


Figure 5-23: Effect of practice, magnitude, and sound type on response time

Note: Error bars represent the 95% Confidence Interval of the mean, and do so throughout

Table 5-4: Mean response time for Sound Type, Magnitude Difference, and Practice (curve shape)

Variable	Mean	SD
Cello	5.15	0.95
Physical	5.20	0.98
Sine	5.24	1.22
Small	5.74	1.25
Medium	5.00	0.87
Large	4.85	0.77
Before Practice	5.45	1.16
After Practice	4.94	0.90
All	5.20	1.05

There was no interaction between practice and either sound type or magnitude difference, $F(2,38)=0.10$, $p=0.90$, partial $\eta^2=0.01$, and $F(1.93,36.72)=0.82$, $p=0.44$, partial $\eta^2=0.04$, respectively. There was however a significant interaction effect of sound type and magnitude difference (see Figure 5-24), $F(2.86,54.36)=4.51$, $p=0.01$, partial $\eta^2=0.19$. Pre-planned contrasts revealed that the decrease in response time for large magnitude difference (compared with small) was significantly less for 'Cello' and 'Physical' (compared to 'Sine), $F(1,19)=10.00$, $p=0.01$, partial $\eta^2=0.35$, and $F(1,19)=6.81$, $p=0.02$, partial $\eta^2=0.26$, respectively.

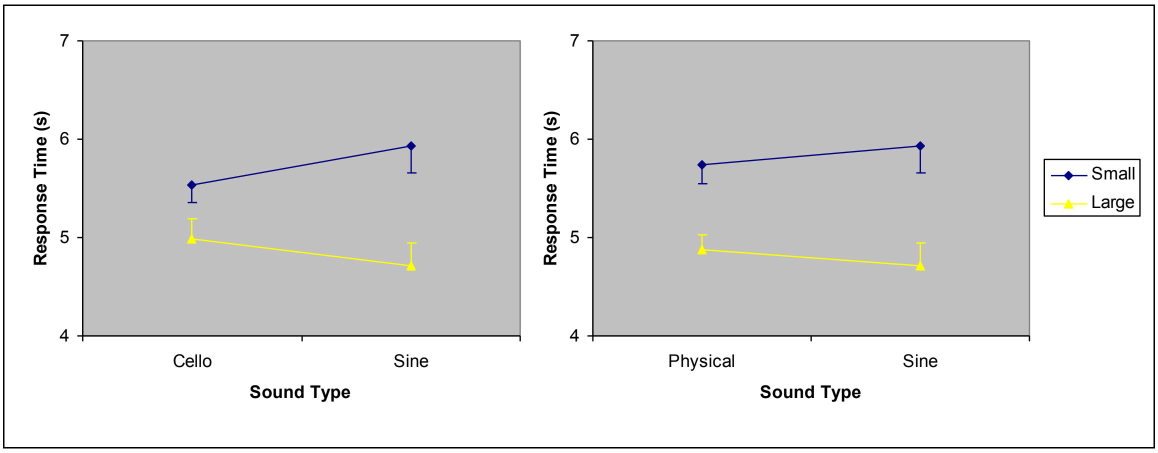


Figure 5-24: Interaction between sound type (curve shape) and magnitude difference

5.3.3.1.2 Curvature

A repeated-measures ANOVA compared the effect on response time of practice, sound type, and magnitude difference (see Figure 5-25[11] and Table 5-5). This revealed that there was a main effect of practice, such that after practice, response time was significantly higher (4.54s, SD=0.95) than before (4.91s, SD=1.18), $F(1,18)=7.39$, $p=0.01$, partial $\eta^2=0.29$. There was no main effect of magnitude difference, $F(1.51,27.19)=1.29$, $p=0.29$, partial $\eta^2=0.07$, but there was of sound type, $F(2,36)=13.61$, $p<0.01$, partial $\eta^2=0.43$. Pre-planned contrasts showed that 'Sine' (4.53s, SD=1.00) had a significantly shorter response times than 'Physical' (5.04s, SD=1.25), $F(1,18)=23.37$, $p<0.01$, partial $\eta^2=0.56$. There was no significant difference between response times for 'Sine' and 'Cello' (4.60s, SD=0.87), $F(1,18)=0.64$, $p=0.43$, partial $\eta^2=0.034$.

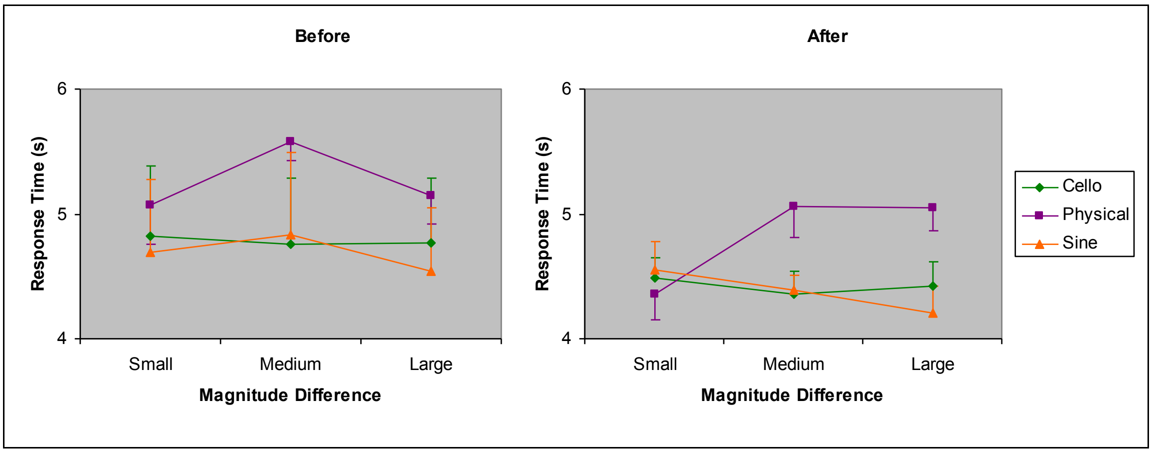


Figure 5-25: Effect of practice, sound type and magnitude difference on response time

Table 5-5: Mean response time for Sound Type, Magnitude Difference, and Practice (curvature)

Variable	Mean	SD
Cello	4.60	0.87
Physical	5.04	1.25
Sine	4.53	1.00
Small	4.66	0.96
Medium	4.83	1.28
Large	4.69	0.94
Before Practice	4.91	1.18
After Practice	4.54	0.95
All	4.73	1.07

[11] There has been a change of display (magnitude is represented on the x-axis) for the curvature graphs as there is no main effect of magnitude difference, but there is of sound type which is better illustrated in this way.

There was no significant interaction between practice and either sound type or magnitude difference, F(1.27,22.81)=0.31, p=0.63, partial η^2=0.17, and F(1.25,22.50)=0.45, p=0.55, partial η^2=0.025. There was however a significant interaction effect of sound type and magnitude difference (see Figure 5-26), F(2.51,45.25)=5.59, p<0.01, partial η^2=0.24. Pre-planned contrasts revealed that the increase in response time for 'Physical' (compared with 'Sine') was significantly greater for medium and large differences (as opposed to small), F(1,18)=17.47, p<0.01, partial η^2=0.49, and, F(1,18)=15.21, p<0.01, partial η^2=0.46, respectively. There was no significant interaction effect when 'Cello' and 'Sine' were compared for small/medium and small/large magnitude differences, F(1,18)=0.34, p=0.57, partial η^2=0.02, and F(1,18)=1.42, p=0.25, partial η^2=0.07, respectively.

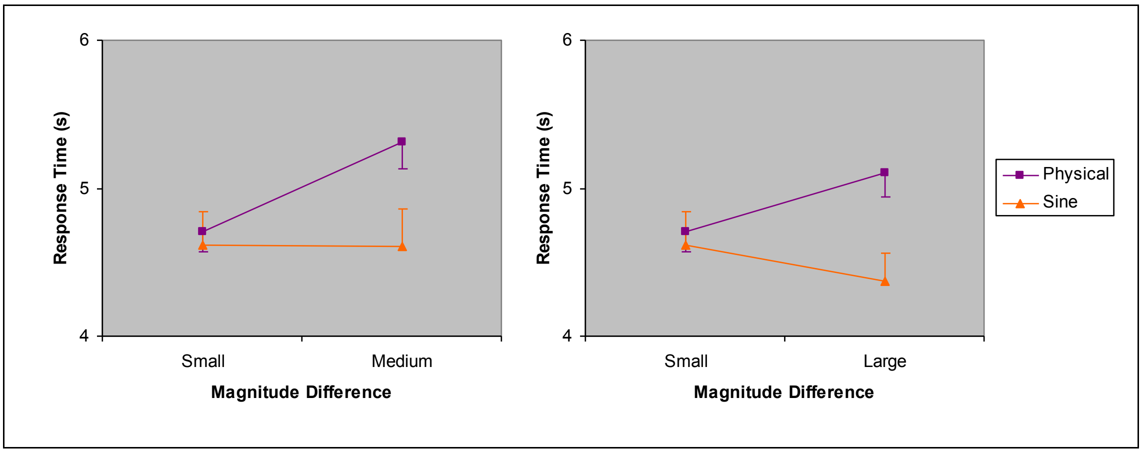


Figure 5-26: Interaction between sound type (curvature) and magnitude difference

5.3.3.2 Accuracy

5.3.3.2.1 Curve Shape

A repeated-measures ANOVA compared the effect on accuracy of practice, sound type, and magnitude difference (see Figure 5-27 and Table 5-6). This revealed that there was a main effect of practice, F(1.19)=5.03, p=0.037, partial η^2=0.21. After practice, accuracy (95.56%, SD=15.14) was significantly higher than before (93.19%, SD=11.51), F(1,19)=5.03, p=0.037, partial η^2=0.21. There was no main effect of sound type, meaning that any observed differences in accuracy are due to chance, F(2,38)=0.99, p=0.38, partial η^2=0.05. There was a main effect of magnitude difference, F(1.05,19.95)=15.46, p<0.01, partial η^2=0.45. Pre-planned contrasts showed that accuracy for small differences (86.88%, SD=19.16) was significantly less than when compared to medium (97.50%, SD=7.53) or large differences (98.75, SD=5.47), F(1,19)=15.74, p<0.01, partial η^2=0.45, and F(1,19)=15.70, p<0.01, partial η^2=0.45, respectively.

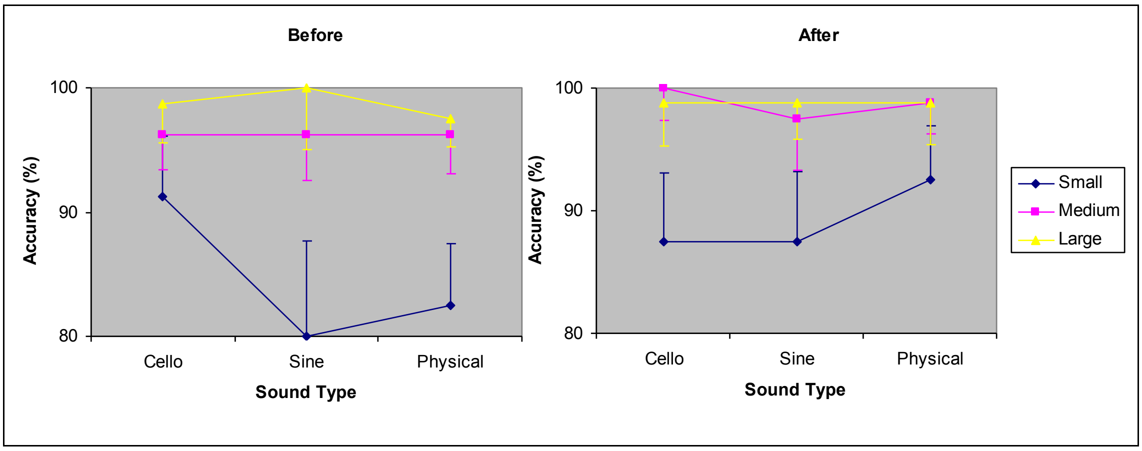


Figure 5-27: Effect of practice, sound type (curve shape), and magnitude difference on accuracy

Table 5-6: Mean accuracy for Sound Type, Magnitude Difference, and Practice (curve shape)

Variable	Mean	SD
Cello	95.42	12.55
Physical	94.38	11.89
Sine	93.33	15.45
Small	86.88	19.16
Medium	97.50	7.53
Large	98.75	5.47
Before Practice	93.19	15.14
After Practice	95.56	11.51
All	94.38	13.38

There was no interaction between practice and either sound type or magnitude difference, F(2,38)=1.75 , p=0.19, partial η^2=0.08, and F(1.39,25.42)=2.75 , p=0.10, partial η^2=0.13, respectively. There was also no interaction between sound type and magnitude difference, F(2.48,46.50)=1.15, p=0.057, partial η^2=0.06.

5.3.3.2.2 Curvature

A repeated-measures ANOVA compared the effect on accuracy of practice, sound type, and magnitude difference (see Figure 5-28 and Table 5-7). This revealed that there was a significant main effect of practice on accuracy, F(1,19)=11.86, p<0.01, partial η^2=0.38. Accuracy was significantly better after practice (94.58%, SD=11.72) than before (90.69%, SD=18.88). There was a main effect of sound type, F(1.33,25.29)=25.57, p<0.01, partial η^2=0.57. Pre-planned contrasts revealed that ‘Sine’ (96.67%, SD=11.20) was more accurate than ‘Physical’ (83.96%, SD=21.69), F(1,19)=25.81, p<0.01, partial

η^2=0.58, but that there was no significant difference in accuracy between 'Sine' and 'Cello' (97.29%, SD=9.05), F(1,19)=0.30, p=0.59, partial η^2=0.016. There was no main effect of magnitude difference, F(1.5,28.55)=0.23, p=0.73, partial η^2=0.01.

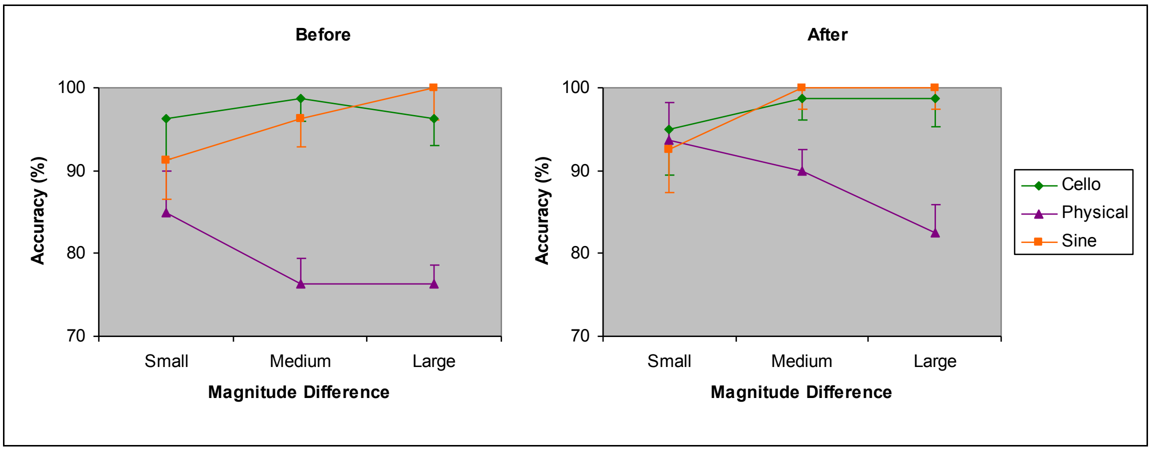


Figure 5-28: Effect of practice, sound type (curvature) and magnitude difference on accuracy

Table 5-7: Mean accuracy for Sound Type, Magnitude Difference, and Practice (curvature)

Variable	Mean	SD
Cello	97.29	9.05
Physical	83.96	21.69
Sine	96.67	11.20
Small	92.29	17.69
Medium	93.33	15.11
Large	92.29	14.31
Before Practice	90.69	18.88
After Practice	94.58	11.72
All	92.64	16.20

There was an interaction effect between practice and sound type, F(1.55,29.39)=5.22, p=0.017, partial η^2=0.22. This indicates that sound had a different effect on accuracy dependent upon whether it was heard before or after practice (see Figure 5-29). Pre-planned contrasts revealed that the increase in accuracy after practice was significantly greater for 'Physical' than for 'Sine', F(1,19)=4.46, p=0.048, partial η^2=0.19. There was no significant difference in the increase of accuracy after practice when comparing 'Sine' and 'Cello', F(1,19)=0.30, p=0.59, partial η^2=0.016. There was no interaction effect between practice and magnitude difference, F(2,38)=0.10, p=0.37, partial η^2=0.05.

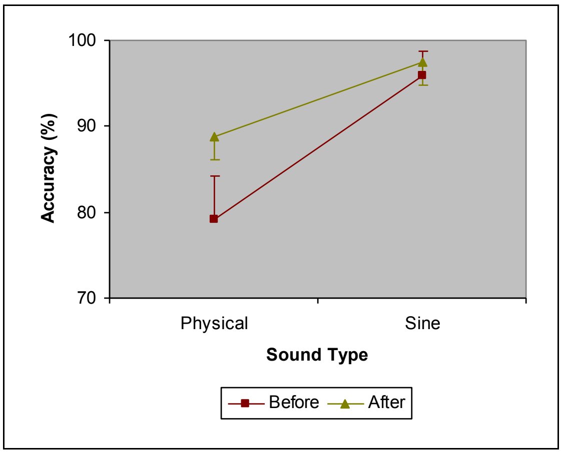


Figure 5-29: Interaction Effect of Practice and Sound Type on Accuracy

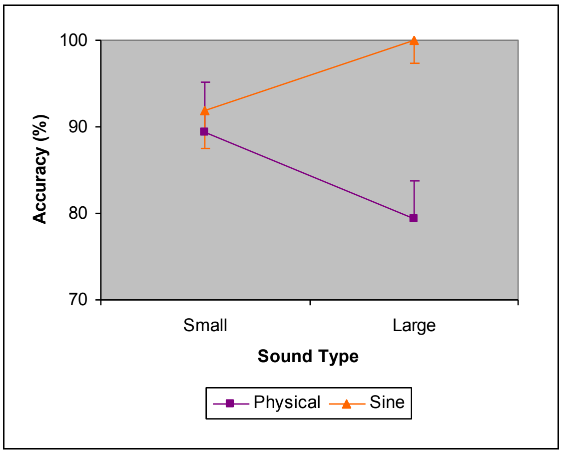


Figure 5-30: Interaction Effect of Sound Type and Magnitude Difference on Accuracy

There was an interaction effect between sound type and magnitude difference, $F(2.35,44.70)=5.18$, $p=0.01$, partial $\eta^2=0.21$. Pre-planned contrasts revealed that the decrease in accuracy for 'Physical' (compared to 'Sine') was significantly greater for large magnitude differences (compared with small), $F(1,19)=10.66$, $p<0.01$, partial $\eta^2=0.36$. This difference is quite marked (see Figure 5-30). Accuracy for large magnitudes with the physical sound are lower than those for small magnitudes, whereas with sine the accuracy is as expected (greater accuracy for large magnitude differences). All other contrasts were non-significant ($p>0.05$).

5.3.3.3 Confidence Level

5.3.3.3.1 Curve Shape

A repeated-measures ANOVA compared the effect on confidence level of practice, sound type, and magnitude difference (see Figure 5-31 and Table 5-8). This revealed that there was a main effect of Sound type, $F(2,18)=5.99$, $p=0.01$, partial $\eta^2=0.40$. Pre-planned contrasts showed that 'Cello' (86.25%, SD=23.66) had a significantly higher confidence level than 'Sine' (77.08%, SD=31.65), $F(1,9)=12.84$, $p=0.01$, partial $\eta^2=0.59$. There was no significant difference between 'Sine' and 'Physical' (83.33%, SD=23.32), $F(1,9)=0.29$, $p=0.60$, partial $\eta^2=0.03$. There was a main effect of magnitude difference, $F(1.14,10.27)=25.05$, $p<0.01$, partial $\eta^2=0.73$. Pre-planned contrasts showed that confidence when judging small (60.00% SD=31.29) magnitude differences was significantly less than for either medium (90.00%, SD=17.95) or large (96.67%, SD=8.57) differences, $F(1,9)=29.99$, $p<0.01$, partial $\eta^2=0.77$, and $F(1,9)=25.11$, $p<0.01$, partial $\eta^2=0.74$. There was no main effect of practice, $F(1,9)=0.42$, $p=0.53$, partial $\eta^2=0.04$.

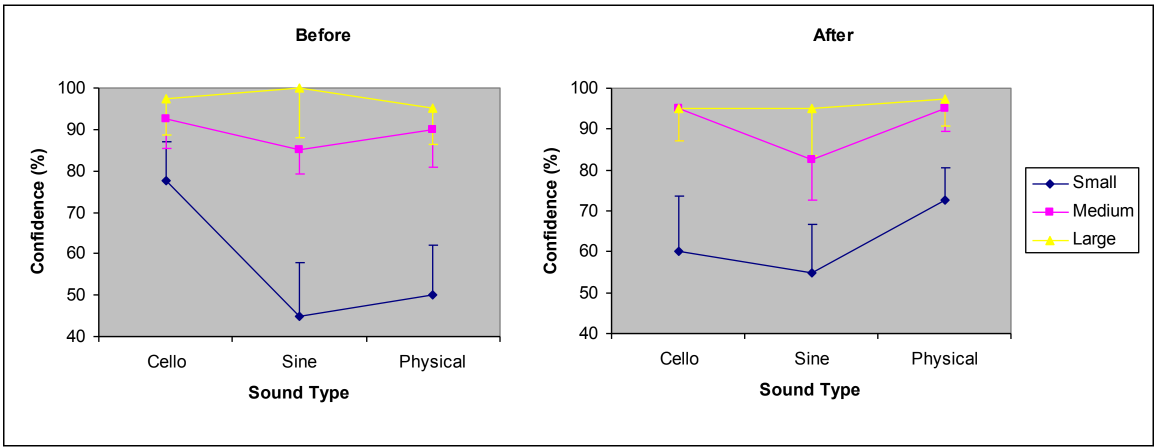


Figure 5-31: Effect of practice, sound type (curve shape), and magnitude difference on confidence

Table 5-8: Mean confidence for Sound Type, Magnitude Difference, and Practice (curve shape)

Variable	Mean	SD
Cello	86.25	23.96
Physical	83.33	23.32
Sine	77.08	31.65
Small	60.00	31.29
Medium	90.00	17.95
Large	96.67	8.57
Before Practice	81.39	27.68
After Practice	83.06	25.65
All	82.22	26.62

There was an interaction effect between sound type and magnitude difference (see Figure 5-32), $F(4,36)=4.96$, $p<0.01$, partial $\eta^2=0.36$. This indicates that sound had a different effect on confidence level dependent upon the magnitude difference. Pre-planned contrasts revealed that the increase in confidence for medium magnitude difference (compared to small) was significantly less for 'Cello' (compared to 'Sine'), $F(1,9)=13.97$, $p=0.01$, partial $\eta^2=0.61$. This was also true for large differences, $F(1,9)=13.50$, $p=0.01$, partial $\eta^2=0.60$.

There was no interaction effect between practice and either sound type or magnitude difference, $F(2,18)=0.90$, $p=0.42$, partial $\eta^2=0.09$, and $F(2,18)=0.96$, $p=0.40$, partial $\eta^2=0.10$, respectively.

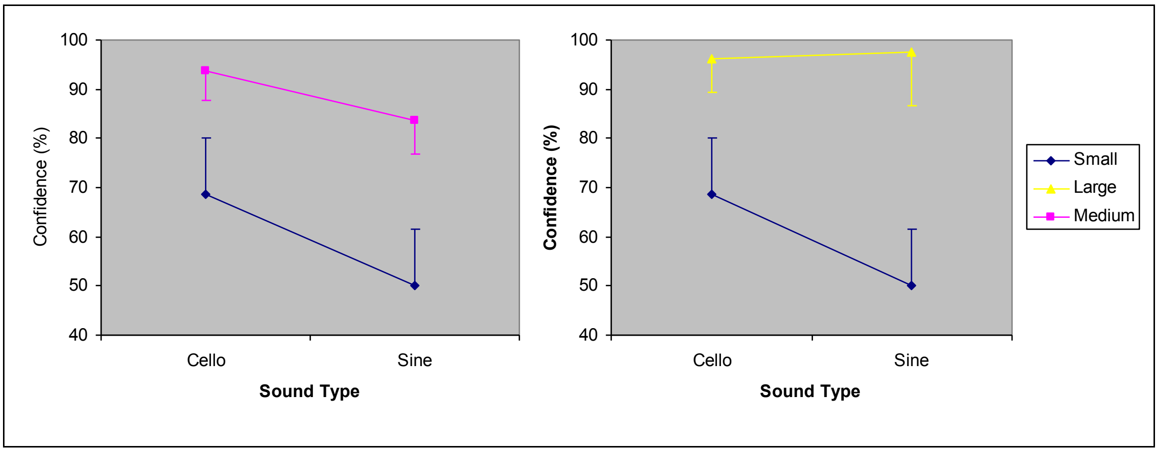


Figure 5-32: Interaction of sound type (curve shape) and magnitude difference on confidence

5.3.3.3.2 Curvature

A repeated-measures ANOVA compared the effect on confidence level of practice, sound type, and magnitude difference (see Figure 5-33 and Table 5-9). This revealed that there was a main effect of Sound type, $F(1.11,9.97)=11.74$, $p=0.01$, partial $\eta^2=0.57$. Pre-planned contrasts revealed that 'Sine' (94.17%, SD=16.83) had a significantly higher confidence level than 'Physical' (78.33%, SD=29.28), $F(1,9)=10.03$, $p=0.01$, partial $\eta^2=0.53$. There was no significant difference between 'Sine' and 'Cello' (94.58%, SD=13.09), $F(1,9)=0.07$, $p=0.80$, partial $\eta^2=0.01$. There was no main effect of practice or magnitude differences, $F(1,9)=2.40$, $p=0.16$, partial $\eta^2=0.21$, and , $F(2,18)=0.26$, $p=0.77$, partial $\eta^2=0.03$, respectively.

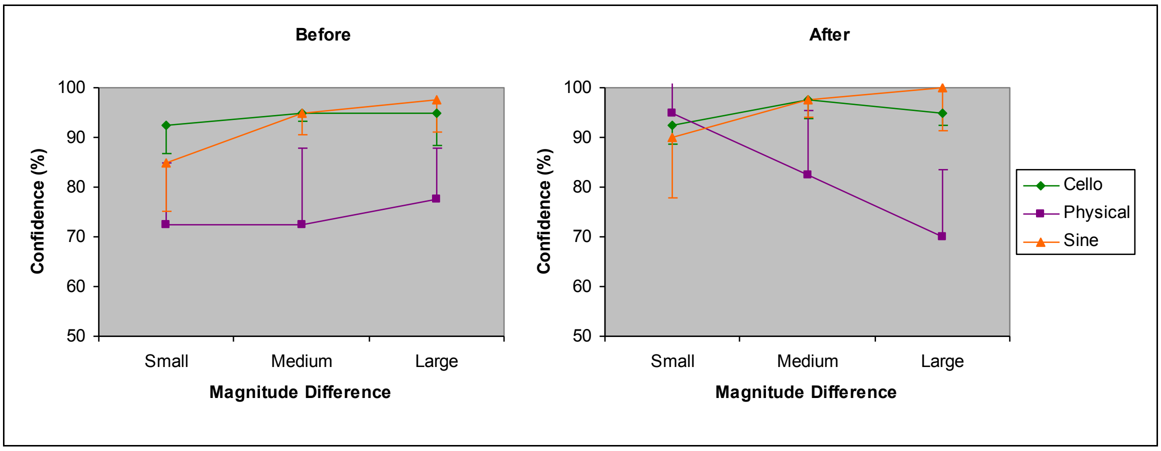


Figure 5-33: The effect of practice, sound type (curvature), and magnitude difference on confidence level

Table 5-9: Mean confidence for Sound Type, Magnitude Difference, and Practice (curvature)

Variable	Mean	SD
Cello	94.58	13.09
Physical	78.33	29.28
Sine	94.17	16.83
Small	87.92	24.57
Medium	90.00	20.17
Large	89.17	21.77
Before Practice	86.94	22.56
After Practice	91.11	21.63
All	89.03	22.14

There was an interaction effect between sound type and magnitude difference (see Figure 5-34), $F(4,36)=3.09$, $p=0.03$, partial $\eta^2=0.25$. Pre-planned contrasts revealed that the decrease in confidence for 'Physical' was significantly greater for medium and large magnitude (as opposed to small), $F(1,9)=8.31$, $p=0.02$, partial $\eta^2=0.48$, and $F(1,9)=6.18$, $p=0.03$, partial $\eta^2=0.41$, respectively.

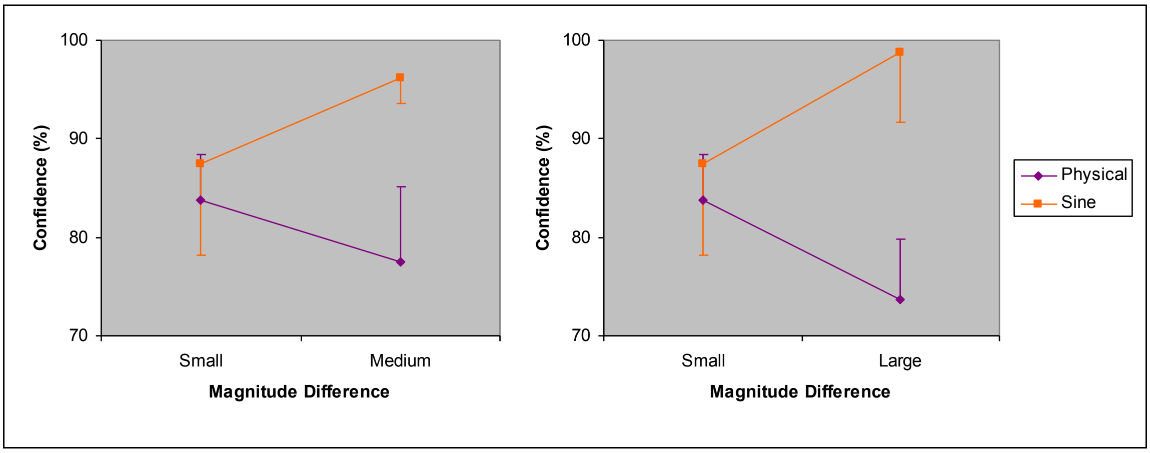


Figure 5-34: The interaction of sound type and magnitude difference on confidence level (Curvature sonification)

There was no interaction effect between practice and either sound type or magnitude difference, $F(1.33,11.96)=1.02$, $p=0.36$, partial $\eta^2=0.10$, and $F(2,18)=1.66$, $p=0.22$, partial $\eta^2=0.15$, respectively.

5.3.3.4 User Experience

5.3.3.4.1 Curve Shape

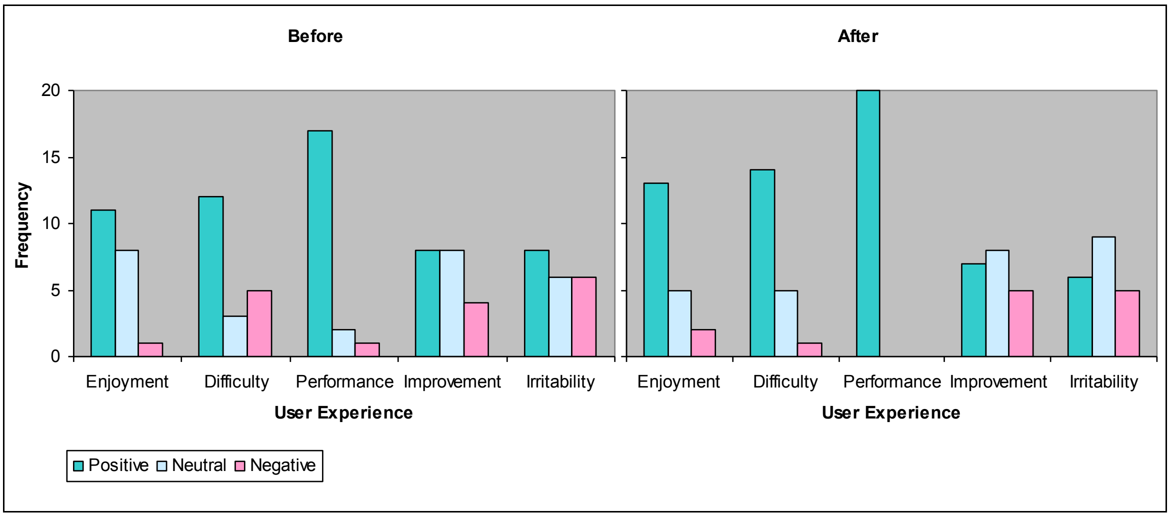


Figure 5-35: User Experience - breakdown of responses for each issue before and after practice (curve shape sonification)

Breakdowns of the frequencies for each question are given in Figure 5-35. The graphs indicate little difference between attitudes before and after practice. Five two-variable chi-squared tests compared the before and after practice frequencies for each user experience question. These confirmed that there was no significant ($p>0.05$) difference between attitudes before and after practice.

Across both practice conditions there appeared to be a positive attitude towards Enjoyment, Difficulty, and Performance, and an even spread of attitudes towards Improvement and Irritability. The data across both practice conditions were collapsed (by averaging the data), and a series of five one-factor chi-squared tests were used to analyse the difference between expected and actual frequencies for each level of attitude (positive, negative, and neutral). The results are shown in Table 5-10, and confirm that there was a positive attitude towards the first three user experience issues and an even spread of attitudes towards the latter two issues.

Table 5-10: Association of frequency with attitude (curve shape)

Issue	n	χ^2	df	Asymp. Sig.
Enjoyment	20	7.14	2	0.03
Difficulty	20	9.10	2	0.01
Performance	20	30.86	2	0.00
Improvement	20	0.86	2	0.76
Irritability	20	0.29	2	0.96

5.3.3.4.2 Curvature

Breakdowns of the frequencies for each question are given in Figure 5-36. The graphs indicate little difference between attitudes before and after practice. Five two-variable chi-squared tests compared the before and after practice frequencies for each user experience question. These confirmed that there was no significant ($p>0.05$) difference between attitudes before and after practice.

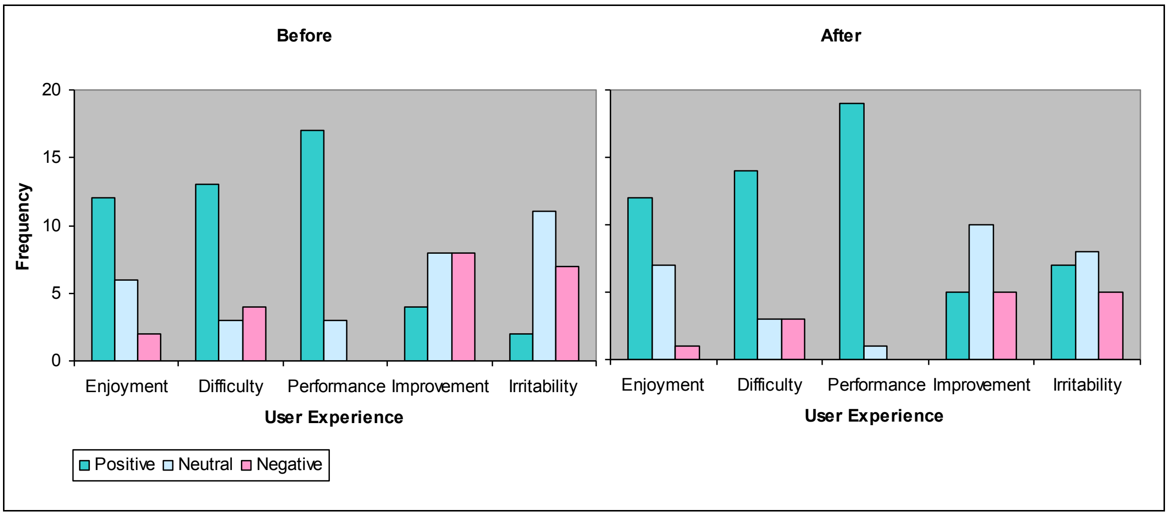


Figure 5-36: User Experience - breakdown of responses for each issue before and after practice (curvature sonification)

Across both practice conditions there appeared to be a positive attitude towards Enjoyment, Difficulty, and Performance, and an even spread of attitudes towards Improvement and Irritability. The data across both practice conditions were collapsed (by averaging the data), and a series of five one-factor chi-squared tests were used to analyses the difference between expected and actual frequencies for each level of attitude (positive, negative, and neutral). The results are shown in Table 5-11, and confirm that there was a positive attitude towards the first three user experience issues and an even spread of attitudes towards the latter two issues.

Table 5-11: Association of frequency with attitude (curvature)

Issue	n	χ^2	df	Exact Sig.
Enjoyment	20	7.14	2	0.03
Difficulty	20	10.57	2	0.01
Performance	20	12.80	2	0.00
Improvement	20	1.14	2	0.61
Irritability	20	2.00	2	0.47

5.3.3.5 Workload

NASA-TLX was used to provide an estimation of the subjective workload experienced by each participant.

5.3.3.5.1 Curve Shape

The mean workload score before practice was 50.40 (SD=14.47) and after was 42.92 (SD=11.54). A dependent t-test showed that there was a significant difference between the two scores, t(19)=3.61, p=0.01, r=0.41. This means that workload was significantly reduced after practice.

A breakdown of the sub-factors influencing workload are shown in Figure 5-37. This gives an idea of the relative contribution each factor made to the participants' sense of workload. Mental demand, performance, and effort are scored more highly than the other factors. As with the workload score, most sub-factors changed little after practice compared to before. Exceptions are mental demand which decreases from 192 (SD=128.99) before practice to 135.75 (SD=112.80) after, and Effort which reduces from 158.25 (SD=95.84) to 118.00 (SD=79.83). Dependent t-tests showed that these differences were significant, t(19)=3.01, p=0.01, r=0.32, and t(19)=2.81, p=0.01, r=0.29, respectively. All other sub-factors when tested (dependent t-tests) showed non-significant differences (p>0.05).

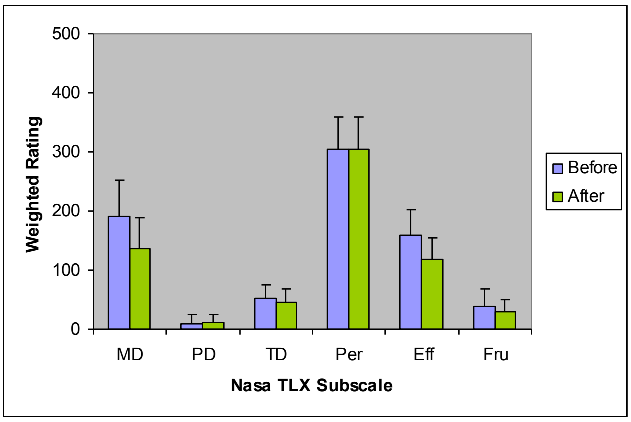


Figure 5-37: Workload Subscales

Note: Mental Demand (MD), Physical Demand (PD), Temporal Demand (TD), Performance (Per), Effort (Eff), and Frustration (Fru)

5.3.3.5.2 Curvature

The mean workload score before practice was 49.00 (SD=15.06) and after was 42.00 (SD=11.70). A dependent t-test showed that there was a significant difference between the two scores, t(19)=3.08, p=0.01, r=0.33. This means that workload was significantly reduced after practice.

A breakdown of the sub-factors influencing workload are shown in Figure 5-38. This gives an idea of the relative contribution each factor made to the participants' sense of workload. Mental demand, performance, and effort are scored more highly than the other factors. As with the workload score, most sub-factors changed little after practice compared to before. Exceptions are mental demand which decreases from 177.50 (SD=125.40) before practice to 129.50 (SD=113.12) after, and Effort which reduces from 145.00 (SD=103.16) to 105.25 (SD=88.18). Dependent t-tests showed that these differences were significant, t(19)=2.57, p=0.02, r=0.26, and t(19)=2.44, p=0.02, r=0.24, respectively. All other sub-factors when tested (dependent t-tests) showed non-significant differences (p>0.05).

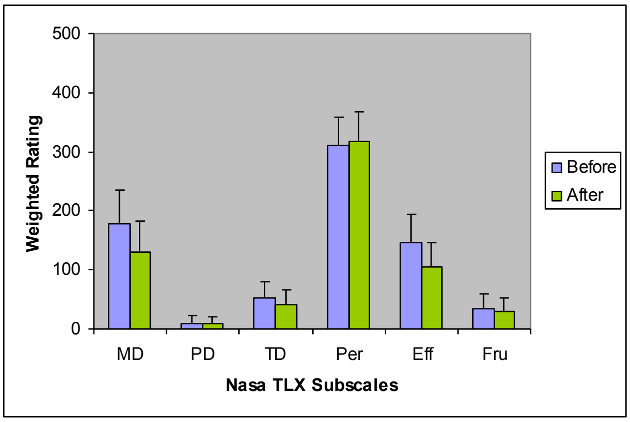


Figure 5-38: Workload Subscales

Note: Mental Demand (MD), Physical Demand (PD), Temporal Demand (TD), Performance (Per), Effort (Eff), and Frustration (Fru)

5.3.3.6 Background

Table 5-12 and Table 5-13 summarise performance measures[12] according to participant background. The results here are combined for all sounds as the primary aim of this analysis is to examine the impact of participant background on performance in general.

---

[12] Due to an error with the ePrime software confidence data were not available for all participants so it was not possible to calculate the effect of background for this.

5.3.3.6.1 Curve Shape

A one-way independent ANOVA revealed that there was no effect of background on response time, F(3,16)=0.94, p=0.44. A second one-way independent ANOVA revealed that there was a significant effect of background on accuracy, F(3,16)=3.36, p=0.04. Pre-planned post-hoc tests (Dunnett's t-test) showed that those with a 'Music' background were significantly more accurate than those with a 'General' background. There was no significant difference between the 'General' category and any other type of background, p>0.05.

Table 5-12: Performance according to participant background (curve shape)

Background	Response Time (s)		Accuracy (%)	
	Mean	SD	Mean	SD
Music	4.89	0.38	98.61	1.96
Physics/Maths	5.10	0.78	97.22	0.98
Product Design	5.17	0.78	92.22	3.62
General	5.62	0.80	89.44	9.55

5.3.3.6.2 Curvature

Two separate one-way independent ANOVAs revealed that there was no effect of background on response time or accuracy, F(3,15)=0.98, p=0.43, and F(3,16)=2.98, p=0.06, respectively.

Table 5-13: Performance according to participant background (curvature)

Background	Response Time (s)		Accuracy (%)	
	Mean	SD	Mean	SD
Music	4.50	0.33	97.50	1.81
Physics/Maths	4.91	0.92	97.22	4.05
Product Design	4.30	1.00	87.22	10.27
General	5.10	0.77	88.61	8.75

5.3.4 Discussion

5.3.4.1 The Appropriateness of Sound

The experiment sought to address the following questions in relation to the appropriateness of sound:

- Is sound an appropriate medium through which to convey information about curve magnitude? That is, can participants confidently, quickly, and accurately identify the magnitude of a curve? Does performance change with magnitude difference? Does performance change with practice?

- Is there an optimum sound type for conveying this information?
- How do sonification methods compare?

Participants were able to quickly, accurately and confidently assess the relative magnitudes of curves through the use of sound (see Table 5-14 for summary of results). The level of performance was high and there was little difference between using 'Curve Shape' or 'Curvature' sonifications. This suggests that the use of sound was fairly intuitive and that participants could easily perceive differences in curve magnitude.

Whilst performance was high it was significantly better after participants had undertaken a few hours practice. However, the improvements were relatively small; a one second decrease in response time for both sonification methods and an increase in accuracy of 3% for 'Curve Shape' and 4% for 'Curvature'. This indicates that performance was improved through acquainting users with sound exploration, but that the performance improvements were not substantial and did not boost confidence. It should also be noted that the increase in performance after practice was similar, regardless of the sound type used. However there was one minor exception. With 'Curvature', the effect of practice was significantly greater with the 'Physical' sound type. So whilst there was an improvement in performance with all sounds, more improvement was shown with those sounds that had initially been less accurate to use.

Table 5-14: Comparison of Performance for Curve Shape and Curvature Sonifications

Performance Measure	Curve Shape				Curvature
	Small	Medium	Large	All Magnitudes	All Magnitudes
Response Time	5.7	5	4.8	5.2	4.7
Accuracy	87	97	99	94	93
Confidence	60	90	97	82	89

Note: There was shown to be no effect of magnitude difference on performance for curvature so only the mean result for all magnitudes is given. For 'Curve Shape', small magnitude differences were found to be significantly different to both medium and large differences for all performance measures.

The results so far discussed show little difference in performance with either sonification method. However there was disparity in performance between the sonification methods for various magnitude differences. With 'Curve Shape' participants were significantly slower, less accurate, and less confident at judging small magnitude differences than either medium or large differences. With 'Curvature' there were no such differences, and this indicates that small differences in magnitude were as easy to judge as medium or large. It is likely that this difference was due to the perceptual qualities of the two sonification methods. The 'Curve Shape' sonification provided a graduated rising then falling tone as the curve was traversed. The differences in audio feedback for the two curves would be subtle at small magnitude differences, and may only have become more obvious as the difference between the curves increased, i.e. when there were medium or large differences. The nature of the ‘Curvature’ sonification was different. With this a continuous tone was produced relative to the curvature value of the examined curve. As humans can perceive differences in tone of as little as 0.1% (Wolfe, Kluender et al. 2006), then even when two very similar curves are sonified the difference should be relatively easy to distinguish, as was the case here.

With ‘Curve Shape’ the type of sound used to explore curve magnitude was found to have a significant effect on participant’s confidence. Confidence was significantly higher when using ‘Cello’ (86%) compared to when ‘Sine’ (78%) was used. There was found to be no difference in confidence between ‘Sine’ and ‘Physical’ (83%). It is difficult to know what particular quality of the ‘Cello’ sound increased participant confidence over that of ‘Sine’. Both sound types provide a relatively clear signal and their performance levels for accuracy and response times were similar. The characteristic difference between them is the ‘Cello’ sound’s harmonic frequencies, and this may have played a role in boosting confidence (perhaps this pleasantness is somehow reassuring). However, it is difficult to suggest why ‘Sine’ did not give rise to more confidence than ‘Physical’ given the perceptual problems already discussed in relation to the latter. The behaviour here was therefore confusing, and so further work would be needed in order to assess the relationship between sound and confidence in this context. However in relationship to response time (M=5s) and accuracy (M=94%) there was no significant effect of sound type and all enabled high levels of performance.

The results for ‘Curvature’ sonification are more understandable relative to the characteristics of the different sound types. For all performance measures (response time, accuracy, confidence), ‘Sine’ (4.5s, 97%, 94%) was significantly better than ‘Physical’ (5s, 84%, 78%), but enabled the same level of performance as ‘Cello’ (4.6s, 97%, 95%). Given the relative clarity of ‘Sine’ compared to ‘Physical’ it is likely that participants found it easier to perceive and so performance was significantly improved. Whereas ‘Cello’ enabled a similar level of performance and indicates that this sound type was equally as well perceived as a simple sine wave.

In comparing the two sonification methods there was a difference between the ways in which sound types affected performance. With ‘Curve Shape’ performance was not dependent upon the sound type used (discounting confidence), whereas for ‘Curvature’ the simpler sounds enabled a higher level of performance than the more complex sound of ‘Physical’. The differences observed might be explained by the relative difficulty of the task undertaken using the different sonification methods. In essence, when comparing magnitudes using the ‘Curvature’ sonification it was required that two tones should be differentiated. This, as explained earlier, is a relatively easy perceptual task and as such would involve the comparison of two tones. Therefore, the noise associated with the ‘Physical’ sound may have made perception of the tone more difficult and given rise to longer response times, less accuracy, and less confidence. Conversely, the relative difficulty of the ‘Curve Shape’ sonification may have masked any small differences between the sound types, and so no significant differences were observed. The fact that the effect of sound types on performance was different for each sonification method makes it more difficult to identify an optimum choice from this experiment. Whilst it is difficult to choose between types it is clear that the less complex ‘Sine’ and ‘Cello’ did outperform ‘Physical’ with ‘Curvature’ sonification.

What might also help decide between the sound types is a consideration of the interaction effect observed between practice and magnitude difference. For the ‘Curve Shape’ sonification there was no interaction between sound type and practice on performance. This meant that there was a similar level of improvement in performance for each sound type after practice. This was also seen for response time and confidence with ‘Curvature’. However, this was not the case for accuracy. Here it was seen that performance with the ‘Physical’ sound improved more than that with the ‘Sine’. This is not to say that performance with ‘Physical’ was greater than ‘Sine’ after practice but that there was more to improve because the initial performance with ‘Physical’ was less accurate. This coheres with the main effect for sound type, and points to the relatively poorer suitability for the physical sound in this task.

For sound types it was observed that there was an interaction with magnitude differences. That is, the performance differences seen between small and medium or large magnitude differences were different dependent upon the sound type used. For curve shape, the decrease in response time for large magnitudes (as opposed to small) was greater for ‘Sine’ than when either ‘Cello’ or ‘Physical’ were used. Similarly the increase in confidence for medium and large magnitude differences (as opposed to small) was greater for ‘Sine’ than for ‘Cello’. There was no interaction of sound and magnitude in the case of accuracy.

With ‘Curvature’ it was seen that the increased response time, decreased accuracy, and decreased confidence associated with the ‘Physical’ sound was greater (compare to

'Sine') when assessing medium and large magnitude differences (compared to small). Effectively this means that as the difference between tones increased performance with the 'Physical' sound became poorer. This effect is not understood, but it does further indicate the unsuitability of the 'Physical' sound relative to 'Sine'.

5.3.4.2 User Experience and Workload

The experiment sought to address the following questions in relation to the user experience and workload:

- What attitude do participants express towards their experience of using sound? Does this change with practice?
- What level of workload do participants experience? Does this change with practice?

The attitude toward sound was similar regardless of sonification method or level of practice. It was found that the attitudes expressed towards 'Enjoyment', 'Performance', and 'Improvement' were significantly positive. This contrasted to the attitudes expressed towards 'Difficulty' and 'Irritability' which were found to be more evenly spread. This meant that equal numbers of people were positive as negative, and that a similar number expressed a neutral attitude. Whilst this meant that around two-thirds of participants did not feel negatively, a substantial proportion of those that undertook the task felt that it was difficult and found the sounds irritating.

User experience therefore seems to be somewhat mixed. Whilst there was little disagreement with the view that the task was enjoyable and a belief by many that they performed accurately and improved overtime, there were a proportion of participants who were irritated by the sounds and found the task difficult to complete. These attitudes did not change with practice, and so greater familiarity with the task did not help overcome some of these negative aspects.

The level of workload experience by participants was moderate with a mean NASA-TLX score of 47 for 'Curve Shape' and 45 for 'Curvature'. The NASA-TLX score ranges from 0 to 100, so the score here would indicate that participants were operating fairly comfortably within the middle of the range. It is interesting to note that despite the differences in the sonification methods the workloads remain similar. There was a significant effect of practice on the level of workload experienced by participants. Before practice the scores were 50 and 49, and after practice 43 and 42, for 'Curve Shape' and

'Curvature' respectively. An examination of sub-factors showed that there was a significant reduction in Mental Demand and Effort following practice for both sonification methods.

5.3.4.3 User Background

Within the experimental design background had not been considered as a separate factor, and steps were taken to ensure that there was a range of aptitudes in the participants undertaking the task. During the course of the experiment it was observed that the 'General' category seemed less able in performing the task. Participants within this category had no formal musical training and this may have affected their ability to undertake the task. Whereas there was a high-level of training within the 'Music' category as these had been selected from students undertaking a music degree, and there was at least a moderate level of musical training in both the 'Maths/Physics' and 'Product Design' categories as these participants had indicated undertaking graded exams (between grades 2 to 6). For this reason it was decided to explore the data to identify any performance differences between the 'General' group and the other three categories.

For 'Curve Shape' it was found that there was no significant difference between categories for response time, but that there was a significant difference in accuracy between those with a general background (89%) and those with a musical background (99%). This indicated that those with degree-level musical training were more skilled in relating the variation in sound to curve shape and then to compare differences in variations across shapes to identify changes in magnitude. However, there was not found to be any significant difference between those with a general background and either 'Product Design' (92%) or 'Maths/Physics' (97%) backgrounds. So whilst those with no musical training were around 10% less accurate than those with high-level musical skills, there was no significant difference between them and those with an intermediate level of musical training. There was however no such difference between backgrounds when the 'Curvature' sonification was used. This would suggest that differences between tones were more obvious to detect and required a less practiced listening ability. To this extent the 'Curvature' sonification may be more suited to situations where there is either limited time to train participants or where an interface may be used by a general audience. Even so, the accuracy levels with 'Curve Shape', even for the non-musically trained 'General' category, were of such a level that it is doubtful that background is of much concern.

5.4 Chapter Summary

This chapter detailed work undertaken for Study 2. This study investigated the appropriateness of using sound to convey curve shape and curvature information. Experiment 1 examined how well participants were able to perceive the orientation of a curve through sound, and Experiment 2 examined how well they were able to perceive curve magnitude. Three different types of sound were used to convey the information; a sine wave ('Sine'), a harmonic sound ('Cello'), and a complex sound ('Physical'). Participants were able to perform tasks in both experiments quickly, accurately and confidently and so proved the appropriateness of sound for conveying these aspects of a curve. The type of sound used had little impact although both 'Cello' and 'Sine' were more accurate than the 'Physical' sound. It was thought that the extra 'noise' associated with this more complex sound led to increased participant error. It was also noted that whilst 'Cello' and 'Sine' had similar levels of performance, participant confidence was significantly higher with the 'Cello' sound. There was some improvement in performance after practice, although this was relatively small. The effect of practice was also reflected in the fact that those with an advanced musical background performed more accurately than those without any musical training ('General'). However, there was no difference between those from the general background and those with a moderate level of musical training ('Maths/Physics'). It was therefore thought that background was of limited influence, and performance for participants was high in any case. In Experiment 2 two sonification methods were used; 'Curve Shape' and 'Curvature'. Performance using both these was at similar levels; however with 'Curve Shape' participants were significantly slower, less accurate, and less confident at judging small magnitude differences than either medium or large differences. Finally, across both experiments user experience was mixed. Whilst use of sound to explore curve properties was found to be generally enjoyable a significant proportion of participants found the sounds irritating and the tasks were considered to be difficult. However, in relation to this later point, workload (NASA-TLX) was found to be moderate and participants were operating comfortably in the middle of the range. Overall, sound was found to be an appropriate medium through which to convey curve orientation and magnitude.

5.5 Related Chapters

The work here considered simplified tasks in order to examine fundamental issues. The use of sound is explored with a more complex task in Chapter 6. The effect of the 'Cello' sound in combination with other modalities is considered in Chapter 7. The findings here also contribute to the general discussion (see Chapter 10).

Chapter 6: Sound in an Applied Context

6.1 Introduction and Rationale

The work here was part of a study examining the suitability of sound for communicating curve shape and curvature information (see Chapter 5 section 5.1.1 for study structure). The previous chapter described a series of experiments that looked at the fundamental issue of communicating orientation and magnitude of a curve. The results from these were encouraging and indicated that sound could be used to convey these fundamental aspects. However, it was recognised that these experiments were quite simplistic and whilst dealing well with the fundamental issues may not necessarily engage with the complexity experienced when judging curves in reality. Therefore, the purpose of the experiments reported here was to explore performance through a more context-relevant task.

6.1.1 The Nature of a Curve in Reality

Unlike the curves in the previous experiments which had a fixed curvature, most curves in reality have a changing curvature. In fact it is these changes in curvature that make a line curvy; high curvature produces tight curves and low curvature gives more open curves, until at zero curvature you have a straight line. It is important to note that we cannot see 'curvature' as it is a mathematical construct, however we can experience it second hand as we observe changes in the curve of a shape. Even so, it is difficult to fully appreciate these changes in curvature as we can only imagine them as we see or feel fluctuations in a curve. As mentioned previously (see Chapter 1 section 1.1.3), for product designers these changes are of great importance, so much so that they have devised a means of visualising curvature through the use of 'porcupine' analysis. As shown in Figure 6-1, the curvature at a given point along the curve is visualised by a line; short lines for low curvatures, and long lines for high curvatures. This gives an immediate sense of the rise and falls in

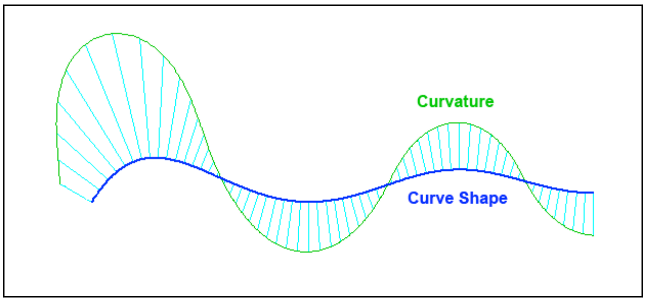


Figure 6-1: Curvature 'porcupine' analysis

curvature along a curve (as shown by the green line). In essence what this illustration shows is the two ways in which we can understand a curve; directly through curve shape (blue line) and abstractly through curvature (green line). This also highlights a potential area of difficulty; when visualised the abstraction of curvature becomes yet another curve shape which is relatively easy to understand. However, when sonified this immediacy may be lost and the mental translation may cause excessive workload.

The curves illustrated in Figure 6-1 are more complicated than the simple arc shapes that were used in the previous experiments. They also provide validity to the real-world context, and the approach adopted by the SATIN project which was to sonify rather than visualise curvature data. The experiments in this chapter will seek to understand how well sound conveys these more complex curve shapes and their associated curvatures, and whether curvature sonification increases workload unduly.

6.1.2 Aims

Six research questions were posed at the start of this study:

1. Is sound an appropriate medium through which to convey curve shape and curvature information? That is, can participants confidently, quickly, and accurately identify the correct curve from auditory feedback?
2. Is there an optimum sound type for conveying this information?
3. How do sonification methods compare?
4. What attitude do participants express towards their experience of using sound?
5. What level of workload do participants experience?
6. Are there any specific interaction strategies employed by participants in completing tasks?

6.2 Method (Study 2 – Experiment 3)

6.2.1 Participants

These were the same as in experiment 2 (see Chapter 4 section 5.2.2.1).

6.2.2 Equipment and Setup

The experiment took place in a dedicated usability lab. Participants were seated in front of a laptop computer on which the experiment would be run. The laptop's screen displayed multiple curved shape images that could be interacted with through a Wacom tablet. This interaction produced either the curved shape or curvature sound which was delivered through a pair of Behringer MS15 speakers positioned either side of the laptop. A second computer was used to run the experiment. The installed E-Prime software randomly generated the trial conditions (sound type, curve set, and Kinetic mode) and allowed the facilitator to record participant responses.

Figure 6-2: Experimental setup showing laptop with curve shape images and Wacom tablet for interacting with images to produce curve shape or curvature sonifications.

6.2.3 Procedure

The experiment was undertaken in two sessions corresponding to the different sonification methods; Curve Shape and Curvature. These were undertaken on separate days, with a week between the two sessions (see Chapter 5 Figure 5-1). The procedure for each of the experimental sessions was the same.

The participant read through a description of the experiment and was given the opportunity to ask questions about this by the facilitator. The equipment was demonstrated to the participant by the facilitator, who also explained the interaction between the tablet, graphical screen image, and sound display. They were then given five minutes to work through a number of examples. This ensured that they would become familiar with interaction via the Wacom tablet and the relationship of visual and auditory displays before commencing the experiment.

The experiment consisted of a series of 75 trials. These were delivered in sets of 25 trials with a few minutes break between each set. Each set took approximately 20 minutes to complete. Following each set the participant was asked to complete a NASA-TLX workload assessment sheet, they also completed a user experience questionnaire following the final set of trials. At the beginning of each trial the participant was instructed to configure the sound type (Cello, Physical, or Sine), curve set and the state of the kinetic module (on or off). Once configuration was completed the participant was presented with the multiple choice screen showing four curve images and they were asked to begin (response time was measured from this point until they gave an answer or they had timed out). The participant was given up to one minute to judge which of the curve shape images was related to the sound produced as they interacted with them via the Wacom tablet. Having given an answer they were then asked if they had 'high' (sure) or 'low' (not sure or don't know) confidence in this judgement. This sequence repeated until all trials were completed. All documentation used to support this study can be found on the accompanying CD as detailed in Appendix G.

6.2.4 Task Object

The purpose of the task was for the participant to correctly relate a curve to its sonified shape or curvature (see 6.2.5 for details of the sonification method). To this end the participants were presented with four curves (A to D) of which only one was related to the sound heard during interaction. The participants interacted with the curves via the Wacom tablet. As they moved the stylus from left to right across the tablet, a pointer (red circle) on each of the curves also moved from left to right tracing a path across the curve. The interaction was simultaneous for all of the curves, and curves were not interacted with separately.

For each trial the participant was presented with one correct curve, i.e. the one from which the sonified sound was produced, and three variant curves. For the curve shape experiment variants were generated by mirroring, amplifying, and shifting the correct curve. The correct curve plus its three variants went to make up a curve set. In all there were 15 curve sets which were presented once for each sound type giving a total of 75 randomised trials. For the curvature experiment correct and variant curves were generated slightly differently.

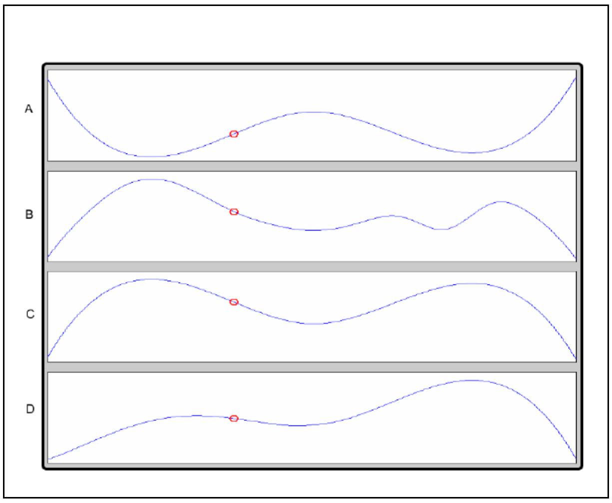


Figure 6-3: Screen presenting multiple choice options.

The sound produced as the participant interacts through the Wacom tablet is generated from only one of the curves displayed; they must decide which one.

Because of the difficulty of ensuring viable curvatures from randomly adjusted curves, the correct curve could not be used to generate its variants. Instead a pool of 20 curves (and associated curvature data) was generated. From these, five were selected as correct curves and 3 variants were randomly assigned from the remaining 15 curves to produce 5 curve sets (A1 to E1). This process was repeated a further two times to give a further 10 curve sets (A2 to E2, and A3 to E3). As with the curve shape experiment each curve set was presented once for each sound type, giving a total of 75 randomised trials.

The curve images, curve shape and curvature data were all developed by colleagues from the Technische Universiteit Eindhoven. They integrated these into a max/msp deliverable which presented the selected curve sets and interactively produced the auditory sonifications of curve shape and curvature data. This was presented to the participants on the laptop (as described in 6.2.2).

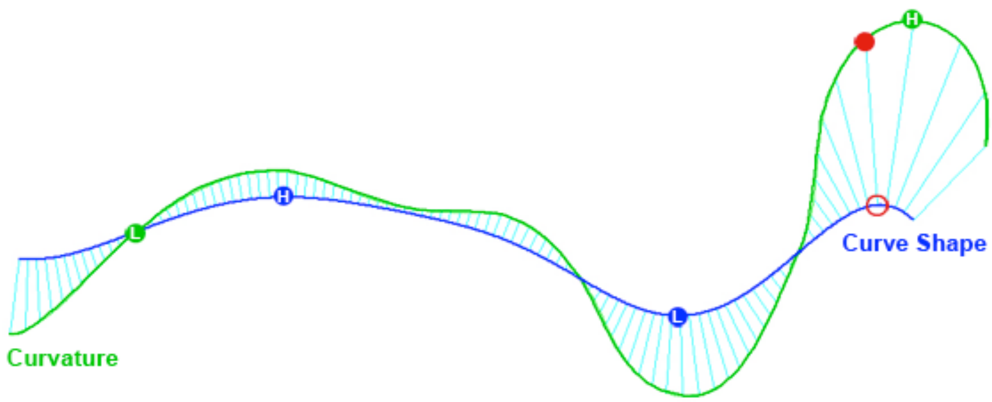
6.2.5 Sonification

6.2.5.1 Sound Types

The sounds were derived from those used in the SATIN prototype evaluations (see Chapter 1 section 1.1.4) and were designed by colleagues from Technische Universiteit Eindhoven. Five different types of sound were used for this experiment; Cello, Cello Kinetic, Physical, Physical Kinetic, and Sine. These are the sounds used in Experiment 2, and a description of Cello, Physical, and Sine can be found in Chapter 5 section 5.2.2.5.1. The two additional sounds were made by the addition of a ‘Kinetic’ module to the basic Cello and Physical sounds. The Kinetic module was applied to the sound so that it varied dependent upon the speed and pressure of the interaction with the Wacom tablet. The effect was somewhat akin to that experienced when touching a surface, say the top of a desk. The sound changes depending on how hard we press on to it or how fast we move across it.

6.2.5.2 Sound Mapping

Two mappings were used in this experiment; curve shape and curvature. For both mappings the frequency ranges from 100Hz to 400Hz, which is approximately a two octave range. For the curve shape sonification the minimum frequency is mapped to the lowest point of the curve, and the highest frequency is mapped to the highest point. This is illustrated in Figure 6-4, so that as the curve is explored the sound heard rises and falls dependent upon the position along the line of the curve. For the curvature sonification the mapping is similar; the minimum point has minimum frequency and the maximum point has maximum frequency. However, this time the sonification relates to the curvature value at a given point (see Figure 6-4 for explanation of how curvature relates to curve shape). So, as illustrated in Figure 6-4, as the curve is explored the sound rises and falls dependent upon the curvature value at a given point along the line of the curve.



The curve shape sonification was produced as the participant explored along the line of the displayed curve (blue line) with the Wacom tablet. The maximum frequency (400Hz) was heard at the highest position of the curve (blue H), and the minimum frequency (100Hz) was heard at the lowest position of the curve (blue L). The curvature sonification was again produced as the participant explored along the displayed curve (blue), however what was heard was the curvature value (green line) at that point. So if the curve was touched at the point marked by the red circle, then what was heard was the curvature value marked by the red dot. The maximum frequency (400Hz) was heard at the point of highest curvature (green H), and the minimum frequency (100Hz) was heard at the point of lowest curvature (green L). For the curvature sonification what was heard is very different to what was seen as illustrated by the blue and green lines above.

Figure 6-4: Curve Shape and Curvature Sonification

For a more technical description of the way the sound was mapped to the curve see the paper co-authored with Shelley et al. (2009).

6.2.6 Design and Hypotheses

The experiment was a repeated-measures design with one independent variable: sound type (Cello, Cello Kinetic, Physical, Physical Kinetic, Sine). The effect of this would be measured on the dependent variables: response time, accuracy, and confidence. In order to answer the research questions posed at the start (see 6.1.2) of this experiment, the following hypotheses were examined:

6.2.6.1 Performance (H1-H3)

H1-H3: There is an effect of sound type on response time; accuracy; confidence.

6.2.6.2 User experience Issues (H4-H8)

H4-H8: There is a significant difference between expected and observed attitudes towards enjoyment; difficulty; performance; improvement; irritability.

6.2.6.3 Workload (H9)

H9: There is an effect of presentation set on workload score.

6.2.7 Data Analysis

The ePrime data were collated for analysis in SPSS. In some cases variables were found to have a non-normal distribution; either positive or negative skew, which is usual for these types of data. This was not a problem as ANOVAs are considered to be a robust method of statistical analysis (Davies 1956; Field 2009)(see Appendix F for discussion of this). A number of pre-planned contrasts were used in the analysis of these data. All sound types were contrasted against Sine as this was considered to be the optimum sound.

6.3 Results (Study 2 – Experiment 3)

This section details all results in detail, a summary of the results in relation to the hypotheses can be found in Appendix D.

6.3.1 Performance

6.3.1.1 Response Time

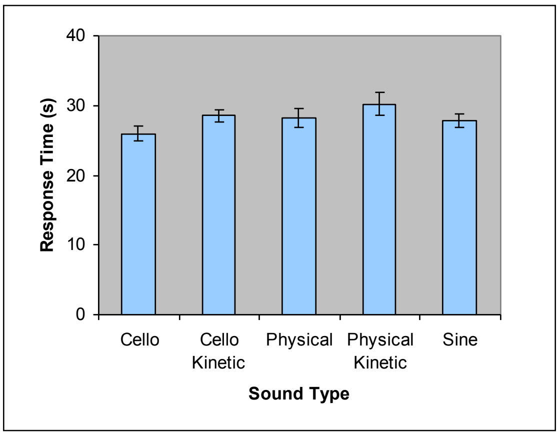


Figure 6-5: Effect of sound type (curve shape) on response time

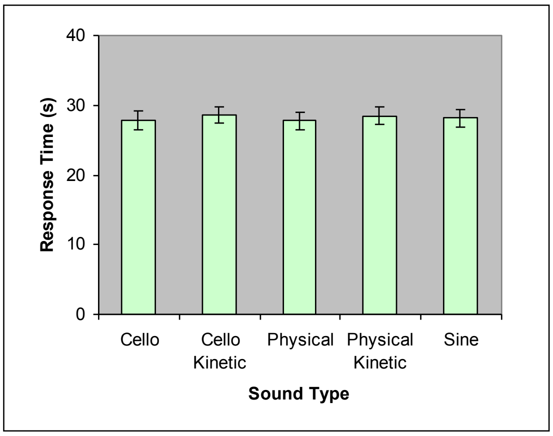


Figure 6-6: Effect of sound type (curvature) on response time

Note: Error bars represent the 95% Confidence Interval of the mean, and do so throughout.

6.3.1.1.1 Curve Shape

For ‘Curve Shape’ the mean response time was 28.17s (SD=7.90). The effect of sound type on response time can be seen in Figure 6-5. A repeated-measures ANOVA revealed that there was a significant effect of sound type on response time, $F(4,76) = 5.72$, $p<0.01$, partial $\eta^2=0.23$. Pre-planned contrasts showed that Sine (28.17s, SD=7.90) was significantly slower than Cello, $F(1,19) = 8.61$, $p=0.01$, partial $\eta^2=0.31$. There was no significant difference between Sine and any other sound type ($p>0.05$).

6.3.1.1.2 Curvature

For ‘Curvature’ the mean response time was 28.17s (SD=6.19). The effect of sound type on response time can be seen in Figure 6-6. A repeated-measures ANOVA revealed that there was no significant effect of sound type on response time, $F(4,76) = 0.27$, $p=0.90$, partial $\eta^2=0.01$.

6.3.1.2 Accuracy

6.3.1.2.1 Curve Shape

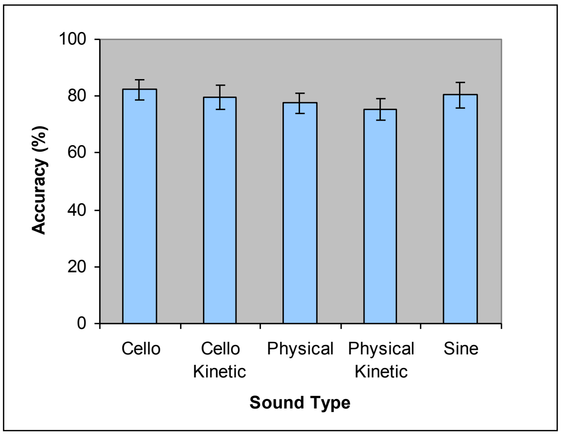


Figure 6-7: Effect of sound type (curve shape) on accuracy

For ‘Curve Shape’ the mean accuracy was 79.07% (SD=18.42). The effect of different sound types on accuracy can be seen in Figure 6-7. A repeated-measures ANOVA revealed that there was no significant effect of sound on accuracy, $F(4,76)=1.59$, $p=0.19$, partial $\eta^2=0.07$.

6.3.1.2.2 Curvature

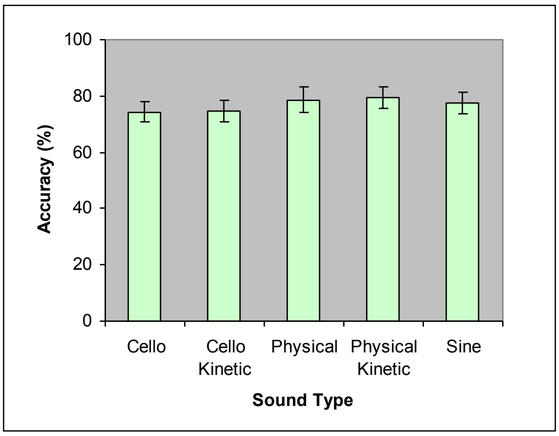


Figure 6-8: Effect of sound type (curvature) on accuracy

For ‘Curvature’ the mean accuracy was 76.93% (SD=17.01). The effect of different sound types on accuracy can be seen in Figure 6-8. A repeated-measures ANOVA revealed that there was no significant effect of sound on accuracy, F(4, 76)=1.23, p=0.31, partial η^2=0.06.

6.3.1.3 Confidence

6.3.1.3.1 Curve Shape

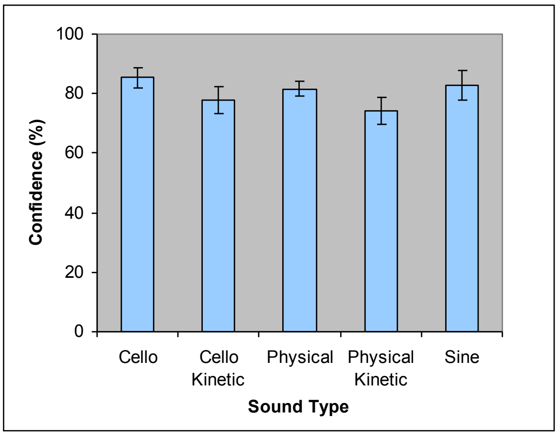


Figure 6-9: Effect of sound type (curve shape) on confidence

For ‘Curve Shape’ the mean confidence was 80.40% (SD=16.71). The effect of different sound types on confidence can be seen in Figure 6-9. A repeated-measures ANOVA revealed that there was a significant effect of sound type, F(2.54,48.25)=3.97, p=0.02, partial η^2=0.17. Pre-planned contrasts showed that there was a significant difference in confidence between Sine (83.00%, SD=11.74) and Physical Kinetic, F(1,19)=4.65, p=0.04, partial η^2=0.20. There was no significant difference between Sine and any other sound type (p>0.05).

6.3.1.3.2 Curvature

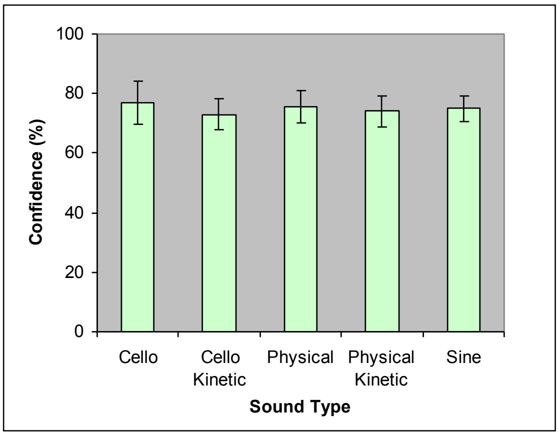


Figure 6-10: Effect of sound type (curvature) on accuracy

For 'Curvature' the mean confidence was 74.93% (SD=18.23). The effect of different sound types on accuracy can be seen in Figure 6-10. A repeated-measures ANOVA revealed that there was no significant effect of sound on confidence, F(4, 76)=0.26, p=0.90, partial η^2=0.01.

6.3.2 User Experience

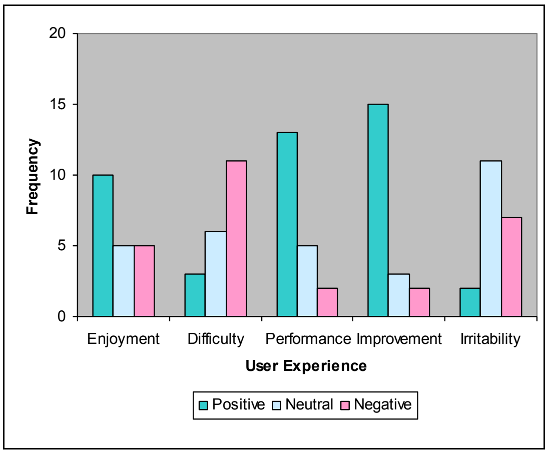


Figure 6-11: User Experience - breakdown of responses for each issue (curve shape)

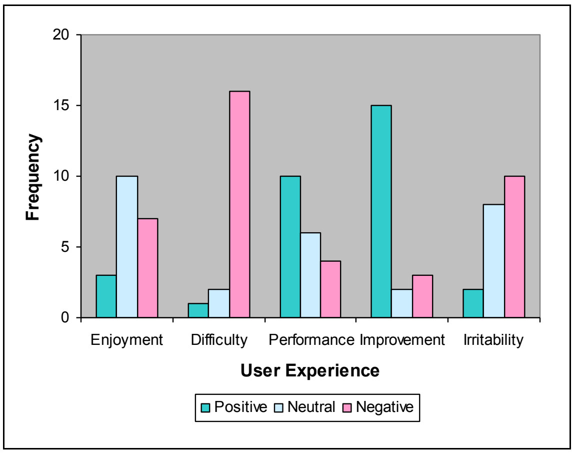


Figure 6-12: User Experience - breakdown of responses for each issue (curvature)

6.3.2.1 Curve Shape

Breakdowns of the frequencies for each user experience issue are given in Figure 6-11. A series of five one-factor chi-squared tests were used to analyses the difference between expected and actual frequencies for each level of attitude (positive, negative, and neutral). The results are shown in Table 5-11, and reveal that there was a positive attitude towards 'Performance' and 'Improvement', and no significant distribution of attitudes towards 'Enjoyment', 'Difficulty', and 'Irritability'.

Table 6-1: Association of frequency with attitude (curve shape)

Issue	n	χ^2	df	Exact. Sig.
Enjoyment	20	2.5	2	0.33
Difficulty	20	4.9	2	0.10
Performance	20	9.70	2	0.01
Improvement	20	15.70	2	0.00
Irritability	20	6.1	2	0.06

6.3.2.2 Curvature

Breakdowns of the frequencies for each question are given in Figure 6-12. A series of five one-factor chi-squared tests were used to analyse the difference between expected and actual frequencies for each level of attitude (positive, negative, and neutral). The results are shown in Table 6-2, and reveal that there was a positive attitude towards 'Improvement', a negative attitude towards 'Difficulty', and no significant distribution of attitudes towards 'Enjoyment', 'Performance', and 'Irritability'.

Table 6-2: Association of frequency with attitude (curvature)

Issue	n	χ^2	df	Exact. Sig.
Enjoyment	20	3.7	2	0.18
Difficulty	20	22.21	2	0.00
Performance	20	2.8	2	0.29
Improvement	20	15.70	2	0.00
Irritability	20	5.2	2	0.07

6.3.3 Workload

6.3.3.1 Curve Shape

NASA-TLX was used to provide an estimation of the subjective workload experienced by each participant. The measure was applied after 25, 50, and 75 presentations. A workload score was calculated for each of these as shown in Figure 6-13. A dependent

ANOVA confirmed that there was no significant effect of presentation set on workload, $F(1.32,25.05)=0.83$, $p=0.40$, partial $\eta^2=0.04$.

A breakdown of the sub-factors influencing workload are shown in Figure 6-14. This gives an idea of the relative contribution each factor made to the participants' sense of workload. Mental demand, performance, and effort are scored more highly than the other factors and so contribute more to the overall workload score. The other factors seem to have had little effect, and particularly 'Physical Demand' which is hardly apparent. The mean workload score across all presentations was 57.46 (SD=11.28).

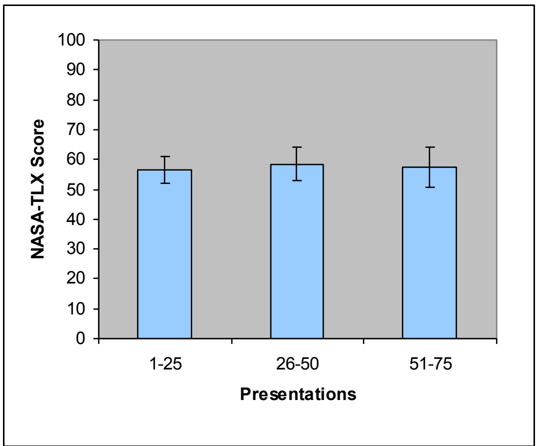


Figure 6-13: NASA TLX scores for each set of 25 presentations (curve shape).

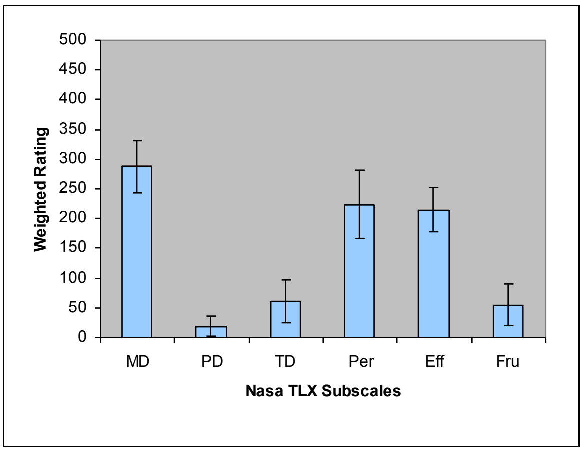


Figure 6-14: Workload Subscales for all sounds (curve shape)

Note: Mental Demand (MD), Physical Demand (PD), Temporal Demand (TD), Performance (Per), Effort (Eff), and Frustration (Fru)

6.3.3.2 Curvature

A workload score was calculated for each set of presentations as shown in Figure 6-15. This indicated that the workload experienced was consistent throughout the experimental sessions. A dependent ANOVA confirmed that there was no significant effect of presentation set on workload, $F(1.44,27.48)=0.17$, $p=0.84$, partial $\eta^2=0.01$.

A breakdown of the sub-factors influencing workload are shown in Figure 6-16. This gives an idea of the relative contribution each factor made to the participants' sense of workload. 'Mental Demand', 'Performance', and 'Effort' are scored more highly than the other factors and so contribute more to the overall workload score. The other factors seem to have had little effect, and particularly 'Physical Demand' which is hardly apparent. The mean workload score across all presentations was 56.41 (SD=8.81).

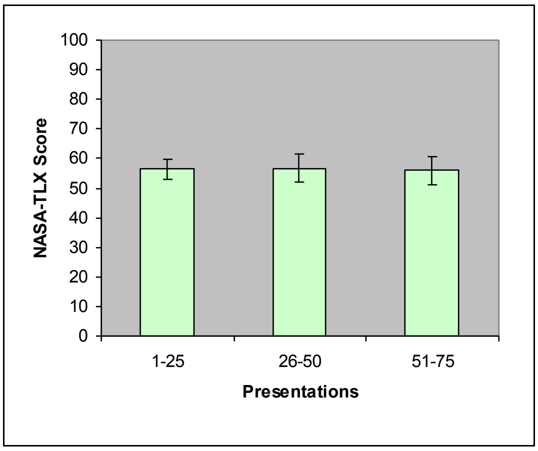


Figure 6-15: NASA TLX scores for each set of 25 presentations (curvature).

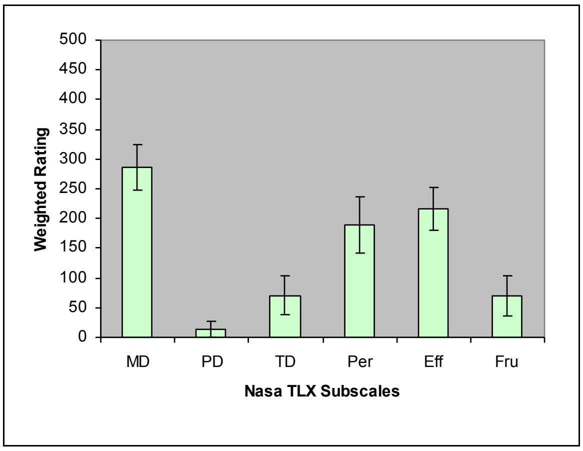


Figure 6-16: Workload Subscales for all sounds (curvature)

Note: Mental Demand (MD), Physical Demand (PD), Temporal Demand (TD), Performance (Per), Effort (Eff), and Frustration (Fru)

6.3.4 Background

During the course of the experiment it was observed that some participants seemed to be undertaking the task more easily than others. Because of this it was thought worthwhile to explore the effect of participant background on performance. Table 6-3 and Table 6-4 summarise performance measures according to participant background for each of the sonification methods. The results here are combined for all sounds as the primary aim of this analysis is to examine the impact of participant background on performance in general.

6.3.4.1 Curve Shape

A separate one-way independent ANOVA was undertaken for each of the performance measures. These revealed that there was a significant effect of background on response time, $F(3,16)=4.90$, $p=0.01$. Pre-planned post-hoc tests (Dunnett's t-test) showed that those with a 'General' background were not significantly different to those from any other background ($p<0.05$). The cause of significance in the ANOVA must therefore have been between other groups.

There was an effect of participant background on accuracy, $F(3,16)=10.95$, $p<0.01$. Pre-planned post-hoc tests (Dunnett's t-test) showed that those with a 'General' background were significantly less accurate than those with a 'Music' or 'Physics/maths' background, p<0.05. Finally, there was no significant effect of background on confidence, F(3,16)=1.60, p=0.23.

Table 6-3: Effect of background on performance (curve shape)

Background	Response Time (s)		Accuracy (%)		Confidence (%)	
	Mean	SD	Mean	SD	Mean	SD
Music	21.50	3.47	89.60	6.69	88.53	8.93
Physics/Maths	27.14	4.99	94.13	6.75	85.07	18.49
Product Design	35.74	9.28	70.67	12.91	72.80	16.66
General	28.32	5.21	61.87	20.23	75.20	16.50

6.3.4.2 Curvature

For curvature a one-way independent ANOVA was undertaken for each of the performance measures. These showed that there was no significant effect of participant background on any of the measures, F(3,16)=0.73, p=0.55, F(3,16)=0.87, p=0.48, and F(3,19)=0.12, p=0.12, respectively.

Table 6-4: Effect of background on performance (curvature)

Background	Response Time (s)		Accuracy (%)		Confidence (%)	
	Mean	SD	Mean	SD	Mean	SD
Music	26.93	5.40	72.53	16.70	76.53	16.00
Physics/Maths	25.76	6.64	86.40	12.98	76.53	20.83
Product Design	29.32	4.56	74.13	13.10	75.20	17.59
General	30.66	6.97	74.67	21.17	71.47	19.00

6.3.5 Interaction Strategies

A number of different interaction strategies were observed (see Figure 6-17). These can be broken down into two broad styles; sweep, and point. The first is where the participant moves the point of contact to produce a changing or sweeping tone (a, b, and c); the second is where the participant touches at a particular point to produce a single tone (d and e). Each of these can be further divided by the sub-style of interaction adopted; total (a), tracking (b and d), and comparison (c and e). With tracking the participant traces a path along the whole of the curve shape, and with comparison they compare two areas of the curve.

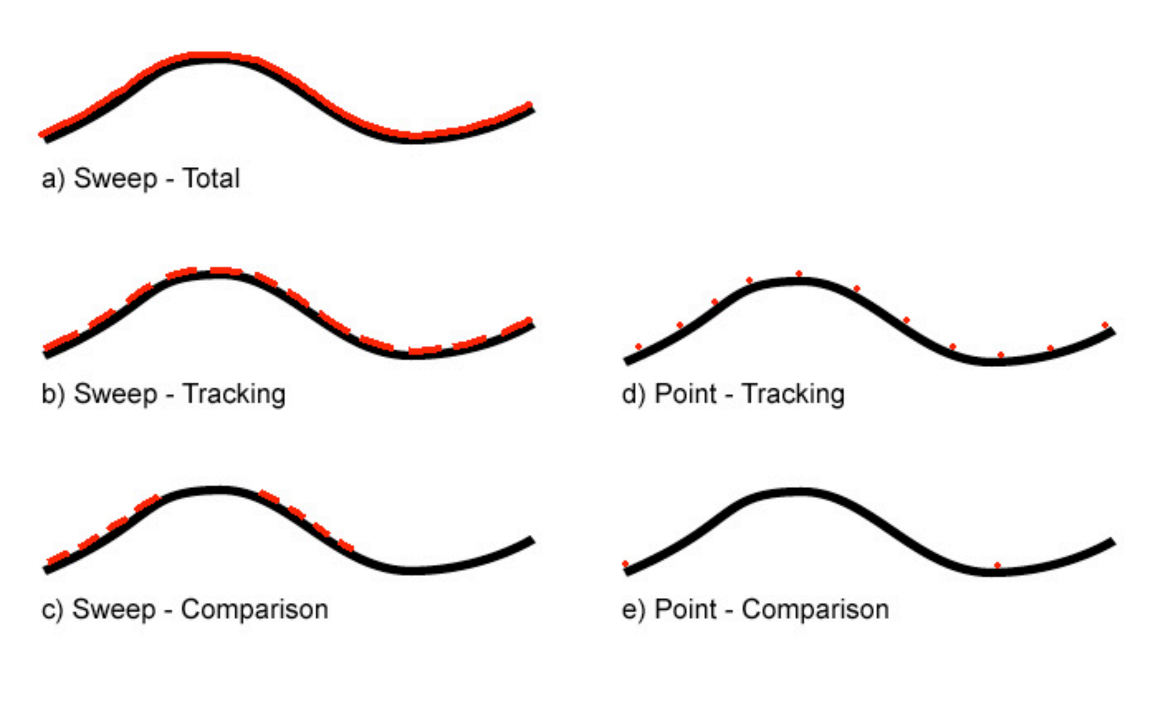


Figure 6-17: Interaction Strategies

The interaction used by participants seems to be adaptive. At the beginning of sessions the participants were observed to use the 'Total Sweep' style. Over the course of a session they would experiment with different types of interaction and seemed to gravitate from continuous auditory feedback to styles that produced discontinuous or discrete sounds, such as 'tracking' and 'comparison'.

6.4 Discussion (Study 2 – Experiment 3)

6.4.1.1 The Appropriateness of Sound

The experiment sought to address the following questions in relation to the appropriateness of sound:

- Was sound an appropriate medium through which to convey curve shape and curvature information? That is, can participants confidently, quickly, and accurately identify the correct curve from auditory feedback?
- Was there an optimum sound type for conveying this information?

The results showed that the different sonification methods, curve shape and curvature, performed similarly across the different performance measures when sound types were conflated. The mean response time was 28 seconds for both, accuracy was 79% for 'Curve Shape' and '77%' for 'Curvature', whilst confidence was 80% and 75%

respectively. Performance for both sonifications was relatively high and indicates that sound was a suitable means of conveying both curve shape and curvature information.

It had not been anticipated that the performance for the different sonifications would be so close. It must be remembered that for the ‘Curve Shape’ sonification the auditory feedback was a direct representation of one of the curves shown, whereas for the ‘Curvature’ sonification the auditory feedback related to the unseen property of curvature. So in the latter case, this meant that in comparing the auditory feedback to the graphs, the participant first had to translate the seen curve shape into an imagined curvature against which the sound was then compared. Therefore given the further complexity of this it is not only a surprise that the accuracy figures are similar but that they were achieved within the same time frame. This goes further to underline the appropriateness of sound for this type of task, in that whilst there is an inherent amount of cognitive load associated with the task, additional complexity does not appear to significantly increase this. What might be an indicator of the increased difficulty is the slightly lower confidence reported with the Curvature sonification. The similarity may also be due to a learning effect as the ‘Curvature’ condition was undertaken after the ‘Curve Shape’ condition.

The type of sound used with the sonifications seems to have had little overall effect. For curve shape there was some effect of sound type on response time and confidence; responses were quicker when using Cello (26s) than with Sine (28s), and there was more confidence with Sine (83%) than with Physical Kinetic (74%). For curvature there was no difference between sounds in terms of response time (28s), accuracy (77%), or confidence (75%). This was potentially a result of learning effects, since all participants completed the ‘Curvature’ evaluation after they had taken part in the ‘Curve Shape’ evaluation. By the time participants undertook the ‘Curvature’ evaluation they may have become accustomed to the sound types to such an extent that their performance was no longer affected by differences between the sounds. This strengthens the observation that sound type had little impact on performance.

6.4.1.2 User Experience and Workload

The experiment sought to address the following questions in relation to user experience issues and workload:

- What attitude did participants express towards their experience of using sound?
- What level of workload did participants experience?

For curve shape, participants felt that they had performed well and that they had improved over time. There was a more even spread of attitude towards ‘Difficulty’ and ‘Irritability’. So whilst some disagreed that the task was difficult or that the sounds were irritating, there was a substantial proportion of participants that expressed a counter attitude. For curvature, there was a similar attitude towards irritation but an increased feeling that the task was difficult (the result here was now significant). Even so participants maintained a sense that they had improved over time, but the view was more mixed about their level of performance. It is therefore not surprising that for both ‘Curve Shape’ and ‘Curvature’ there was a split in attitude towards having had a sense of enjoyment in completing the task.

Across the two sonification methods there were similar attitudes expressed. However, there was a shift towards more negativity for the ‘Curvature’ sonification. This may be a result of the fact that as the task grew more difficult any dissatisfaction or irritation was magnified. Despite this participants seemed to remain positive about their sense of improving in what was perceived to be a difficult task.

Whilst there were differences in attitude to the difficulty of the task, with ‘Curve Shape’ and ‘Curvature’ feedback the NASA-TLX scores for both were similar. For ‘Curve Shape’ the score was 57, and for ‘Curvature’ it was 56. So whilst the subjective experiences reported in terms of task difficulty were different the NASA-TLX scores (also subjective) would indicate that the workloads were actually the same. These subjective feelings, whilst different, are not necessarily contradictory, as a sense of ‘difficulty’ may not map easily against the sub-factors explored by the NASA-TLX. The workload score remained constant across the experimental session under both sonification methods. This suggests that whilst there may have been some learning involved, the nature of the task was not so taxing that this showed up as increased workload at the beginning compared to the end of each the session. The NASA-TLX score ranges from 0 to 100, so the score here would indicate that participants were operating fairly comfortably within the middle of the range.

A breakdown of sub-factors showed higher levels of ‘Mental Demand’, ‘Performance’, and ‘Effort’ compared to the other factors. The levels for each of these were similar for both sonification methods. As mentioned in respect of the performance results this is a little surprising given the ‘unseen’ nature of the task when using the ‘Curvature’ sonification. However, given that the ‘Mental Demand’ was similar in both cases the preconceived difficulty of this task was not apparent within the actual score. It may therefore be concluded that discriminating curve shape and curvature from auditory feedback presented a similar level of workload, and the fact that one was not visible did not induce additional demands.

6.4.1.3 User Background

Within the experimental design, background had not been considered as a separate factor, and steps were taken to ensure that there was a range of aptitudes in the participants undertaking the task. During the course of the experiment it was observed that the 'General' category seemed less able in performing the task. Participants within this category had no formal musical training and this may have affected their ability to undertake the task. Whereas there was a high-level of training within the 'Music' category since these had been selected from students undertaking a music degree, there was at least a moderate level of musical training in both the 'Maths/Physics' and 'Product Design' categories since these participants had indicated undertaking graded exams (between grades 2 to 6). For this reason it was decided to explore the data to identify any performance differences between the 'General' group and the other three categories.

With 'Curve Shape' there was no difference in performance between those with a general background and others for response time and confidence. However, there was an effect of background on accuracy. Those with a 'General' background performed less accurately (62%) than those with a 'Music' (90%) or 'Maths/Physics' (94%) background. With 'Curvature' there was no effect of background on performance. This result was interesting in that the effect of background was not consistent across the two sonification methods. It can be seen that the variance for the curve shape task was lower for those with a music or maths/physics background than the other two backgrounds. However, for the curvature task the variance for all backgrounds is much higher. This indicates that whilst a musical training was helpful, and increased accuracy substantially, compared to those without, the nature of the curvature task nullified any advantage. There could be a few reasons for this; auditory feedback was poorly understood and was improved through direct visual comparison, or curvature was difficult to conceptualise and led to increased error. There was certainly indication of the increased difficulty of the curvature task from the user experience results, and it is doubtful that the auditory feedback was poorly understood since results here and elsewhere indicates otherwise. Whilst no significant differences were found, those with a 'Maths/Physics' background scored more highly (86%) than those with a 'Music' background (72%), who perform similarly to the other backgrounds. It was therefore likely that those with a 'Maths/Physics' background were more easily able to conceptualise curvature, although this was not a significant advantage. It was therefore likely that participants had difficulty conceptualising curvature and this contributed to a lowering of accuracy compared to ‘Curve Shape’.

6.4.1.4 Interaction

The experiment sought to address the following question in relation to interaction:

- Were there any specific interaction strategies employed by participants in completing tasks?

For the first two experiments of this study (see Chapter 5) auditory feedback was constrained, with the participants only being able to passively listen to the sounds. The situation in this experiment was different in that participants were able to control auditory feedback through interaction with the visual representation (via the Wacom tablet). This meant they could choose how, when and for what duration auditory information was provided.

A number of different interaction strategies were observed during the course of ‘Curve Shape’ and ‘Curvature’ exploration. Of the five types of interaction observed, four of these were concerned with breaking the feedback down into small auditory chunks. So rather than listening to the whole of the curve, much of the interaction revolved around exploring and comparing smaller sections of it. This meant that participants opted to receive discreet auditory feedback rather than a continuous stream of information. In this way small sections of the presented curves could be explored in order to identify the correct representation. This type of exploration was particularly noticeable for the curvature sonification. It is thought that long streams of continuous feedback were difficult to interpret and that the exhibited behaviour was an effort to translate these into more manageable chunks. So overall, the interaction is an adaptation away from complexity (continuous sound) and an expression of a need for more simple feedback (discreet tones). However, what is unclear is whether this is a facet of this particular task or an indication of a more general preference.

6.5 Chapter Summary

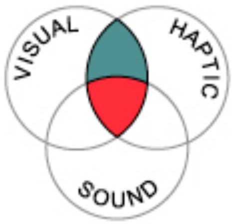
This chapter reported on Experiment 3 of Study 2. This continued the investigation of the appropriateness of sound in conveying curve properties. Participants were presented with four different curves and were required to match auditory feedback to one of the curves displayed. As with the previous experiments, there were different sound types (‘Cello’, ‘Cello Kinetic’, ‘Physical’, ‘Physical Kinetic’, and ‘Sine’) and two sonification methods (‘Curve Shape’ and ‘Curvature’). It was found that participants could quickly, accurately, and confidently match the auditory feedback to the correct visual representation. The type of sound used had little effect on task performance except for minor differences when using the ‘Curve Shape’ sonification (responses with ‘Cello’ were faster, and confidence was higher with ‘Physical Kinetic’ than with ‘Sine’). Overall there was no difference in performance between the two sonification methods. However, when participant background was considered those with a ‘Physics/Maths’ background performed more highly than the others (although not significantly). This may indicate that the ‘Curvature’ sonification was conceptually more difficult to grasp and so led to poorer performances. Even so, there was no difference in the workload attributed to the different sonification methods, with participants operating within the mid-range of the NASA-TLX scale. The user experience results did reflect a slight difference in attitude between the sonification methods, with ‘Curvature’ showing a higher sense that the task was difficult. Finally, it was seen that when participants were allowed to freely interact (as opposed to Experiments 1 and 2 where they were constrained) there was a desire to simplify the feedback by the way in which interaction was executed.

6.6 Related Chapters

The work here and in Chapter 5 has considered the appropriateness of using sound to convey curve information. The effect of sound in combination with other modalities is considered in Chapter 7. The findings here also contribute to the general discussion (see Chapter 10)

Chapter 7: The Effect of Sound on Visual-Haptic Perception, Performance, and Interaction

7.1 Introduction and Rationale



Prior to this chapter there has been a consideration of the perception of curvature (Chapters 3 and 4), performance in haptic, visual, and combined modalities (Chapter 3), and the appropriateness and performance of sound in curvature/curve shape identification (Chapters 5 and 6). In this chapter these different strands will be brought together to evaluate how sound might affect visual-haptic perception, performance and interaction. Additionally, in Chapter 8, the theory that threshold gradient could be a predictor of stimulus gradient (as postulated in Chapter 4) will be explored by further analysis of results obtained in this study.

7.1.1 How this Differs from Studies 1 and 2

This study has many similarities to Study 1 (Chapter 3) but does differ in important ways. The previous study considered visual and haptic modalities across three conditions (visual only, haptic only, and visual-haptic) whereas here only the combined visual-haptic modality was considered separately and in combination with different levels of sound (Curve Shape and Curvature). This reflects the focus of this study, which was to ascertain the effect of the addition of the third modality; sound. Whereas Study 1 sought to understand the effect of combining visual and haptic modalities, and Study 2 examined sound firstly in isolation, and then in combination with vision.

Another important difference from previous studies was that gradient was adopted as the measure of difference between stimuli rather than curvature. This change reflects the analysis undertaken in Chapter 4 which indicated that gradient was the probable mechanism of difference perception. Data from this study therefore underwent further analysis in order to substantiate this hypothesis (see Chapter 8).

Finally, the stimuli used in this study were contained within a singular object, whereas for Study 1 the stimuli were presented as two separate objects. This change was made for a number of reasons. Firstly, this was more reflective of the real-world situation where differences were judged within rather than between objects. Secondly, the need to track finger position to generate sound was more efficiently provided for in this way.

7.1.2 The Stimuli

As outlined above, for the purposes of this experiment differences between stimuli will be expressed with reference to the gradient. However, it should be noted that dimensions of the stimulus such as gradient, base-to-peak height, and curvature are inseparable and co-vary with each other. In Study 1 it was convenient to talk about the stimulus in terms of being ‘High’ and ‘Low’ curvature as this provided a good reference point through which to view differences in performance. However, this also had its limitations when trying to gain a deeper understanding of stimuli effect on performance. This was because in focusing upon the representation of small and large objects through the ‘High’ and ‘Low’ curvatures little consideration was given to the control of stimulus dimensions. Without this control it was difficult in some cases to give proper interpretation to the experimental results. For example, the ‘High’ curvature was situated on a 5cm wide block whilst the ‘Low’ curvature was on a 20cm wide block. Response times were found to be quicker for ‘High’ curvatures. However it is difficult to disambiguate this from the fact that this was also the shorter block; so was the response time affected by the width of the block, the level of the curvature, or a combination of both? Further uncertainty about exactly what was happening was added by the fact that when response time is divided by width the wider block had the quicker response times.

In this study, whilst the primary interest is the effect of gradient, the stimuli were controlled for a number of dimensions; gradient, curvature, and width (see Table 7-1 for dimensions that were held in common; the stimuli themselves are more fully described in 7.2.4). Gradient has already been discussed, but curvature was used because it had applied relevance and width because it is uncertain as to the effect of this on performance. The similarities and differences between the stimuli should allow for more certain interpretation of the data.

Table 7-1: Common dimensions of stimuli

Dimensions	Stimulus		
	D	E	F
Gradient	x	✓	✓
Curvature	✓	✓	x
Width	✓	x	✓

Since the dimensions of the stimulus have been controlled it is possible to make inferences based on these and identify the 'dimension of difference'. For example, Blocks E and F have the same gradient, if contrasts show a significant difference between F and D, but not between E and F, then we may assume that gradient is the dimension of difference. So if F and D are found to be similar, and are significantly different from E, then the dimension of difference is width. All the inferences used in this study are detailed in Table 7-2, Table 7-3, and Table 7-4. The control of dimensions, and inferences based upon them, were designed to be of particular help in disambiguating the effect of stimulus dimensions on performance (see Chapter 8). It was thought that they may also be of use in considering any interactions between sound and stimuli.

Table 7-2: Inferences for two significant differences between stimuli

Stimulus	D	E	F	Inference
D		✓	✓	D and E differed in gradient and width. D and F differed in gradient and curvature. Common difference was gradient, can infer that **gradient** was dimension of effect if E and F do not show a significant difference
E	✓		✓	E and D differed in gradient and width. E and F differed in curvature and width. Common difference was length, can infer that **width** was dimension of effect if D and F do not show a significant difference.
F	✓	✓		F and D differed in gradient and curvature. F and E differed in curvature and width. Common difference was curvature, can infer that **curvature** was dimension of effect if D and E do not show a significant difference.

Notes: ✓ indicates a significant difference between stimuli

Table 7-3: Inferences for one significant difference between stimuli

Stimulus	D	E	F	Inference
D			✓	D and F differed in gradient and curvature. D and E differed in **gradient** and width. F and E differed in **curvature** and width. Common similarity was length, can infer that width was not dimension of effect if DE and EF do not show a significant difference. Dimension of effect is either gradient or curvature.
E	✓			E and D differed in gradient and width. E and F differed in **width** and curvature. D and F differed in **gradient** and curvature. Common similarity was curvature, can infer that curvature was not dimension of effect if EF and DF do not show a significant difference. Dimension of effect is either gradient or width.
F		✓		F and E differed in curvature and width. F and D differed in **curvature** and gradient. E and D differed in **width** and gradient. Common similarity was gradient, can infer that gradient was not dimension of effect if FD and ED do not show a significant difference. Dimension of effect is either curvature or width.

Notes: ✓ indicates a significant difference between stimuli

Table 7-4: Inferences for all or no significant differences between stimuli

Stimulus	D	E	F	Inference
D E F	All Significant			There is more than one dimension of effect, or the effect is unclear
D E F	None Significant			There is no effect of these dimensions

7.1.3 Uniformity of Magnitude Difference

Study 1 indicated that performance was affected by the degree of magnitude difference. For that study difference was classified into small (6%), medium (21 and 27%), and large (36% and 55%) differences. Whilst this classification was convenient and was able to show the effect, there was not a straightforward relationship between these measures. This meant that it was difficult to understand the increase in performance in terms of factors of increase; for example, performance was greater for large differences, however there is little sense of the scale of increase from either small or medium differences. In order to address this it was decided that for this study magnitude differences would be standardised on a base of 3.25% difference from the standard stimulus, and increased as a factor of this to give 1x, 2x, 4x, and 8x difference in magnitude.

7.1.4 Performance Characteristics

The previous studies (1 and 2) indicated a number of characteristics associated with performance in various modalities. The characteristics described differences in response times (RT) between Correct and Incorrect, High and Low Confidence, and High Confidence Correct (HCC) with each of Low Confidence Correct (LCC), High Confidence Incorrect (HCI), and Low Confidence Incorrect (LCI). The findings for each of the modalities are summarised in Table 7-5 and Table 7-6.

Table 7-5: Performance characteristics from previous studies (differences)

Modality	Correct RT < Incorrect RT	High Confidence RT > Low Confidence RT	HCC RT <		
			LCC	HCI	LCI
Visual-Haptic	✓	✓✓	✓✓	x	✓✓
Sound (Cello Curve Shape)	✓	✓✓	✓✓	x	✓✓
Sound (Cello Curvature)	✓✓	✓✓	✓✓	✓✓	✓✓

Note: Significant differences are indicated by ✓ (>0.05) and ✓✓ (>0.01)

Table 7-6: Performance characteristics from previous studies (correlations)

Modality	Accuracy (%) and RT	Confidence (%) and RT	Accuracy (%) and Confidence (%)
Visual-Haptic	x	x	✓
Sound (Cello Curve Shape)	✓✓	✓	✓
Sound (Cello Curvature)	x	x	x

Note: Significant differences are indicated by ✓ (>0.05) and ✓✓ (>0.01)

These showed that visual-haptic responses were characterised by a difference between correct/incorrect and high/low confidence, but that there was no correlation between response time and either accuracy (%) or confidence (%). This differed from the results for sound where there was a negative correlation between response time and both accuracy or confidence. Therefore for sound, when accuracy and confidence increased the response time reduced. What was of interest to this study was to see if these characteristics were preserved or disrupted when combining the visual-haptic and sound modalities.

7.1.5 Observing Interaction

There were a number of observations made about interaction from the first two studies. In Study 1 it was observed that interaction was quicker for longer stimuli (when normalised). This suggested that interaction might not be consistent across the surface of the whole stimuli. It may be the case that interaction is focused on the ends of the stimuli where slope is more prominent, at the centre where there is a change in shape, or it may be evenly distributed across these locations. One purpose of this study was therefore to track interaction across the stimulus to detect if interaction was greater at any particular location and if this was affected by the presence or absence of sound.

In Study 2 it was noticed that two different interaction types were adopted; sweep and point (see Chapter 6 section 6.3.5). The task for that study was matching sound feedback to its visual counter-part. The styles observed may be isolated to that task or they may be indicative of more widely applicable interactions[13]. To further understand the frequency and application of these styles they were used to categorise the interaction found in this study.

[13] Whilst not systematically analysed, these types of interaction were observed in the formative evaluation of the SATIN prototype undertaken in July 2009.

7.1.6 Aims

Four research questions were posed at the start of this study:

1. Does sound affect perceptual acuity, and is this effect dependent upon the sonification used? Does this acuity vary depending upon the gradient level examined?
2. Is performance (response time, accuracy, confidence) affected by the addition of sound? Does the effect of sound vary dependent upon the gradient, or magnitude difference judged? Is any one sonification better than the other? Are performance characteristics changed by the use of sound?
3. Are particular interaction styles adopted dependent upon the sound used? Is the whole of the stimulus explored equally or is exploration focused in some areas more than others? Is the extent of interaction (exploration duration) affected by sound and location?
4. What are participants' subjective experiences of using sound to judge differences? Do the senses used change in the presence of sound? Are the senses perceived as providing conflicting or similar information? Is their mental workload increased by the addition of sound, and is this dependent upon the particular sonification used?

7.2 Method (Study 3)

7.2.1 Participants

Eight participants, four female and four male, completed the study. Their ages ranged between 18 and 55. The results of the sound experiments (Studies 2 and 3) had shown some effect of background and practice. It was therefore decided to ensure a mix of backgrounds to minimise this effect. Four participants were students from The University of Nottingham; two studying Mathematics or Physics and two studying Music. The remaining four participants had no recent or advanced Music, Mathematics or Physics experience, and came from the general population. All participants were recruited from those that had been involved in sound experiments 2 and 3 (Study 2) in order to minimise practice effect. On completion of the study, participants were compensated for their time with a payment of £120.

7.2.2 Equipment and Set Up

The equipment was set up as shown in Figure 7-1. Participants were seated on a chair throughout the experiment. At the beginning of the session this was adjusted in order to ensure that they were comfortable throughout the session. The chair offered back support, but no arm support (to ensure free movement of arm). The stimuli blocks were presented in a clear holder that was fixed against a black surface. This was arranged so that the participant could easily reach the blocks to be examined.

Figure 7-1: Experimental setup

Figure 7-2: Experiment control and data recording equipment

The experiment was recorded with a Sony camcorder positioned to the left of the work area and focused on the stimuli block. Sound was provided by two Behringer Ms16 Monitor speakers, these were the same as used in the previous sound studies (described in Chapters 5 and 6). They were positioned to the left and right of the work area, and provided stereo feedback. It should be noted that from the participant's position (seated on the chair) the view of the experimenter's monitors is obscured by a screen. Figure 7-2 shows the set up on the experimenter's side of the screen. There were two computers; a laptop which controlled the sound and recording of haptic interaction, and another computer which ran the e-Prime software controlling the experiment. The stimulus blocks can be seen stored below a low table prior to being used – note the labels to ensure the correct block is used.

In order to track the user's interaction with the stimulus each block was fitted with an Eowave position sensor (see Appendix H for technical data sheet). This was connected to the computer via an Eowave Eobody2 USB 8 SensorBox. It was necessary to develop a specific software programme in order to track movement using this hardware. For this purpose a Max/MSP patch was created by colleagues at the Technische Universiteit Eindhoven from functional and interface specifications provided by the author (see CD

/other-documents/functional-interface-specification.doc). This enabled tracking of position and duration corresponding to five zones on the stimuli (see Figure 7-3).

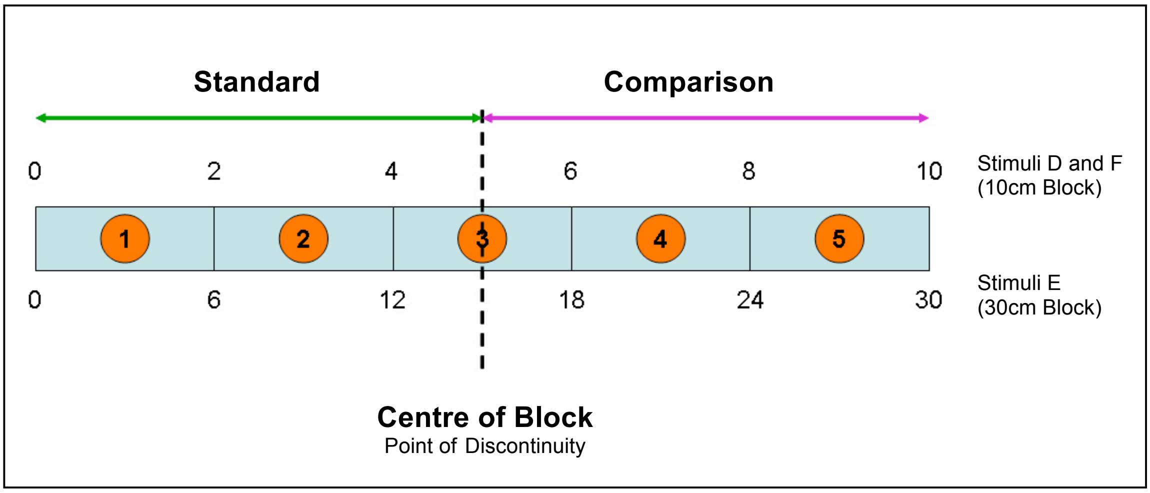


Figure 7-3: Interaction zones

7.2.3 Procedure

The structure of the experiment can be seen in Figure 7-4. Each stimulus set (D, E, and F, see 7.2.4 Stimuli for details) was presented in separate sessions due to time and setup constraints. A session would take between two and half to three hours to complete. The sessions were held on different days, and there was at least one day but no more than seven days between sessions. Each session was broken down into three evaluations; one for each of the sound conditions (no sound, curve shape, and curvature). The evaluations lasted approximately fifty minutes and there was a ten minute break between evaluations. The order of the stimulus (sessions) and sound (trials) conditions were randomised using a Latin square to counter-balance for order effects.

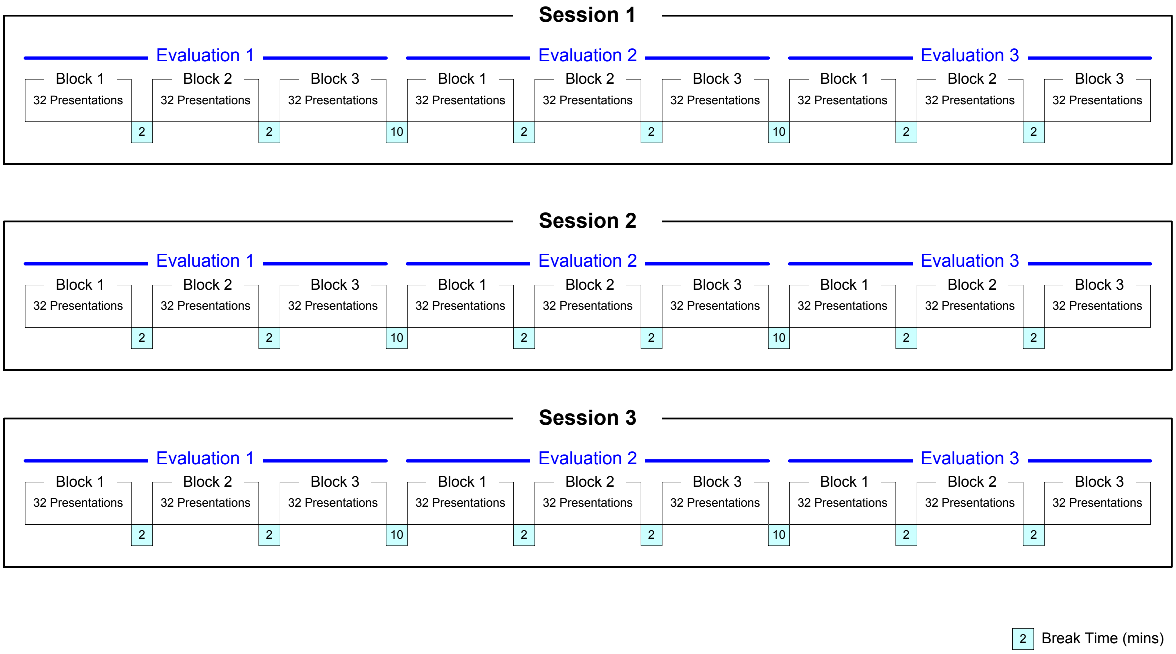


Figure 7-4: Experiment structure

Within an evaluation individual stimulus blocks from a particular set were presented. Each stimulus block had a standard and a comparison half, and there were eight blocks in total for each set. The stimuli were presented twelve times each (six times with the standard to the right and six times with standard to the left).This gave a total of 96 presentations for each trial which were randomised using the ePrime software. This was too many to present comfortably in one block, so presentations were divided into three blocks of 32 presentations with a break of two minutes between each. The presentation blocks took between 10 and 15 minutes to complete dependent upon the conditions and participant.

The procedure for all sessions was the same, except for the first where participants were asked to sign a consent form and complete a demographic questionnaire. At the beginning of the session the participant would be welcomed and asked to make themselves comfortable on the seat provided. They were then asked to read through an introduction to that day's session which explained that they would undertake three evaluations.

The first evaluation commenced by asking the participant to read through an introductory document specific to the sound condition being evaluated. The facilitator followed this up by summarising what was expected and the nature of the sound that would be heard (if relevant). The participant was then asked some test questions in order to ensure that they had understood the notion of ‘more or less curved’. Following this the stimuli were presented to the participant.

The procedure for presentation of each stimulus was the same. For each presentation participants were asked to judge if the right-side of the stimulus was ‘more’ or ‘less’ curved than the left-side. They could examine the stimuli as many times as they wished, but had to comply with the following restrictions;

- They had to use only the forefinger of the dominant hand
- They should not lean the arm or hand on the table
- They should sit in an upright position, at a comfortable reaching distance from the block.
- They should not stoop down to view the block in profile

Once they gave their answer they were then asked to state how confident they were in that answer, responding either ‘high’ or ‘low’ confidence. For each presentation a record was kept of their answers, the response time, and the duration in each of five stimulus zones. The evaluation ended once all 96 presentations had been made. At the end of each evaluation the participant was asked to complete a workload and user experience questionnaire. All documentation used to support this study can be found on the accompanying CD as detailed in Appendix I.

7.2.4 Stimuli

There were three different stimuli used in this study. The key dimensions of these stimuli are given in Table 7-7 (full details of all dimensions can be found in Appendix C). The stimuli were designed to have some dimensions in common, for example stimuli D and F are the same length, but differ from E in this respect (see Table 7-1). The purpose of this was to enable identification of the dimension of ‘difference’ (see 7.1.2, for explanation of the rationale behind this). Where performance differed between stimuli the fact that the stimuli shared or had dimensions in difference would help to identify which facet of the stimulus had an effect on performance. For example, if stimulus D and F were found to have significantly different response times, since they had width in common it is unlikely that this was the factor of difference. This would leave either curvature or gradient (or an unknown factor) as facets that may have affected this difference. Since stimuli D and E have curvature, and E and F have gradient in common, then by also examining these results it was considered possible to infer the facet of stimulus that is effective in promoting difference in response time (see Table 7-2 to Table 7-4 for inferences).

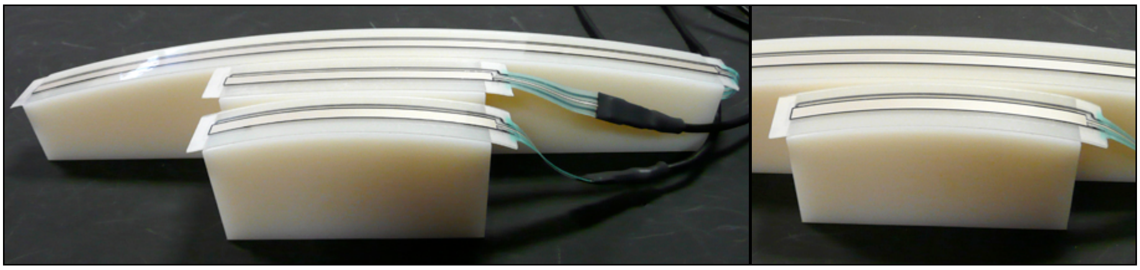


Figure 7-5: Stimulus blocks (large image shows blocks D, E, and F. small image shows close up of blocks E and F which have the same gradient)

Table 7-7: Stimulus dimensions

Dimensions	Standard Stimuli		
	D	E	F
Gradient	0.037	0.111	0.111
Curvature	1.46/m	1.46/m	4.38/m
Width	10cm	30cm	10cm

As with Study 1, each set of stimuli had one standard and eight comparison stimuli. However, an important difference from Study 1 is that the standard and comparison stimuli were contained within the same block (as opposed to having separate blocks for each). The purpose of this was to more closely mimic detection of object difference in an applied setting. The comparison gradients were derived as a proportion of the standard. This proportion was initially 3.25% and was doubled for each subsequent comparison, thus giving comparisons of 3.25% (1x), 6.5% (2x), 13% (4x), and 26% (8x) gradient difference from the standard. The comparisons were made above and below the standard so giving eight comparisons in total for each of the stimuli D, E, and F.

It should be noted that there was some difficulty in deciding on the range of comparison stimuli. The problem was that the stimuli needed to encompass a sufficient range to be able to derive thresholds for all sound conditions. From the prediction model derived in Chapter 4 it was possible to have a good approximation of the difference threshold for ‘no sound’ (visual-haptic only), hence the upper limit of 26%. However, the lower limit was more difficult to judge as it was uncertain how much improvement (if any) would be given by the addition of sound. The only guide was to compare the performance of visual-haptic in Study 1 against that achieved by sound in Study 2 (Experiment 2 - magnitude). For the visual-haptic modality accuracy was 90%, compared to 94% for the curved shape and 97% for the curvature sounds. It was therefore felt that a lower limit of 3.25% would easily encompass all thresholds.

The stimulus blocks were made from a nylon composite material to ensure that the surface texture of each block was uniform and smooth. The dimensions for all stimuli were the same in respect of the block's central-height (5cm) and depth (2.5cm), other dimensions varied as shown in Table 7-7. The stimuli block contained one half of its width that was the standard stimulus and one half that was the comparison. This is illustrated in Figure 7-6, the differences between the two are exaggerated here and were far more subtle in reality as seen in Figure 7-5.

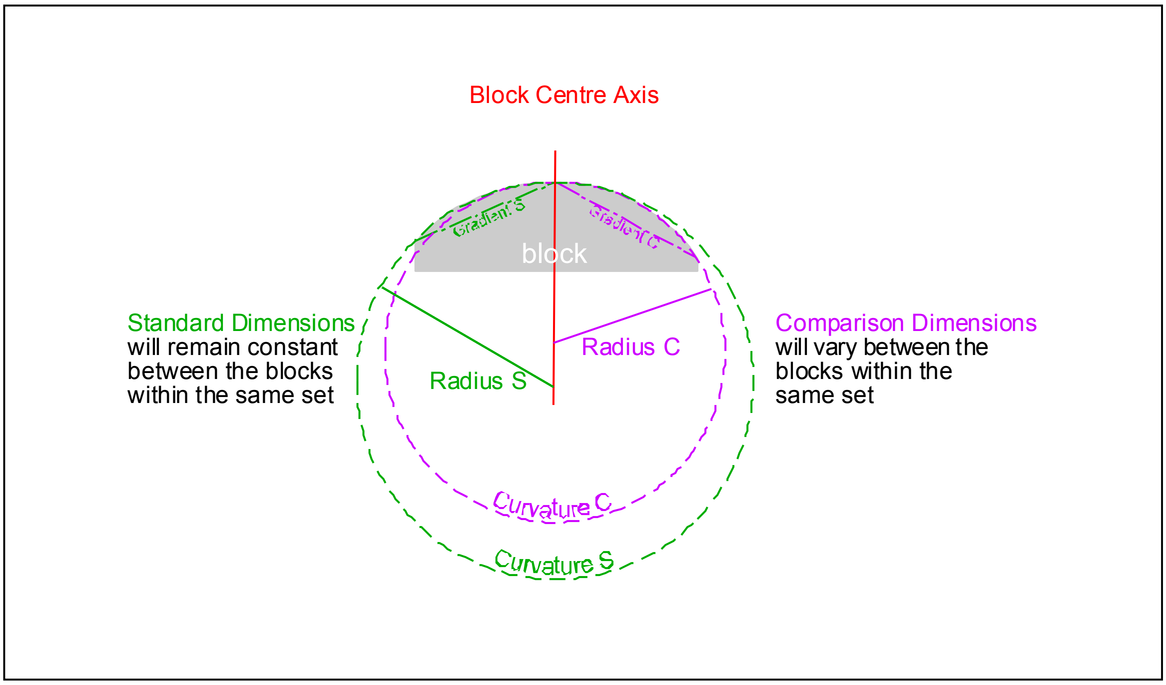


Figure 7-6: Standard and comparison dimensions and how they apply to the block

7.2.5 Sonification

7.2.5.1 Choice of Sound

The sound used in this study was chosen from those examined in the sound study (see Chapters 5 and 6). In that study three sounds, 'Cello', 'Physical', and 'Sine', were assessed to determine which facilitated the better performance when conveying object characteristics such as curve shape and curvature. It was concluded that there was little difference between the 'Cello' and 'Sine' sounds. However, for the curve shape mapping, 'Cello' did evoke better confidence when judging curvature magnitude. For this reason the 'Cello' sound was selected for use in this study. The 'Cello' sound was a sampled cello with the decay removed. This gave a sound with a constant frequency of 131Hz which is equivalent to C3 (an octave below middle C). This was designed by colleagues from Technische Universiteit Eindhoven who had expertise in sound design. For a more technical description of these sounds see the paper co-authored with Shelley et al. (2009).

7.2.5.2 Sound Mapping

Two types of mappings were used in this experiment; curved shape and curvature. These are the same as those used in experiment 2 of the Study 2. For both mappings the frequency ranges from 100Hz to 400Hz, which is about a two octave range. The 'Curved Shape' sonification follows the shape of the curve from left to right. The mapping is anchored on the start point of the standard stimulus (see Figure 7-7). So imagine that this is on the left-hand side, the sound begins with a frequency of 100Hz and rises in frequency to the midpoint where it is 400Hz. It then decreases in frequency back to around 100Hz at the right end of the curve (the comparison side). The end-point frequency will change depending on the comparison stimuli sonified. The 'P' stimuli will be slightly greater than 100Hz whilst 'M' stimuli will be slightly less than 100Hz.

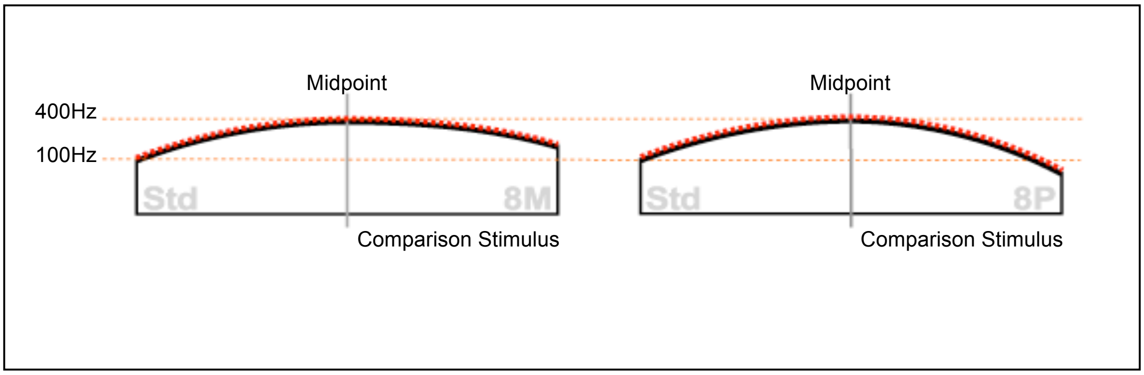


Figure 7-7: Curve shape sonification

The 'Curvature' sonification is mapped to the curvature value of the curve. For each stimuli set (D, E, and F) the comparison with the lowest curvature is block 8M and this is mapped to a frequency of 100Hz. The comparison with the highest curvature is block 8P and is mapped to a frequency of 400Hz. Because the curvature at any point in a circle is the same, the effect is that the standard half produces one continuous tone and the comparison half produces another. The tone in the comparison half will be higher or lower than the standard half depending on whether the comparison curvature is higher or lower (see Figure 7-8). The standard curvature, being the mid-point between the highest and lowest comparison curvatures, had a frequency of 200Hz (this is one octave above 100Hz and one octave below 400Hz). Intermediate changes in curvature (for other stimuli) where mapped to logarithmic changes in frequency. The exact frequencies for each stimulus are not so much the concern as the percentage change between stimuli which because of the mapping is the same as differences between curvatures i.e. 3.25%, 6.5%, 13% and 26%. This means that, for example, the sonified frequency for the 1M stimulus had a difference of 3.25% in frequency from the standard.

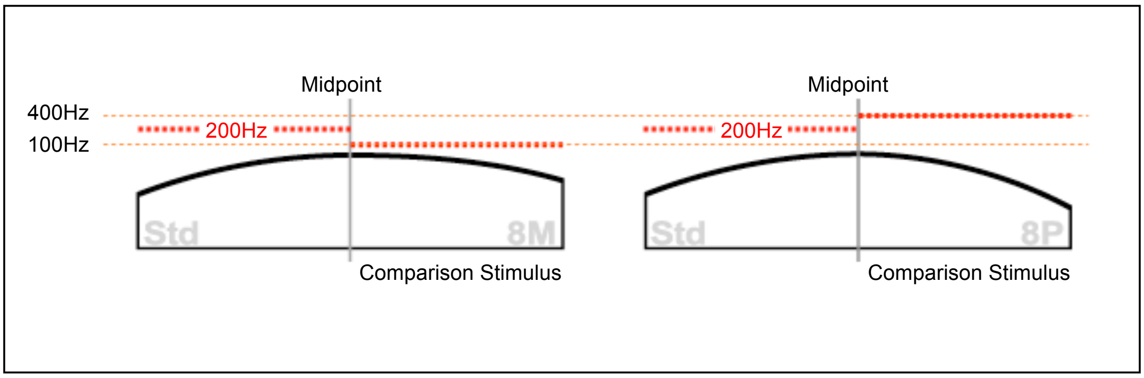


Figure 7-8: Curvature sonification

7.2.6 Design and Hypotheses

The experiment was a repeated-measures design, and applied the psychophysical method of constant stimuli to establish the difference threshold for gradient in No Sound, Curve Shape, and Curvature conditions.

It also sought to examine the effects of three independent variables; Stimuli (D, E and F), Sound (No Sound, Curve Shape, and Curvature), and magnitude difference (3.25, 6.5, 13, and 26 percent) on performance (response time, accuracy, and confidence), and interaction (exploration duration). In order to answer the research questions posed (see 7.1.6) the following hypotheses were examined:

7.2.6.1 Perception (H1-H3)

H1: There is an effect of sound on difference threshold.
H2: There is an effect of stimulus on difference threshold.
H3: There is an interaction affect between stimuli and sound on difference threshold.

7.2.6.2 Performance (H4-H21)

H4-H7: There is an effect of sound on response time; accuracy; confidence.
H8-H9: There is an effect of stimulus on response time; accuracy; confidence.

H10-12: There is an effect of magnitude difference on response time; accuracy; confidence.
H13-H15: There is an interaction effect between sound and stimulus on response time; accuracy; confidence.
H16-H18: There is an interaction effect between sound and magnitude difference on response time; accuracy; confidence.
H19-H21: There is an interaction effect between sound, stimulus, and magnitude difference on response time; accuracy, and confidence.

7.2.6.3 Performance Characteristics (H22-H39)

H22-H24: For each sound (NoS, CS, CURV) there is a significant difference in response time between correct and incorrect judgements.
H25-H27: For each sound (NoS, CS, CURV) there is a correlation between response time and accuracy.
H28-H30: For each sound (NoS, CS, CURV) there is a significant difference in response time between high and low confidence judgements.
H31-H33: For each sound (NoS, CS, CURV) there is a correlation between response time and confidence.
H34-H36: For each sound (NoS, CS, CURV) there is a significant difference in response time between high confidence correct judgements and low confidence correct; high confidence incorrect; low confidence incorrect.
H37-H39: For each sound (NoS, CS, CURV) there is a correlation between accuracy and confidence.

7.2.6.4 Interaction (H40-H43)

H40: There is an association between sound and interaction style.
H41: There is an effect of sound on exploration duration.
H42: There is an effect of location on exploration duration.
H43: There is an interaction effect between sound and location difference on exploration duration.

7.2.6.5 User Experience (H44-H62)

H44: There is an association between sound and ease of use.
H45: There is an association between sound and difficulty.
H46: There is an association between sound and helpfulness.
H47: There is an association between sound and concentration.
H48: There is an association between sound and senses used.

H49: There is an association between sound and sense conflict.
H50: There is an effect of sound on workload score.
H51-H56: There is an effect of sound on mental demand; physical demand; temporal demand; performance; effort; frustration.

7.2.7 Data Analysis

The performance data (response time, accuracy, and confidence) were found to be non-normally distributed for some variables; either positive or negative skew which is usual for these types of data. This was not a problem as ANOVAs are considered to be a robust method of statistical analysis (Davies 1956; Field 2009)(see Appendix F for discussion of this). A number of pre-planned contrasts were used in the analysis of this data. For sound, all were contrasted to ‘No Sound’ as this was the control modality. For stimuli, all were contrasted with ‘D’ and ‘F’ to ensure that all were contrasted with each other and so inferences could be draw as per Table 7-2, Table 7-3, and Table 7-4 . For magnitude difference all were contrasted with ‘1x’ which was likely to show the lowest performance and so improvement from this could be gauged. For other statistical tests the appropriate parametric analysis methods were used for non-normal variables (the tests used are noted within the results).

7.3 Results (Study 3)

This section details all results in detail, a summary of the results in relation to the hypotheses can be found in Appendix D.

7.3.1 Perception

Threshold gradients were calculated for each of the sound conditions in each stimulus. These were the mean of individual threshold gradients for each participant as shown in Table 7-8, Table 7-9, and Table 7-10 (see Appendix J for calculation of these individual thresholds and model statistics). It should be noted that thresholds for the curvature condition, whilst calculated, used limited data points as the performance level was better than the lowest anticipated difference of 3.25%. Albeit with some reservation[14], the curvature thresholds were calculated in order to provide useable data for the ANOVA.

---

[14] In order to effectively calculate threshold there should be at least three data points of a value greater than 0 and less than 100. The thresholds here were calculated using only two data points (plus and minus 3.25%) with values of 0 and 100. Whilst this does not give an accurate figure for

Table 7-8: Individual JNDs and Weber Fractions for No Sound

P-ID	No Sound D		E		F	
	JND	WF	JND	WF	JND	WF
P1	0.005	14.065	0.013	11.354	0.014	12.388
P2			0.013	12.113	0.013	11.974
P3	0.004	11.922	0.013	11.985	0.012	11.033
P4	0.005	12.592	0.012	11.211	0.013	11.722
P5	0.004	11.732	0.012	11.202	0.013	11.456
P6	0.004	11.054	0.013	11.541	0.013	11.521
P7	0.006	15.719	0.013	11.671	0.014	12.374
P8	0.007	17.630	0.012	11.134	0.013	11.832
M	0.005	13.530	0.013	11.526	0.013	11.788
SD	0.001	2.406	0.000	0.371	0.001	0.462

Table 7-9: Individual JNDs and Weber Fractions for Curve Shape

P-ID	Curve Shape D		E		F	
	JND	WF	JND	WF	JND	WF
P1	0.004	11.711	0.013	11.312	0.014	12.935
P2	0.005	12.715	0.015	13.209	0.013	11.725
P3	0.004	11.850	0.015	13.708	0.013	11.281
P4	0.005	13.727	0.016	14.778	0.017	14.921
P5	0.004	11.822	0.013	12.027	0.014	12.169
P6	0.004	12.135	0.013	11.378	0.012	10.941
P7	0.005	12.750	0.014	12.651	0.012	11.044
P8	0.004	11.926	0.016	14.252	0.014	12.308
M	0.005	12.330	0.014	12.914	0.014	12.165
SD	0.000	0.692	0.001	1.296	0.001	1.306

Table 7-10: Individual JNDs and Weber Fractions for Curvature

P-ID	Curvature D		E		F	
	JND	WF	JND	WF	JND	WF
P1	0.001	2.298	0.005	4.599	0.004	3.957
P2	0.001	2.298	0.003	2.323	0.003	2.485
P3	0.001	2.253	0.003	2.278	0.003	2.278
P4	0.001	2.298	0.003	2.323	0.003	2.323
P5	0.001	2.363	0.003	2.284	0.003	2.323
P6	0.001	2.298	0.003	2.323	0.003	2.278
P7	0.001	2.253	0.003	2.323	0.003	2.323
P8	0.001	2.298	0.003	2.323	0.003	2.323
M	0.001	2.295	0.003	2.597	0.003	2.536
SD	0.000	0.034	0.001	0.809	0.001	0.578

---

the threshold it does reflect an approximation for each participant. As it was known that the true threshold was somewhere below 3.25% (given the almost uniform 100% accuracy at this level), this method allows for the data to be used in analysis as it provides differing means for each participant.

A repeated-measures ANOVA compared the effect of sound and stimulus on the percentage change in gradient (see Figure 7-9). This revealed that there was a significant effect of sound, $F(2,12)=355.01$, $p<0.001$, partial $\eta^2=0.98$. Pre-planned contrasts showed that there was no significant difference between 'No Sound' (12.23%, SD=1.57) and 'Curve Shape' (12.47%, SD=1.13), but that there was between 'No Sound' and 'Curvature' (2.48%, SD=0.56), $F(1,6)=0.20$, $p=0.67$, partial $\eta^2=0.03$, and $F(1,6)=766.53$, $p<0.001$, partial $\eta^2=0.99$, respectively. There was no effect of stimuli, or interaction effect of sound and stimuli, on percentage change in gradient, $F(2,12)=1.89$, $p=0.19$, partial $\eta^2=0.24$, and $F(1.69,10.12)=4.10$, $p=0.055$, partial $\eta^2=0.41$, respectively.

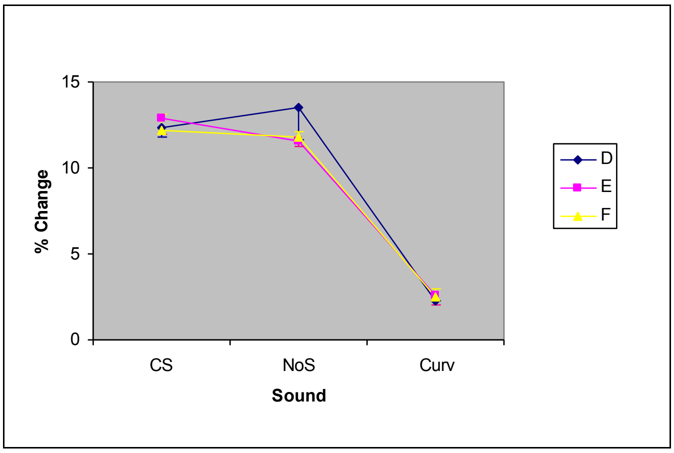


Figure 7-9: Effect of sound and stimulus on Weber Fractions

Note: Error bars represent the 95% Confidence Interval of the mean, and do so throughout

7.3.2 Performance

7.3.2.1 Response Time

A repeated-measures ANOVA compared the effect of sound, stimulus, and magnitude difference on response time (see Figure 7-10 and Table 7-11). This revealed that there was a main effect of sound, $F(2,14)=80.10$, $p<0.01$, partial $\eta^2=0.92$. Pre-planned contrasts showed that 'No Sound' had a significantly slower response time than 'Curvature' but was significantly faster than 'Curve Shape', $F(1,7)=9.26$, $p=0.019$, partial $\eta^2=0.57$, and $F(1,7)=67.47$, $p<0.01$, partial $\eta^2=0.91$, respectively. There was no main effect of stimuli, $F(2,14)=1.79$, $p=0.20$, partial $\eta^2=0.20$. There was a main effect of magnitude difference, $F(3,21)=23.88$, $p<0.01$, partial $\eta^2=0.77$. Pre-planned contrasts showed that '1x' difference, was significantly slower than '8x' but was not significantly different from '2x' or '4x' difference, F(1,7)=52.01, p<0.01, partial η^2=0.88, F(1,7)=0.21, p=0.663, partial η^2=0.03, and F(1,7)=1.26, p=0.299, partial η^2=0.15, respectively.

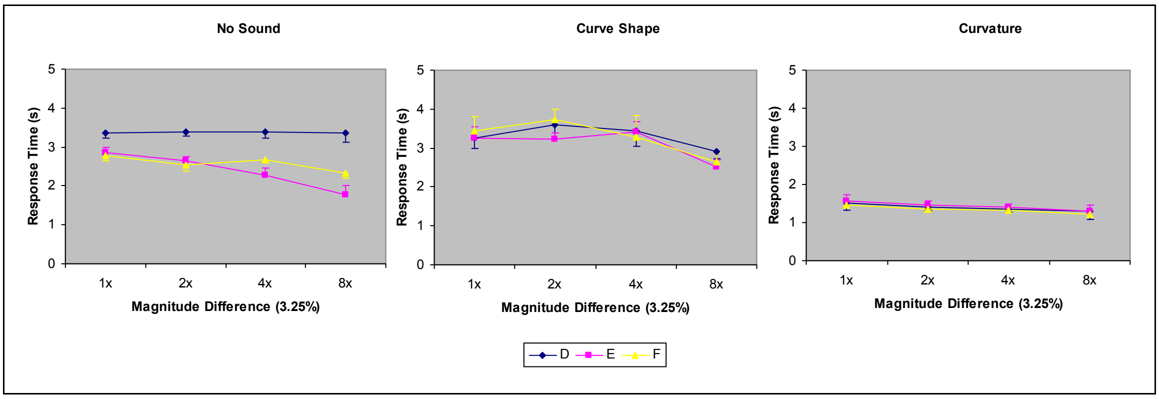


Figure 7-10: Effect of sound, stimulus, and magnitude difference on response time

Note: Error bars represent the 95% Confidence Interval of the mean, and do so throughout

Table 7-11: Mean response time for sound, stimulus, and magnitude difference

Variable	Mean	SD
No Sound	2.78	0.47
Curve Shape	3.23	0.63
Curvature	1.39	0.44
Stimulus D	2.69	0.72
Stimulus E	2.31	0.36
Stimulus F	2.40	0.59
1x Magnitude Difference	2.61	0.55
2x Magnitude Difference	2.59	0.49
4x Magnitude Difference	2.51	0.41
8x Magnitude Difference	2.15	0.43
All	2.47	0.46

There was a significant interaction between sound and stimulus on response time, F(4,28)=4.90, p<0.01, partial η^2=0.41. Pre-planned contrasts revealed that the increase in response time for 'Curve Shape' (compared to 'No Sound') was less for stimulus 'D' compared to either 'E' or 'F' (see Figure 7-11), F(1,7)=16.26, p<0.01, partial η^2=0.70, and F(1,7)=16.26, p<0.01, partial η^2=0.70, respectively. The decrease in response time for 'Curvature' (compared to 'No Sound') was greater for stimulus 'D' than either 'E' or 'F' (see Figure 7-12), F(1,7)=9.98, p=0.01, partial η^2=0.59, and F(1,7)=11.72, p=0.01, partial η^2=0.63, respectively. There was no significant interaction between sound and magnitude difference, F(1.87,13.07)=3.56, p=0.06, partial η^2=0.34.

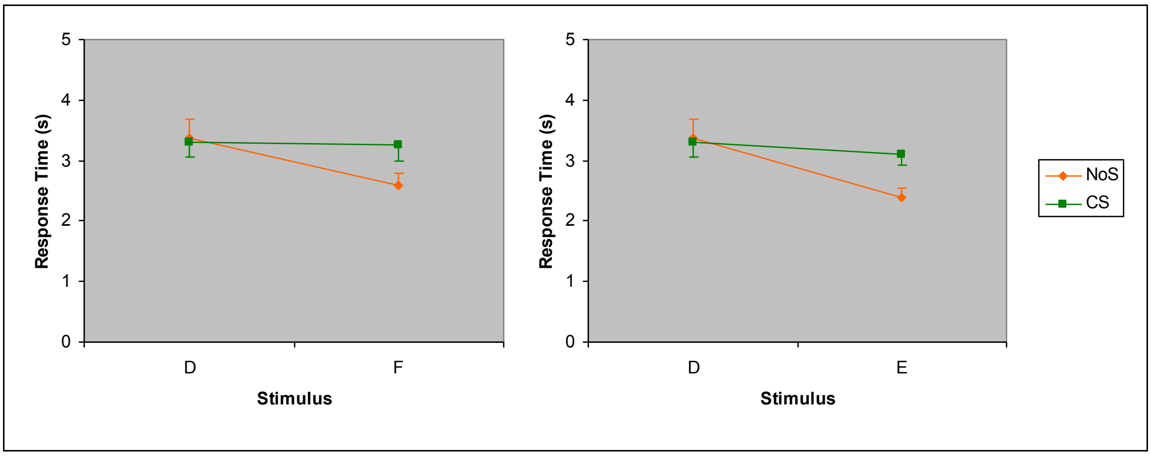


Figure 7-11: Interaction effect of sound and stimulus on response time

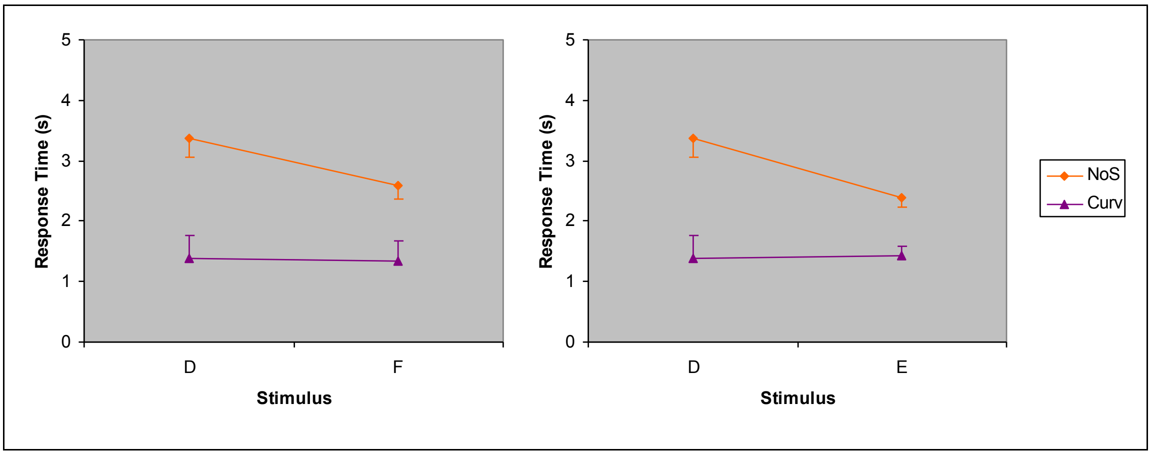


Figure 7-12: Interaction effect of sound and stimulus on response time

There was a significant interaction between stimulus and magnitude difference, $F(6,42)=3.13$, $p=0.013$, partial $\eta^2=0.31$. Pre-planned contrasts showed that the decrease in response time for '8x' difference (compared to '1x') is greater for stimulus 'E' than 'D' (see Figure 7-13), $F(1,7)=8.27$, $p=0.024$, partial $\eta^2=0.54$.

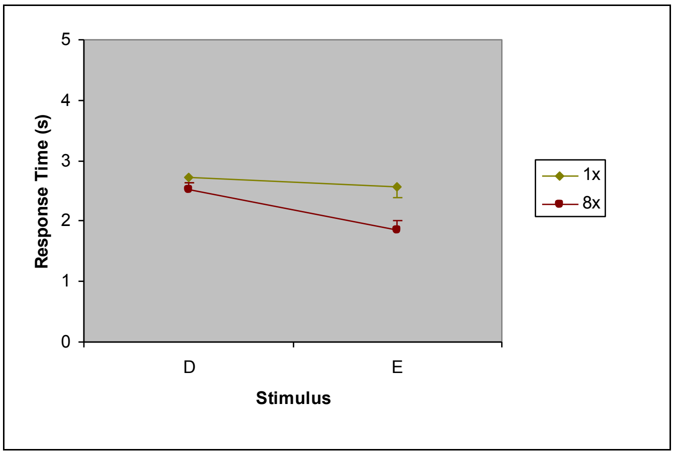


Figure 7-13: Interaction effect of sound and stimulus on response time

Finally, there was a significant three-way interaction between stimuli, sound and magnitude differences, $F(12,84)=2.39$, $p=0.01$, partial $\eta^2=0.25$. Pre-planned contrasts revealed that the interaction between stimuli and magnitude difference varied across different levels of sound (see Figure 7-14). It can be seen that for 'No Sound' (compared to 'Curvature) the decrease in response time for both 4x and 8x magnitude difference (compared to 1x) is significantly greater for stimulus 'E' compared to 'D', $F(1,7)=9.48$, $p=0.018$, partial $\eta^2=0.57$, and $F(1,7)=11.30$, $p=0.012$, partial $\eta^2=0.62$, respectively.

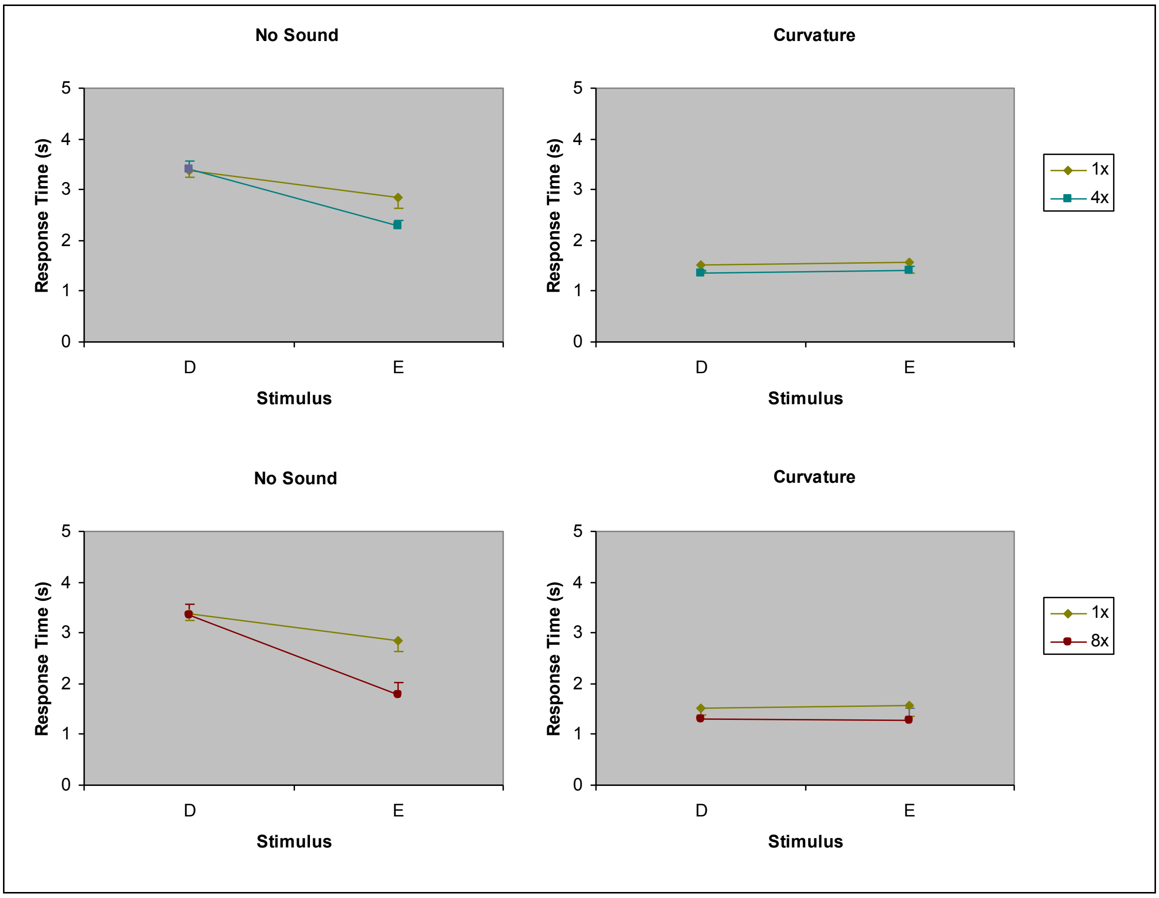


Figure 7-14: Interaction effect of sound, stimulus, and magnitude difference on response time

7.3.2.2 Accuracy

A repeated-measures ANOVA compared the effect of sound, stimulus, and magnitude difference on accuracy (see Figure 7-15 and Table 7-12). This revealed that there was a main effect of sound, $F(2,14)=127.39$, $p<0.01$, partial $\eta^2=0.95$. Pre-planned contrasts showed that 'No Sound' was significantly less accurate than 'Curvature' but showed no significant difference from 'Curve Shape', $F(1,7)=176.09$, $p<0.01$, partial $\eta^2=0.96$, and $F(1,7)=0.24$, $p=0.64$, partial $\eta^2=0.03$, respectively. There was a main effect of stimuli, $F(2,14)=20.39$, $p<0.01$, partial $\eta^2=0.74$. Pre-planned contrasts showed that 'E' was significantly more accurate than either ‘D’ or ‘F’, and ‘F’ was more accurate than ‘D’, F(1,7)=39.03, p<0.01, partial η^2=0.85, F(1,7)=19.44, p<0.01, partial η^2=0.73, and F(1,7)=6.49, p=0.04, partial η^2=0.48, respectively. There was a main effect of magnitude difference, F(3,21)=142.03, p<0.01, partial η^2=0.95. Pre-planned contrasts showed that ‘1x’ difference, was significantly less accurate than either ‘2x’, ‘4x’, or ‘8x’, F(1,7)=13.78, p=0.01, partial η^2=0.66, F(1,7)=107.98, p<0.01, partial η^2=0.94, and F(1,7)=503.17, p<0.01, partial η^2=0.99, respectively.

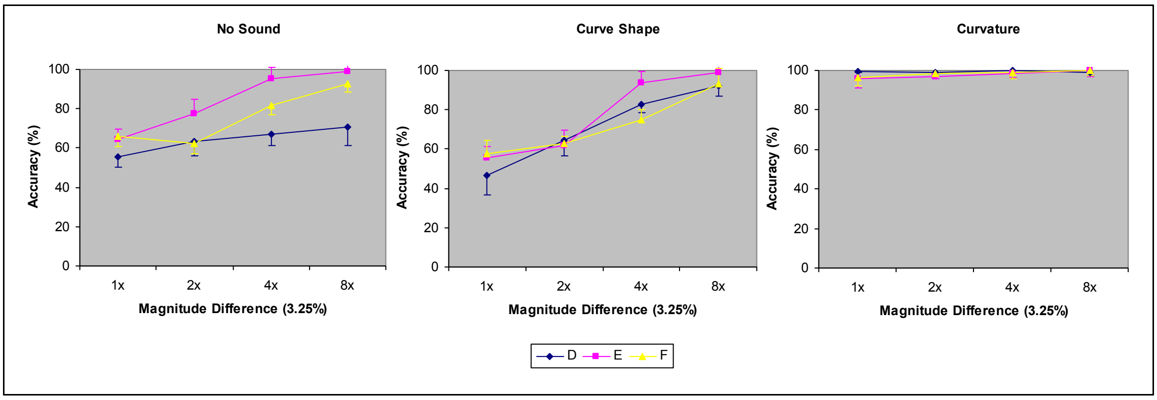


Figure 7-15: Effect of sound, stimulus and magnitude difference on accuracy

Table 7-12: Mean accuracy for sound, stimulus, and magnitude difference

Variable	Mean	SD
No Sound	74.74	4.26
Curve Shape	73.78	2.86
Curvature	98.52	2.39
Stimulus D	78.39	3.16
Stimulus E	86.50	1.98
Stimulus F	82.16	2.54
1x Magnitude Difference	70.89	3.15
2x Magnitude Difference	76.39	2.32
4x Magnitude Difference	88.14	2.64
8x Magnitude Difference	93.98	2.54
All	82.35	1.57

There was a significant interaction between sound and stimulus on accuracy, F(4,28)=9.5, p<0.01, partial η^2=0.58. Pre-planned contrasts revealed that for ‘No Sound’ (compared to ‘Curve Shape’) stimulus ‘E’ and ‘F’ (see Figure 7-16) had a greater effect on accuracy compared to D, F(1,7)=6.28, p=0.04, partial η^2=0.47, and F(1,7)=6.34, p=0.04, partial η^2=0.47, respectively. This was also true when comparing ‘No Sound’ to ‘Curvature’, stimulus ‘E’ and ‘F’ had a greater effect on accuracy compared to ‘D’ (see Figure 7-17), F(1,7)=45.07, p<0.01, partial η^2=0.87, and F(1,7)=19.60, p<0.01, partial η^2=0.74, respectively. Similarly, stimulus ‘F’ (compared to ‘E’) had a greater effect on accuracy with ‘No Sound’ (compared to ‘Curvature’), F(1,7)=29.78, p<0.01, partial η^2=0.81.

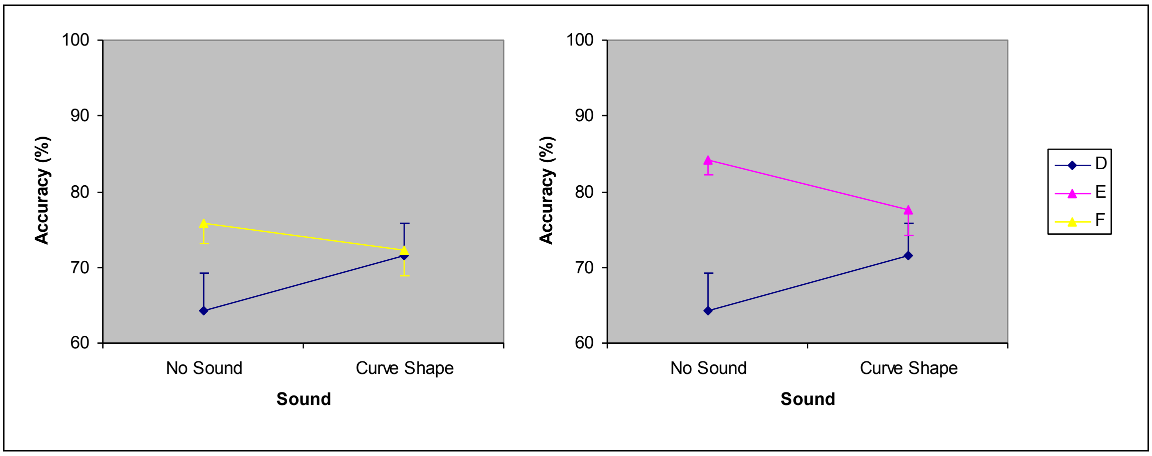


Figure 7-16: Interaction effect of sound and stimulus on accuracy

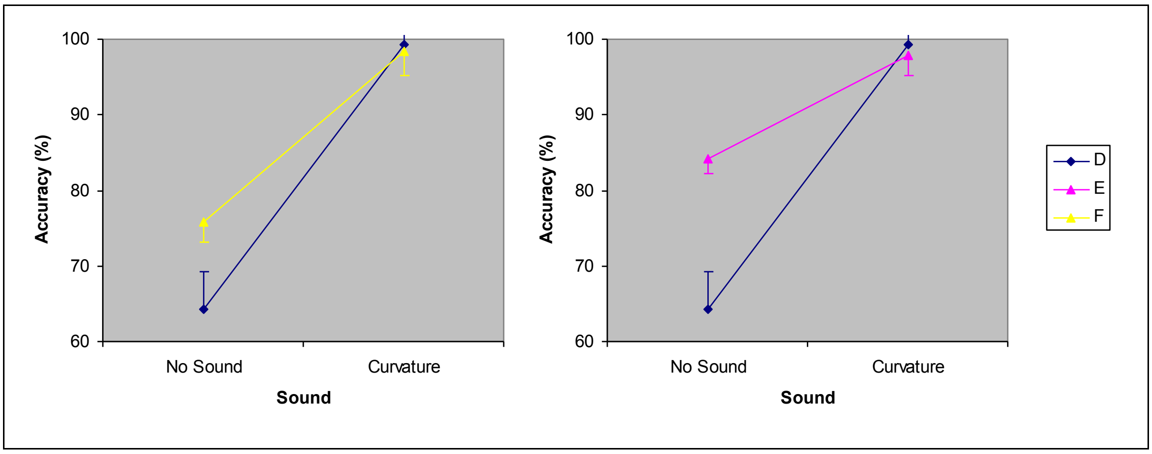


Figure 7-17: Interaction effect of sound and stimulus on accuracy

There was a significant interaction between sound and magnitude difference on accuracy, $F(6,42)=22.25$, $p<0.01$, partial $\eta^2=0.76$. The increase in accuracy for '4x' and '8x' difference (compared to '1x', see Figure 7-18) was greater for 'Curve Shape' (compared to 'No Sound'), $F(1,7)=5.64$, $p=0.049$, partial $\eta^2=0.45$, and $F(1,7)=7.63$, $p=0.02$, partial $\eta^2=0.52$, respectively. However, the increase in accuracy for '4x' and '8x' difference (compared to '1x', see Figure 7-19) was less for 'Curvature' (compared to 'No Sound'), $F(1,7)=93.08$, $p<0.01$, partial $\eta^2=0.93$, and $F(1,7)=45.39$, $p<0.01$, partial $\eta^2=0.87$, respectively.

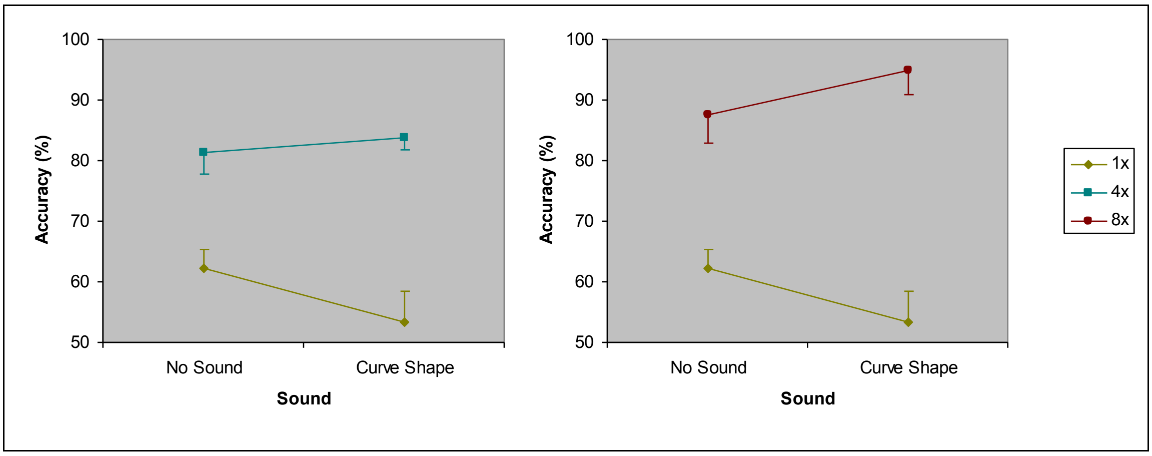


Figure 7-18: Interaction effect of sound and magnitude difference on accuracy

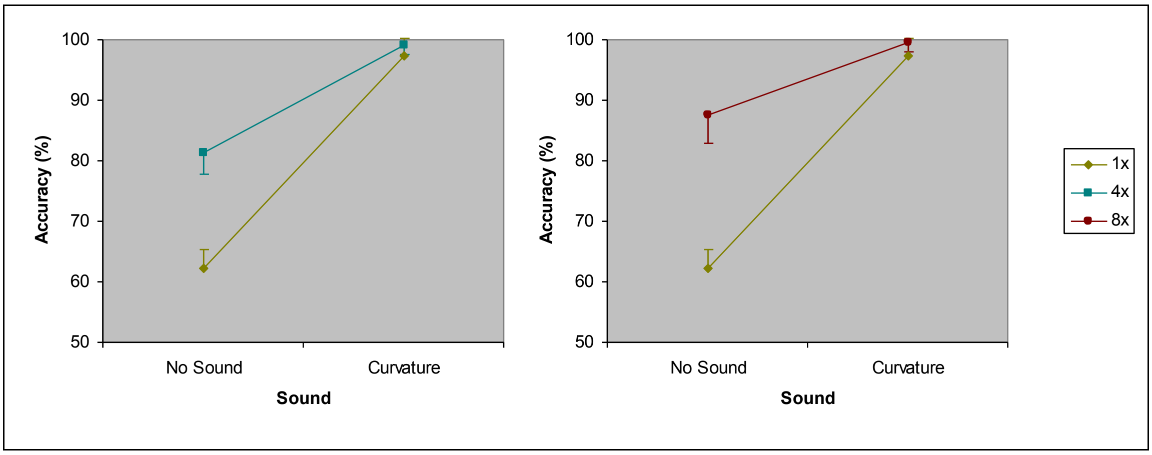


Figure 7-19: Interaction effect of sound and magnitude difference on accuracy

There was a significant interaction between stimulus and magnitude difference on accuracy, $F(6,42)=22.46$, $p<0.01$, partial $\eta^2=0.76$. The increase in accuracy for 'x4' difference (compared to 'x1') was greater for 'E' compared to either 'D' or 'F' (see Figure 7-20), $F(1,7)=6.94$, $p=0.034$, partial $\eta^2=0.50$, and $F(1,7)=18.13$, $p<0.01$, partial $\eta^2=0.72$, respectively. Additionally, the increase in accuracy for '2x' difference (compared to '1x') was greater for stimulus 'D' compared to 'F' (see Figure 7-21), $F(1,7)=6.24$, $p=0.041$, partial $\eta^2=0.47$.

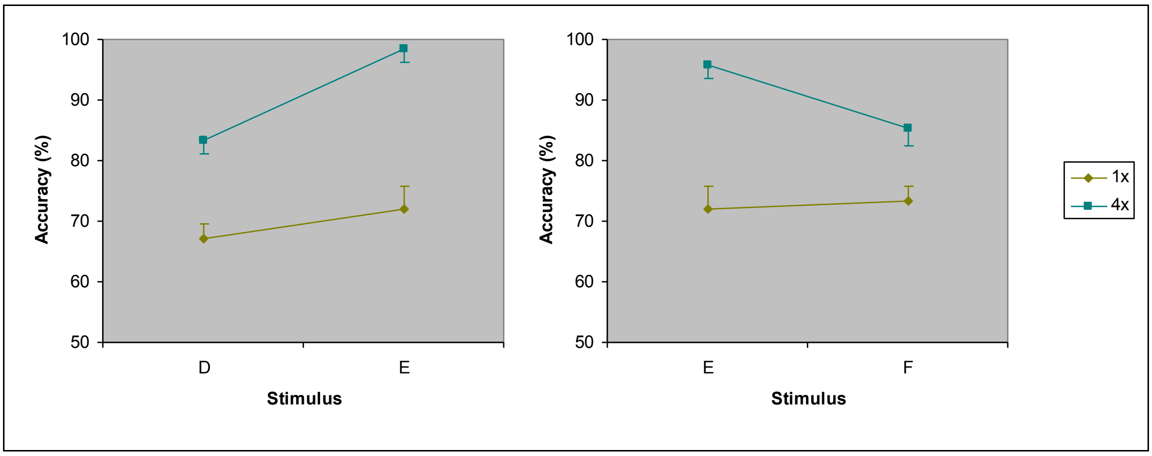


Figure 7-20: Interaction effect of stimulus and magnitude difference on accuracy

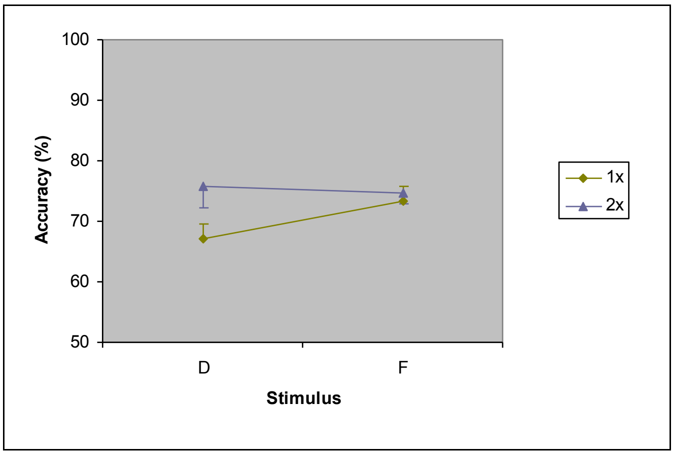


Figure 7-21: Interaction effect of stimulus and magnitude difference on accuracy

Finally there was a significant three-way interaction between stimulus, sound, and magnitude difference on accuracy, $F(12,84)=2.97$, $p<0.01$, partial $\eta^2=0.44$. Pre-planned contrasts revealed that the interaction between stimuli and magnitude difference varied across different levels of sound. It can be seen that for 'No Sound' (compared to 'Curvature') the increase in accuracy for 2x and 4x magnitude difference (compared to 1x) is significantly greater for stimulus 'E' compared to 'F' (see Figure 7-22). It can also be seen that for 'No Sound' (compared to 'Curvature') the increase in accuracy for 4x and 8x magnitude difference (compared to 1x) is significantly greater for stimulus 'E' compared to 'D' (see Figure 7-23). All significant ANOVA statistics are reported in Table 7-13.

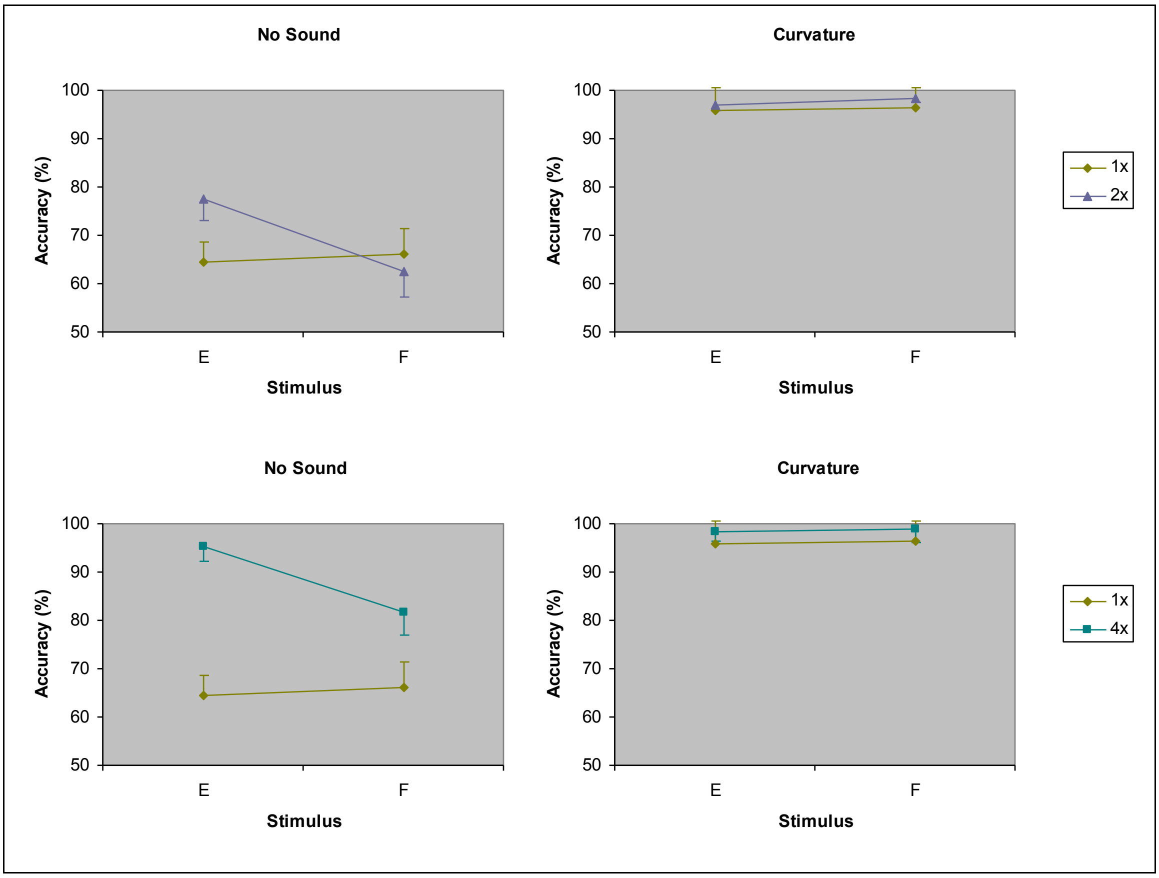


Figure 7-22: Interaction effect of sound, stimulus, and magnitude difference

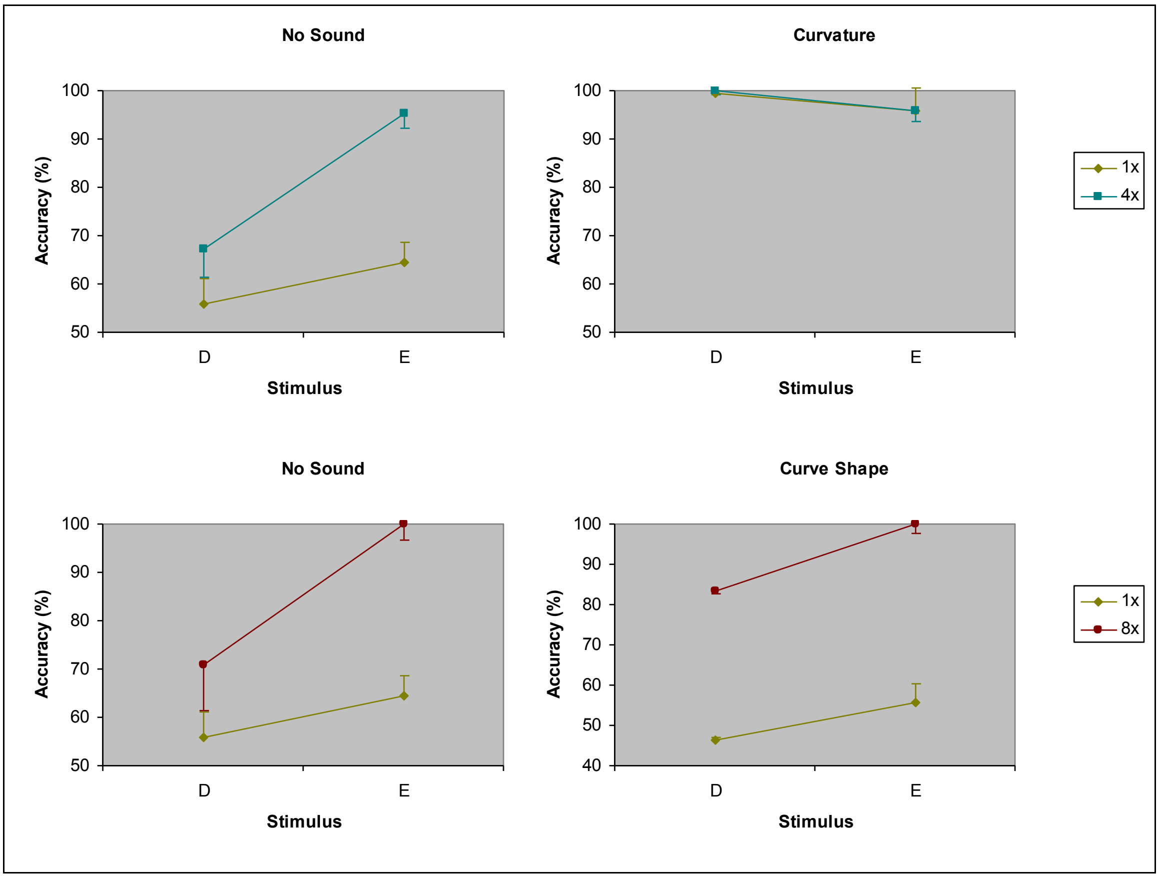


Figure 7-23: Interaction effect of sound, stimulus, and magnitude difference

Table 7-13: ANOVA statistics of significant three-way interactions

Stimulus	Sound	Mag. Diff.	df	F	Sig.	partial η^2
E vs. F	NoS vs. Curv	1x vs. 2x	1,7	6.50	0.038	0.48
		1x vs. 4x	1,7	16.77	0.00	0.71
D vs. E	NoS vs. CS	1x vs. 8x	1,7	6.14	0.042	0.47
	NoS vs. Curv	1x vs. 4x	1,7	9.83	0.02	0.58

7.3.2.3 Confidence

A repeated-measures ANOVA compared the effect of sound, stimulus, and magnitude difference on confidence (see Figure 7-24 and Table 7-14. This revealed that there was a main effect of sound, F(2,14)=11.49, p<0.01, partial η^2=0.62. Pre-planned contrasts showed that ‘Curve Shape’ and ‘Curvature’ were significantly more confident than ‘No Sound’, F(1,7)=6.10, p=0.043, partial η^2=0.46, and F(1,7)=14.76, p=0.01, partial η^2=0.68, respectively. There was no main effect of stimuli, F(2,14)=2.14, p=0.15, partial η^2=0.23. There was a main effect of magnitude difference, F(3,21)=18.86, p<0.01, partial η^2=0.73. Pre-planned contrasts showed that ‘1x’ difference, was significantly less confident than ‘8x’ but was not significantly different from ‘2x’ or ‘4x’ difference, F(1,7)=0.03, p=0.85, partial η^2=0.01, F(1,7)=1.13, p=0.32, partial η^2=0.14, and F(1,7)=34.68, p<0.01, partial η^2=0.83, respectively.

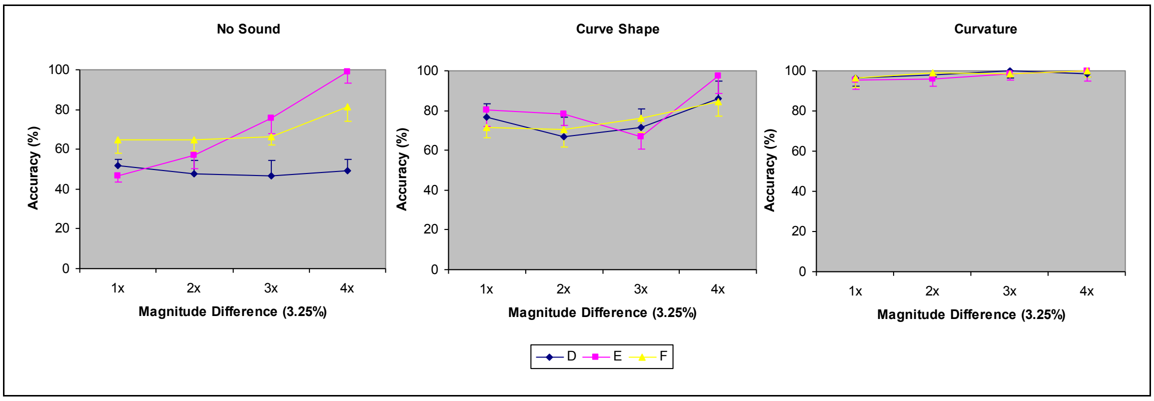


Figure 7-24: Effect of sound, stimulus and magnitude difference on confidence

There was no significant interaction effect between sound and stimulus on confidence, F(2.19,15.52)=3.10, p=0.07, partial η^2=0.31. There was a significant interaction between sound and magnitude difference, F(6,42)=5.07, p<0.01, partial η^2=0.42. Pre-planned contrasts revealed that the increase in confidence for ‘8x’ difference (compared to ‘1x’) was less for ‘Curvature’ compared to ‘No Sound’ (see Figure 7-25), F(1,7)=8.77, p=0.021, partial η^2=0.56.

Table 7-14: Mean confidence for sound, stimulus, and magnitude difference

Variable	Mean	SD
No Sound	62.54	24.30
Curve Shape	77.08	18.54
Curvature	98.00	5.23
Stimulus D	74.09	15.31
Stimulus E	82.51	8.59
Stimulus F	81.03	19.03
1x Magnitude Difference	75.46	14.32
2x Magnitude Difference	75.23	15.32
4x Magnitude Difference	77.72	13.87
8x Magnitude Difference	88.43	10.46
All	79.21	13.15

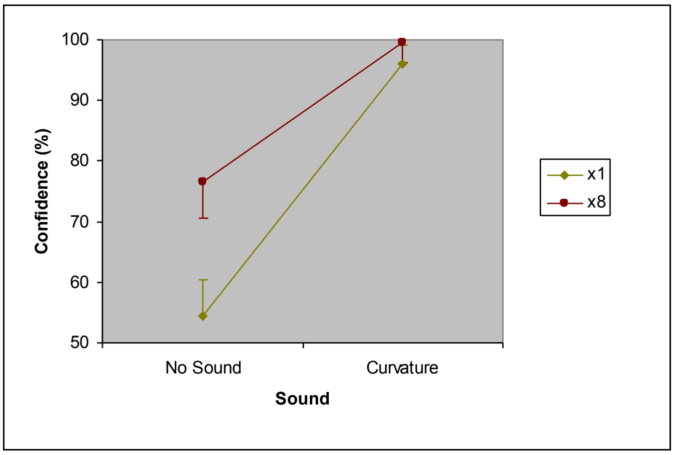


Figure 7-25: Interaction between sound and stimulus on confidence

There was a significant interaction between stimulus and magnitude difference, $F(3.14,21.98)=6.21$, $p<0.01$, partial $\eta^2=0.47$. Pre-planned contrasts revealed that the increase in confidence for '8x' difference (compared to '1x') was greater for 'E' compared to either 'D' or 'F' (see Figure 7-26), $F(1,7)=19.50$, $p<0.01$, partial $\eta^2=0.74$, and $F(1,7)=17.89$, $p<0.01$, partial $\eta^2=0.72$, respectively. It was also shown that the increase in confidence for '4x' difference (compared to '1x') was greater for 'E' than for 'D' (the interaction looks similar to that shown in Figure 7-26 for '8x' difference), $F(1,7)=8.75$, $p=0.021$, partial $\eta^2=0.56$.

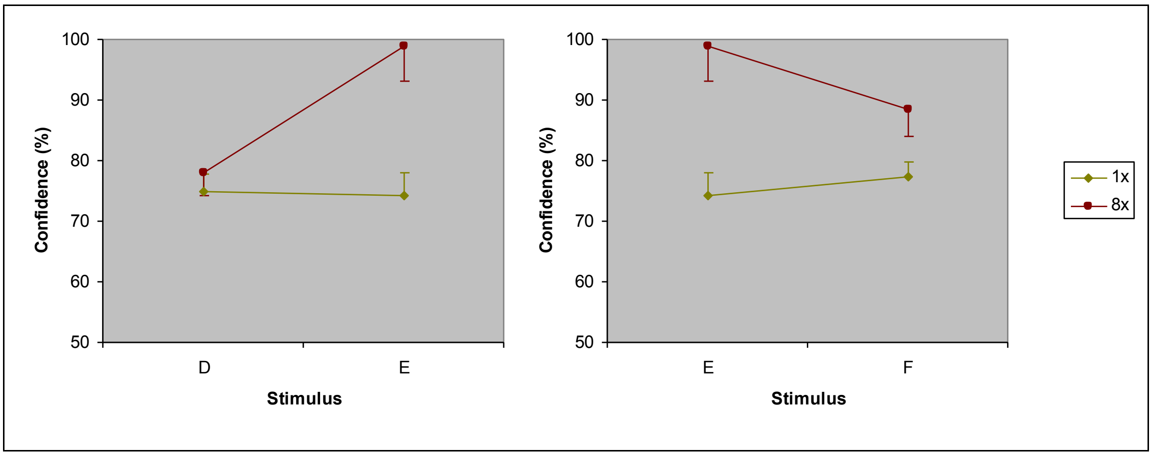


Figure 7-26: Interaction effect between stimulus and magnitude difference

Finally, there was a significant three-way interaction between stimuli, sound and magnitude differences on confidence, $F(12,84)=5.57$, $p<0.01$, partial $\eta^2=0.44$. Pre-planned contrasts revealed that the interaction between stimuli and magnitude difference varied across different levels of sound. It can be seen that for ‘No Sound’ (compared to ‘Curve Shape’ and ‘Curvature’) the increase in confidence for ‘4x’ and ‘8x’ magnitude difference (compared to ‘1x’) is significantly greater for stimulus ‘E’ compared to ‘D’ (see Figure 7-28). Similarly, it can be seen that for ‘No Sound’ (compared to ‘Curvature’) the increase in confidence for ‘4x’ and ‘8x’ magnitude difference (compared to 1x) is significantly greater for stimulus ‘E’ compared to ‘F’. However when comparing ‘No Sound’ to ‘Curve Shape’ the increase in confidence for ‘E’ (compared to ‘F’) is only significant at ‘4x’ magnitude difference[15] (see Figure 7-29). A similar effect is also observed when comparing stimulus ‘F’ with ‘D’, which saw a greater increase in confidence for 8x magnitude difference (compared to 1x) for ‘F’ (see Figure 7-27). All significant ANOVA statistics are reported in Table 7-15.

Table 7-15: ANOVA statistics of significant three-way interactions

Sound	Stimulus	Mag. Diff.	df	F	Sig.	partial η^2
NoS vs. CS	D vs. E	1x vs. 4x	1,7	9.15	0.019	0.57
		1x vs. 8x	1,7	6.63	0.037	0.49
	E vs. F	1x vs. 4x	1,7	8.25	0.024	0.54
NoS vs. Curv	D vs. E	1x vs. 4x	1,7	22.70	0.002	0.76
		1x vs. 8x	1,7	25.69	0.001	0.79
	D vs. F	1x vs. 8x	1,7	8.30	0.024	0.54
	E vs. F	1x vs. 4x	1,7	9.29	0.019	0.57
		1x vs. 8x	1,7	13.07	0.009	0.65

[15] The result for ‘8x’ magnitude difference is marginal, $F(1,7)=5.25$, $p=0.056$, partial $\eta^2=0.43$. This has been included in the results as the effect size is high and Figure 7-29 shows an interaction.

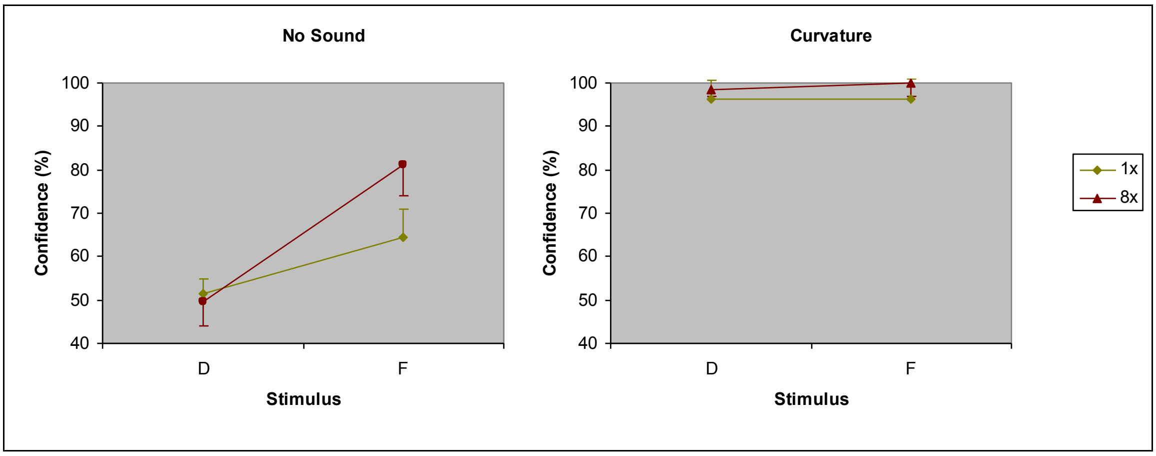


Figure 7-27: Interaction effect of sound, stimulus, and magnitude difference on confidence

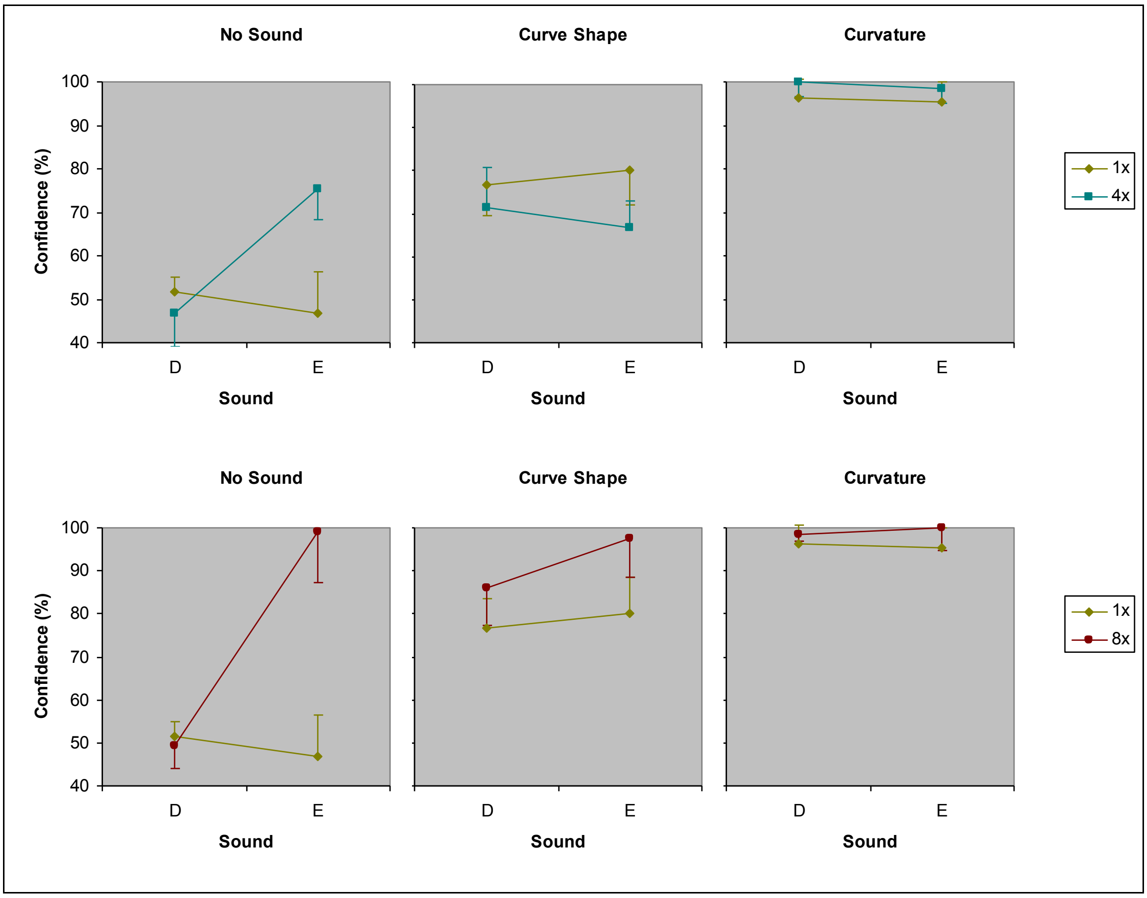


Figure 7-28: Interaction effect of sound, stimulus, and magnitude difference on confidence

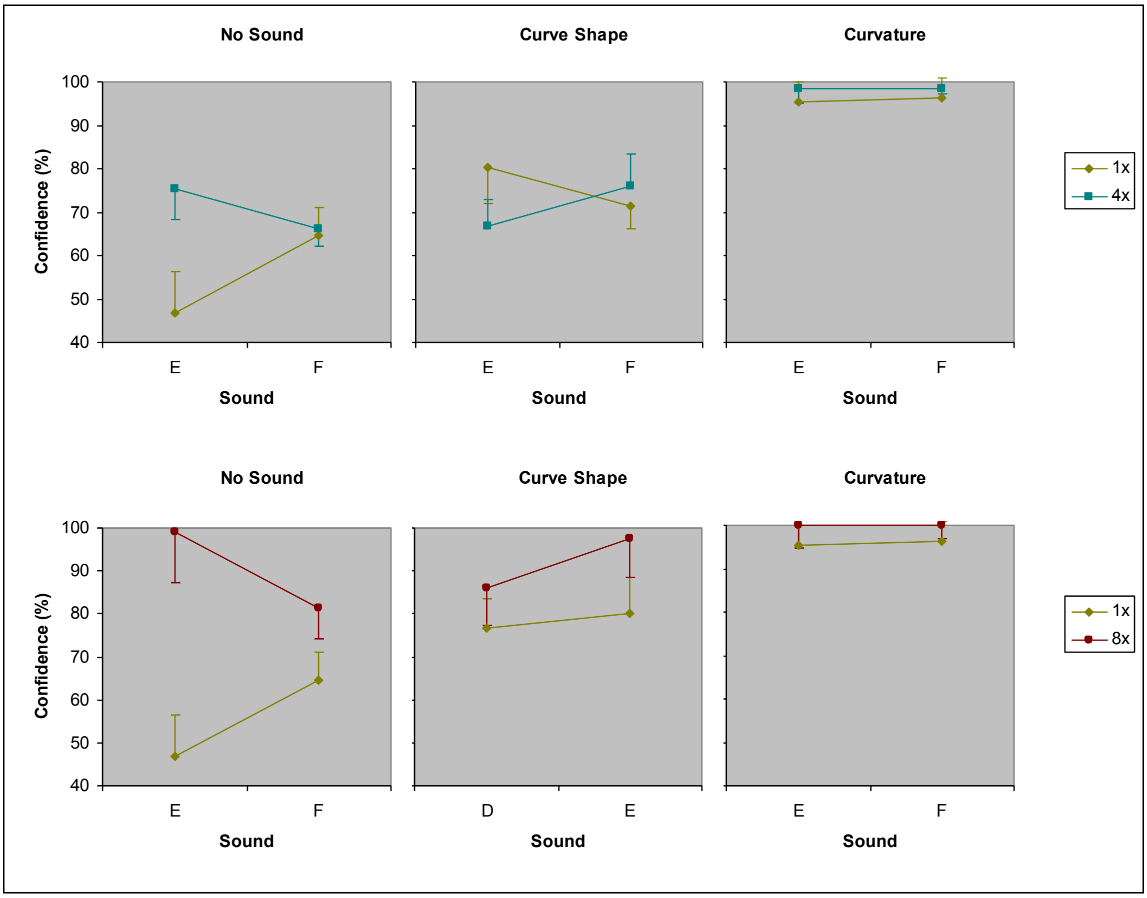


Figure 7-29: Interaction effect of sound, stimulus, and magnitude difference on confidence

7.3.3 Performance Characteristics

7.3.3.1 Accuracy and Response Time

Repeated-measures t-tests were used to examine the differences between correct and incorrect response times for different levels of sound. These revealed that response times for correct answers were significantly faster than incorrect for 'No Sound', and 'Curve Shape'. There was no significant difference between response times for 'Curvature' (see Table 7-16 for mean response times and test statistics).

Table 7-16: Mean RT for Correct and Incorrect judgments with significance of differences

Sound	RT Correct		RT Incorrect		Dependent t-test			
	Mean	SD	Mean	SD	t	df	r	sig. (2-tailed)
No Sound	2.89	0.67	3.40	0.68	-7.56	7	0.94	0.00
Curve Shape	3.39	0.62	3.85	0.95	-3.00	7	0.75	0.02
Curvature	1.41	0.52	1.89	0.82	-2.06	5	0.68	0.09

A series of scatter plots were used to compare response time against accuracy for each sound condition. These did not appear to indicate any correlation between accuracy and response time. Further analysis was undertaken (Pearson's r) which confirmed that there was no correlation between response time and accuracy for any of the variables tested, p>0.05 (see Appendix K for full results).

7.3.3.2 Confidence and Response Time

Repeated-measures t-tests were used to examine the differences between high confidence and low confidence response times for different levels of sound. These revealed that response times for high confidence answers were significantly faster than low confidence for 'No Sound', and 'Curve Shape'[16] (z=-2.52, p<0.01). There was no significant difference between response times for 'Curvature' (see Table 7-17 for mean response times and test statistics).

Table 7-17: Mean RT for high and low confidence with significance of differences

Sound	RT High		RT Low		Dependent t-test			
	Mean	SD	Mean	SD	t	df	r	sig. (2-tailed)
No Sound	2.47	0.58	4.21	1.66	-4.28	7	0.85	0.00
Curve Shape	2.99	0.61	5.32	2.05	-3.41	7	0.79	0.01
Curvature	1.37	0.37	3.06	1.39	-2.41	2	0.86	0.14

A series of scatter plots were used to compare response time against confidence for each sound condition. These did not appear to indicate any correlation between confidence and response time. Further analysis was undertaken (Pearson's r and Spearman's rho[17]) which confirmed that there was no correlation between response time and confidence for any of the variables tested, p>0.05 (see Appendix K for full results).

7.3.3.3 Accuracy and Confidence

For the different levels of sound, repeated-measures t-tests were used to examine the differences between response times for high confidence correct (HCC) and each of three alternatives; low confidence correct (LCC), high confidence incorrect (HCI), low confidence incorrect (LCI). Summary data for each of these comparisons are given in Table 7-18, Table 7-19, and Table 7-20 respectively. It was shown that for 'No Sound'

---

[16] The differences between high confidence and low confidence response times for 'Curve Shape' were non-normally distributed so the *Wilcoxon signed-rank Test* was used.

[17] Spearman's rho was used for 'Curvature' as it was non-normally distributed.

HCC answers were significantly faster than all other types of answer[18]. For 'Curve Shape', HCC was significantly faster than LCI and LCC answers but was the same for HCI. Finally, it was shown that for 'Curvature', HCC answers were not significantly faster than any other type of answer[19].

Table 7-18: Mean RT for HCC and LCC with significance of differences

Sound	RT HCC		RT LCC		Dependent t-test			
	Mean	SD	Mean	SD	t	df	r	sig. (2-tailed)
No Sound	2.39	0.50	4.27	1.94	-3.64	7.00	0.81	0.01
Curve Shape	3.00	0.63	5.51	2.74	-2.59	7.00	0.70	0.04
Curvature[4]	1.35	0.36	2.82	1.91				

Table 7-19: Mean RT for HCC and HCI with significance of differences

Sound	RT HCC		RT HCI		Dependent t-test			
	Mean	SD	Mean	SD	t	df	r	sig. (2-tailed)
No Sound	2.39	0.50	2.85	0.76	-4.74	7.00	0.87	0.00
Curve Shape	3.00	0.63	3.17	0.71	-1.76	7.00	0.55	0.12
Curvature	1.35	0.36	1.84	0.73	-2.14	5.00	0.69	0.09

Table 7-20: Mean RT for HCC and LCI with significance of differences

Sound	RT HCC		RT LCI		Dependent t-test			
	Mean	SD	Mean	SD	t	df	r	sig. (2-tailed)
No Sound[3]	2.39	0.50	4.17	1.52				
Curve Shape	3.00	0.63	5.27	1.96	-3.90	7.00	0.83	0.01
Curvature	1.35	0.36	3.60	0.01	-4.70	1.00	0.98	0.13

A series of scatter plots were used to compare confidence against accuracy for each level of sound. Correlation analysis (Pearson's r and Spearman's rho[20]) revealed that there was a significant correlation between confidence and accuracy for 'Curvature', $r_s=0.90$, $p<0.01$. There was no significant correlation for either 'No Sound' or 'Curve Shape', $p>0.05$ (see Appendix K for full results).

[18] The differences between high confidence correct and Low confidence incorrect response times for 'No Sound' were non-normally distributed so the *Wilcoxon signed-rank Test* was used; z=-2.52, p<0.01.

[19] The differences between high confidence correct and Low confidence correct response times for 'Curvature' were non-normally distributed so the *Wilcoxon signed-rank Test* was used; z=-1.34, p>0.05.

[20] Spearman's rho was used for non-normally distributed variables; 'Curvature'.

7.3.4 Interaction

7.3.4.1 Style

The frequency of each interaction style was calculated for the different levels of sound and is illustrated as a percentage in Figure 7-30. It can be seen that for each sound type the breakdown of styles used was different. A chi-squared test confirmed that there was a significant association between sound and the type of interaction used, $\chi^2(4, n=864)=713.25$, $p<0.01$. By examination of the standardised residuals it was revealed that for 'No Sound' there was significantly more 'Sweep' style than expected, but significantly less 'Point' or 'Both' styles (z=8.9, z=-8.5 and z=-6.6, respectively). For 'Curve Shape' there was significantly more use of 'Both' and as expected use of 'Sweep', but there was significantly less use of the 'Point' style (z=11.9, z=-1.5 and z=-7.0, respectively). Finally, for 'Curvature' there was significantly more use of 'Point', but significantly less of 'Sweep' and 'Both' styles (z=15.5, z=-7.4 and z=-5.3, respectively). From this it is possible to conclude that 'No Sound' is predominantly associated with 'Sweep', 'Curve Shape' with both 'Sweep' and 'Both', and 'Curvature' with 'Point'.

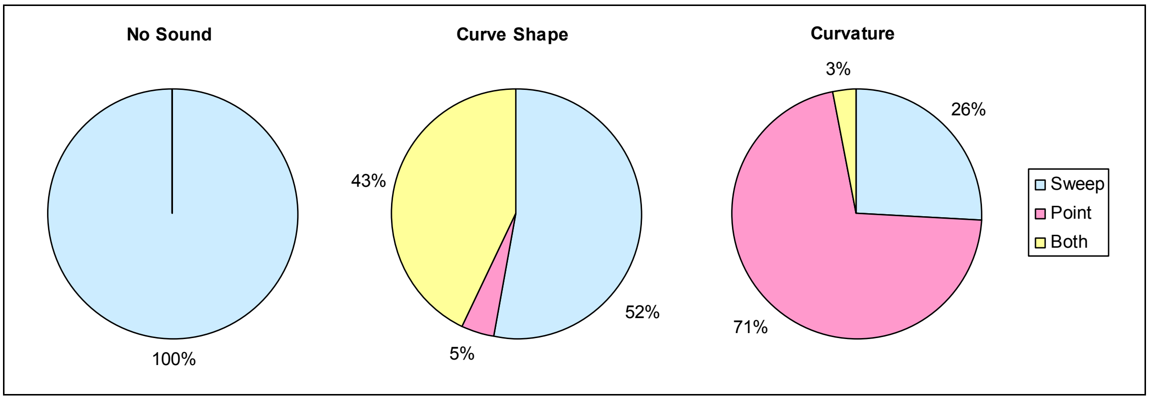


Figure 7-30: Interaction styles (%) observed for each sound condition

7.3.4.2 Exploration Duration

A repeated-measures ANOVA compared the effect of location and sound on exploration duration (see Figure 7-31). This revealed that there was a significant effect of location, $F(2,14)=9.63$, $p<0.01$, partial $\eta^2=0.58$. Pre-planned contrasts revealed that exploration duration was significantly longer at the 'Ends' of the blocks (1.25s, SD=0.43) compared to the 'Middle' (0.81s, SD=0.34), $F(1,7)=5.81$, $p=0.047$, partial $\eta^2=0.45$. It also showed that there was no significant difference in exploration duration between the 'Middle' and the 'Slopes' (0.62s, SD=0.29), $F(1,7)=2.24$, $p=0.178$, partial $\eta^2=0.24$.

There was a significant effect of sound, $F(2,14)=45.53$, $p<0.01$, partial $\eta^2=0.87$. Pre-planned contrasts showed that exploration duration was significantly longer for ‘No Sound’ compared to ‘Curvature’, but there was no significant difference between ‘No Sound’ and ‘Curve Shape’, $F(1,7)=0.003$, $p=0.96$, partial $\eta^2=0.00$, and $F(1,7)=49.19$, $p<0.01$, partial $\eta^2=0.87$, respectively. There was no significant effect of stimulus on exploration duration, $F(2,14)=2.45$, $p=0.12$, partial $\eta^2=0.26$.

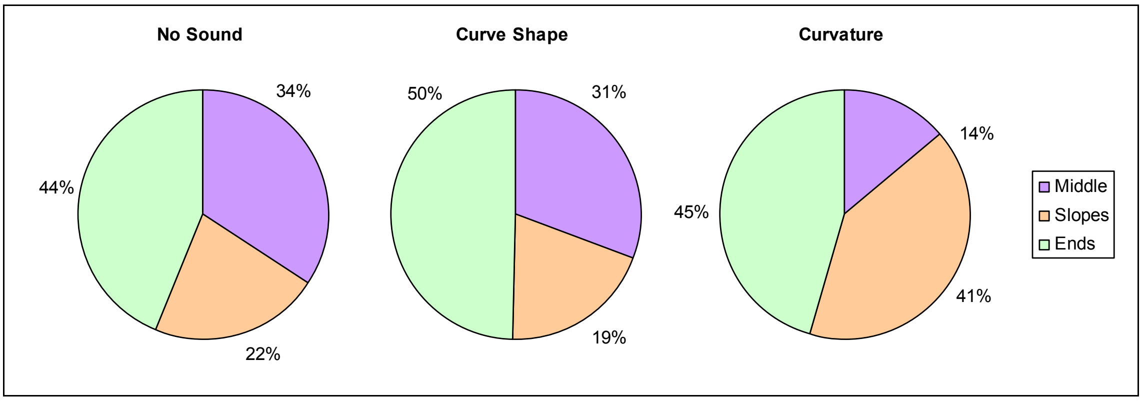


Figure 7-31: Effect of location and sound on exploration duration (%)

There was an interaction between location and sound on exploration duration, $F(2.40,16.77)=6.48$, $p<0.01$, partial $\eta^2=0.48$. Pre-planned contrasts revealed that the decrease in exploration duration for 'Curvature' (compared to 'No Sound') was significantly greater in the middle than on the slopes of the stimulus, $F(1,7)=10.72$, $p=0.014$, partial $\eta^2=0.60$. This can be seen in Figure 7-31 with the increase in percentage duration on the slopes and decrease in the middle between the ‘No Sound’ and ‘Curvature’ conditions. There was no interaction between stimulus and location, or three-way interaction between sound, location, and stimulus, $F(4,28)=2.13$, $p=0.10$, partial $\eta^2=0.23$, and $F(8,56)=1.77$, $p=0.11$, partial $\eta^2=0.20$, respectively.

7.3.5 User Experience

7.3.5.1 Ease, Difficulty, and Helpfulness

Participants were asked to state which of ‘No Sound’, ‘Curve Shape’, and ‘Curvature’ was the easiest and most difficult to use. It was found that there was a significant association between sound and ease of use, $\chi^2(1, n=24)=16.67$, $p<0.01$. 92% of participants found ‘Curvature’ easier to use than either ‘Curve Shape’ (8%) or ‘No Sound’ (0%). In relation to difficulty, there was also a significant association with sound, χ^2(2, n=24)=14.25, p<0.01. For this 67% of participants found ‘No Sound’ more difficult to use than either ‘Curve Shape’ (29%) or ‘Curvature’ (4%).

Participants were also asked whether the presence of the ‘Curve Shape’ and ‘Curvature’ sounds were helpful. For ‘Curve Shape’ there was no significant association between sound and helpfulness with roughly similar numbers saying it was (46%) as was not helpful (54%), $\chi^2(1)=0.17$, $p>0.05$. For ‘Curvature’ 100% of participants indicated that it was helpful. A further Chi-squared test was used to contrast these results. This found that there was a significant association between helpfulness and sound, $\chi^2(1, n=48)=17.83$, $p<0.01$. Examination of standardised residuals revealed that there were significantly more participants who felt that ‘Curve Shape’ was not helpful (z=2.5), and significantly, no participants felt that ‘Curvature’ was not helpful (z=-2.5). In terms of feeling that the sound was helpful, the number of participants expressing this about ‘Curve Shape’ (Z=-1.6) was not significantly different to those expressing this about ‘Curvature’ (z=1.6).

7.3.5.2 Concentration

Participants were asked if, relative to ‘No Sound’, they had to concentrate more or less with ‘Curve Shape’ and ‘Curvature’. For ‘Curve Shape’ there was found to be no significant association between level of concentration and sound, with roughly similar numbers saying they had to concentrate more (62%) or less (38%), $\chi^2(1)=1.5$, $p>0.05$. For ‘Curvature’ 100% of participants indicated that they needed to concentrate less than with ‘No Sound’.

A further Chi-squared test was used to contrast these results in order to reveal the relative performance of ‘Curve Shape’ and ‘Curvature’. This found that there was a significant association between the level of concentration and sound, $\chi^2(1, n=48)=21.82$, $p<0.01$. Examination of standardised residuals revealed that there were significantly more participants who felt that ‘Curve Shape’ required more concentration than ‘No Sound’ (z=2.7), and significantly, no participants felt that ‘Curvature’ required more concentration (z=-2.7). In terms of feeling that the sound required less concentration than ‘No Sound’, the number of participants expressing this about ‘Curve Shape’ (Z=-1.8) was not significantly different to those expressing this about ‘Curvature’ (z=1.8).

7.3.5.3 Use and Conflict of Senses

For each sound condition participants were asked which senses they had used to make a judgement. The proportions of each sense or senses used are shown in Figure 7-32.

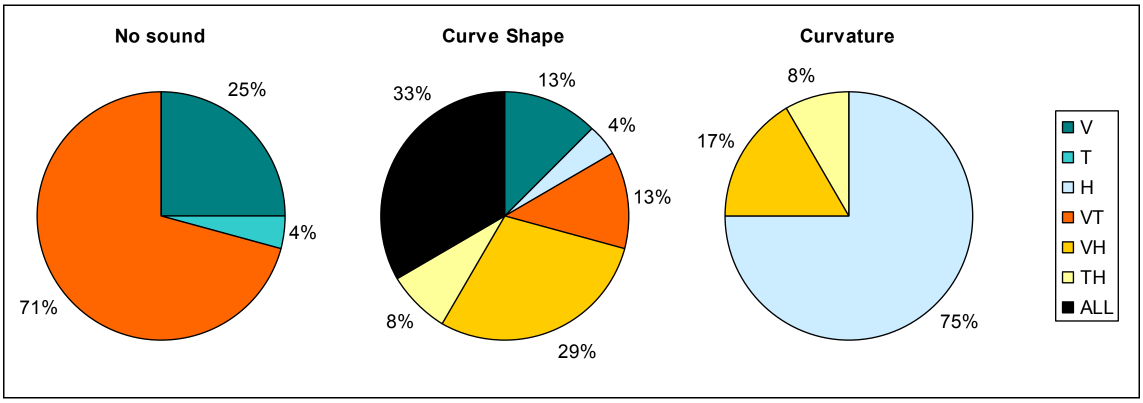


Figure 7-32: Proportion of senses used for each sound condition

For 'No Sound' and 'Curvature' there was a significant difference between expected and observed frequencies, so it can be conclude that all senses were not used equally, $\chi^2(2, n=24)=16.75$, and $\chi^2(2, n=24)=19.00$, respectively and $p<0.01$ for both. The most frequently used senses for 'No Sound' was vision and touch combined (71%), whilst for 'Curvature' hearing (75%) was the predominant sense. Both these conditions used a maximum of three senses or combination of senses. This is in contrast to 'Curve Shape' where a total of six were used, the only omission being the use of touch only. There was no significant difference between expected and observed frequencies for this condition, so we can conclude that for 'Curve Shape' all sense or combination of senses were used equally, $\chi^2(5, n=24)=10.00$, $p>0.05$.

Participants were also asked if they had received the same or conflicting information from the sense that they had used in each sound condition (see Figure 7-33). For each sound condition separately, it was found that there was no significant differences between the observed and expected frequencies, $\chi^2(1, n=24)=0.67$, $\chi^2(1, n=24)=1.50$, and $\chi^2(1, n=24)=2.67$, respectively and $p>0.05$ for all. It can be concluded that the experience of conflict and similarity for each senses was roughly equal. A further Chi-squared test was used to contrast the results across the levels of sound. This found that there was no significant association between the similarity/conflict and sound, $\chi^2(2, n=48)=4.36$, $p>0.05$.

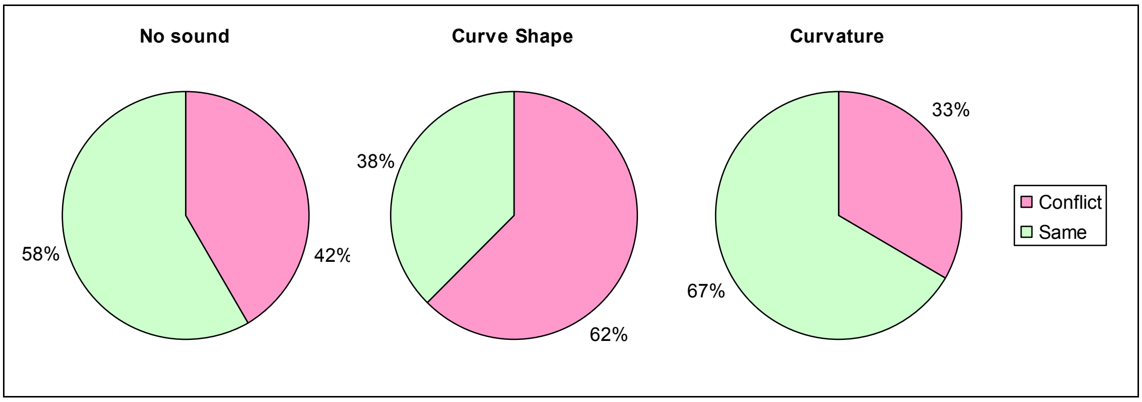


Figure 7-33: User experience of conflict or similarity of senses for each sound condition

The qualitative responses to questions 10 to 12 revealed that participants did experience modalities differently and that this led to conflicting perceptual information (see Appendix L for full transcript). In the 'No Sound' condition participants remarked that there was a difference between what they saw and what they felt, and some felt that they received more information from vision. One participant commented that the conflict in information made them less secure about their visual judgements. Another participant went on to say that they trusted vision over their sense of touch. This sense of conflict increased in the 'Curve Shape' condition. Here most participants felt that what they heard conflicted with what they saw or felt. For one participant this resulted in being in 'two minds' and meant they felt it took longer to respond. For some participants this conflict was resolved by a reliance on vision. Finally, for the 'Curvature' condition whilst many participants reported a conflict between sound and vision, others stated that they only attended to the sound (so reported on conflict). Those that did experience a conflict between vision and sound stated that they 'went with sound'.

7.3.5.4 Workload

NASA-TLX was used to provide an estimation of the subjective workload experienced by each participant. The measure was applied after the end of each evaluation thus giving nine scores; one for each sound/stimuli combination. Figure 7-34 shows the scores for the different sound conditions. A repeated-measures ANOVA compared the effect of sound on the NASA-TLX Scores. This revealed that there was a significant effect of sound on the score, $F(2,14)=4.99$, $p=,0.023$ partial $\eta^2=0.42$. Pre-planned contrasts showed that there was no significant effect between 'No Sound' (47.08, SD=13.06) and either 'Curve Shape' (48.19, SD=16.30) or 'Curvature' (40.76, SD=11.54), $F(1,7)=0.37$, $p=0.56$, partial $\eta^2=0.05$, and $F(1,7)=5.32$, $p=0.54$, partial $\eta^2=0.43$, respectively., The significant result must therefore indicate a difference between scores for ‘Curve Shape’ and ‘Curvature’, a post-hoc pairwise t-test confirmed this to be the case (t=2.55, p=0.04).

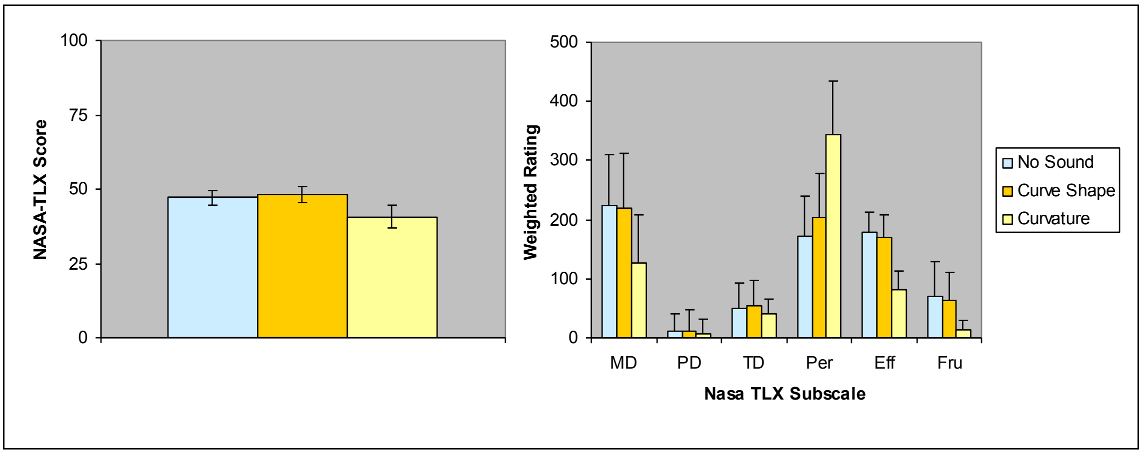


Figure 7-34: Workload experienced for sound and stimulus conditions

Each TLX Score was derived from a number of factors; Mental Demand (MD), Physical Demand (PD), Temporal Demand (TD), Performance (Per), Effort (Eff) and Frustration (Fru). A breakdown of the sub-factors influencing workload for different sounds are shown in Figure 7-34. This gives an idea of the relative contribution each factor made to the participants’ sense of workload. A series of six repeated-measures ANOVA compared the effect of sound on the various sub-factors. These revealed that for Mental Demand, Performance, and Effort there was a significant effect of sound on weighted rating, $F(2,14)=6.39$, $p=0.011$, partial $\eta^2=0.47$, $F(2,14)=14.67$, $p<0.01$, partial $\eta^2=0.68$, and $F(2,14)=30.72$, $p<0.01$, partial $\eta^2=0.81$, respectively. Pre-planned contrasts for Mental Demand and Effort showed that the ratings for ‘Curvature’ (127, 81) were significantly less than for ‘No Sound’ (225, 180), $F(1,7)=6.16$, $p=0.042$, partial $\eta^2=0.47$, and , $F(1,7)=40.84$, $p<0.01$, partial $\eta^2=0.85$, respectively. For Performance they showed that ‘Curvature’ (344) was significantly higher than ‘No Sound’ (171), $F(1,7)=15.09$, $p<0.01$, partial $\eta^2=0.68$. There was no significant effect of Physical Demand, Temporal Demand, or Frustration on sound, $p>0.05$.

7.4 Discussion (Study 3)

7.4.1 Perception

The experiment sought to address the following questions in relation to perception:

- Does sound affect perceptual acuity, and is this effect dependent upon the sonification used?
- Does this acuity vary depending upon the gradient level examined?

The results showed that there was no significant difference between Weber Fractions (% change) for ‘No Sound’ and ‘Curve Shape’. There was however a difference between ‘No Sound’ and ‘Curvature’. The Weber Fraction for ‘Curvature’ was 2.48%; this is a factor of roughly five times better than the acuity achieved in the ‘No Sound’ and ‘Curve Shape’ conditions. It is therefore possible to conclude that sound does improve perceptual acuity when judging gradient differences between shapes. However, this improvement was highly dependent upon the sonification method used. It is clear from the results that ‘Curvature’ provided improvement, whereas ‘Curve Shape’ was the same as using no sound at all. The question that arises is: why should this be the case? A possible answer is that the ‘Curve Shape’ sonification provides similar information to the haptic channel, in that it traces the shape of the curve. This feedback is continual and finely graduated (many levels of sound), whereas in contrast the ‘Curvature’ sonification provides discrete feedback which is grossly graduated (two levels of sound). The ‘Curvature’ sonification therefore makes the task of perceiving curvature change much simpler, and effectively becomes a matter of judging the difference between two tones. This simplification may account for the improvement in acuity.

There should be some caution in accepting the 2.45% Weber Fraction for ‘Curvature’ since the underlying data was not ideal for generating an accurate difference threshold. A more robust estimate would be a Weber Fraction of 3.25%. This figure is based on the fact that for the smallest difference between stimuli, 3.25%, there was almost 100% accuracy from all participants. This means that the Weber Fraction is somewhere below 3.25%. So a conservative assessment of the improvement for ‘Curvature’ would be a factor of about four times. Even at this conservative level this was a substantial improvement over ‘No Sound’ and is a strong indicator of the potential of applying sound in this context. Given that interaction using ‘Curvature’ tends to be predominantly auditory (see 7.4.3 for a discussion of this) then there is the potential that acuity for this type of task could be in the region of 0.1%, which is the level of acuity for pitch differentiation (Wolfe, Kluender et al. 2006).

The study revealed that there was no main effect of stimuli or interaction effect between sound and stimuli on acuity. This meant that perceptual acuity remained constant despite changes in the stimulus, so for 'No Sound' and 'Curve Shape' the amount of change required to detect a difference was about 12%, and for 'Curvature' it was less than 3.25%. This percentage remained the same regardless of the stimulus gradient (these ranged from about 2^o to 6^o). This is an important finding[21] as it means that differentiation of gradient change conforms to Weber's Law; that is the amount of change is a constant proportion of the original stimulus. It also indicates that the relationship of stimulus to detection threshold was not disrupted by the addition of sound (there was no interaction effect).

7.4.2 Performance

The experiment sought to address the following questions in relation to performance:

- Is performance (response time, accuracy, confidence) affected by the addition of sound?
- Does the effect of sound vary dependent upon the gradient, or magnitude difference judged?
- Is anyone sonification better than the other?
- Are performance characteristics changed by the use of sound?

The results showed that judgements made using the 'Curvature' sonification were significantly quicker (1.4s), more accurate (99%), and more confident (98%) than with 'No Sound' (2.8s, 75%, and 63%). For 'Curve Shape' judgements were significantly slower (3.2s), had similar accuracy (74%), and were more confident (77%) than with 'No Sound'. It is clear from these results, that other than for confidence, the utility of sound for improving performance is highly dependent upon the sonification method used. For example whilst response times were quicker with 'Curvature', they were slower with 'Curve Shape'. The effect was also mixed in terms of accuracy, where 'Curvature' improves accuracy but 'Curve Shape' shows no boost in performance over 'No Sound'. The only consistent aspect of sound's effect was in improving the confidence of

[21] The significance of this finding and its implications will not be discussed further here, as the focus is on sound. For further discussion see Chapter 9 which explores the prediction of JNDs.

participants in their judgements. However even here there was a differential effect of the types of sonification method used, with ‘Curvature’ performing more highly than ‘Curve Shape’.

The effect of sound on performance was further complicated by the fact that it had an interaction with both stimulus and magnitude difference. The interaction between sound and stimulus saw a different effect on stimulus ‘D’ in comparison to stimuli ‘E’ and ‘F’ between sound conditions. This difference in effect was seen for both the ‘Curve Shape’ and ‘Curvature’ sonifications. For ease of explanation these effects are illustrated in Figure 7-35. The effect was such that sound reduced the differences between stimuli such that there was no longer a significant difference between them. The strength of this effect was similar for response time, but was much stronger for ‘Curvature’ than ‘Curve Shape’ for accuracy. It was also apparent that the general direction of this effect was to decrease performance for ‘Curve Shape’ and increase performance for ‘Curvature’. As with the main effects it is apparent that the ‘Curvature’ sonification is more successful than the ‘Curve Shape’. A similar interaction effect is seen between sound and magnitude difference (see Figure 7-36). However, whilst curvature nullified the effect of differences in magnitude, ‘Curve Shape’ increased these. So the difference that was seen between ‘1x’ compared to ‘8x’ (or ‘4x’) magnitude difference was significantly increased for ‘Curve Shape’. This significance between the conditions appears to be due to a slight reduction in performance for ‘1x’ magnitude difference and an increase in ‘8x’. The ‘Curve Shape’ sonification therefore seems to have a different effect on accuracy depending on the magnitude difference sonified.

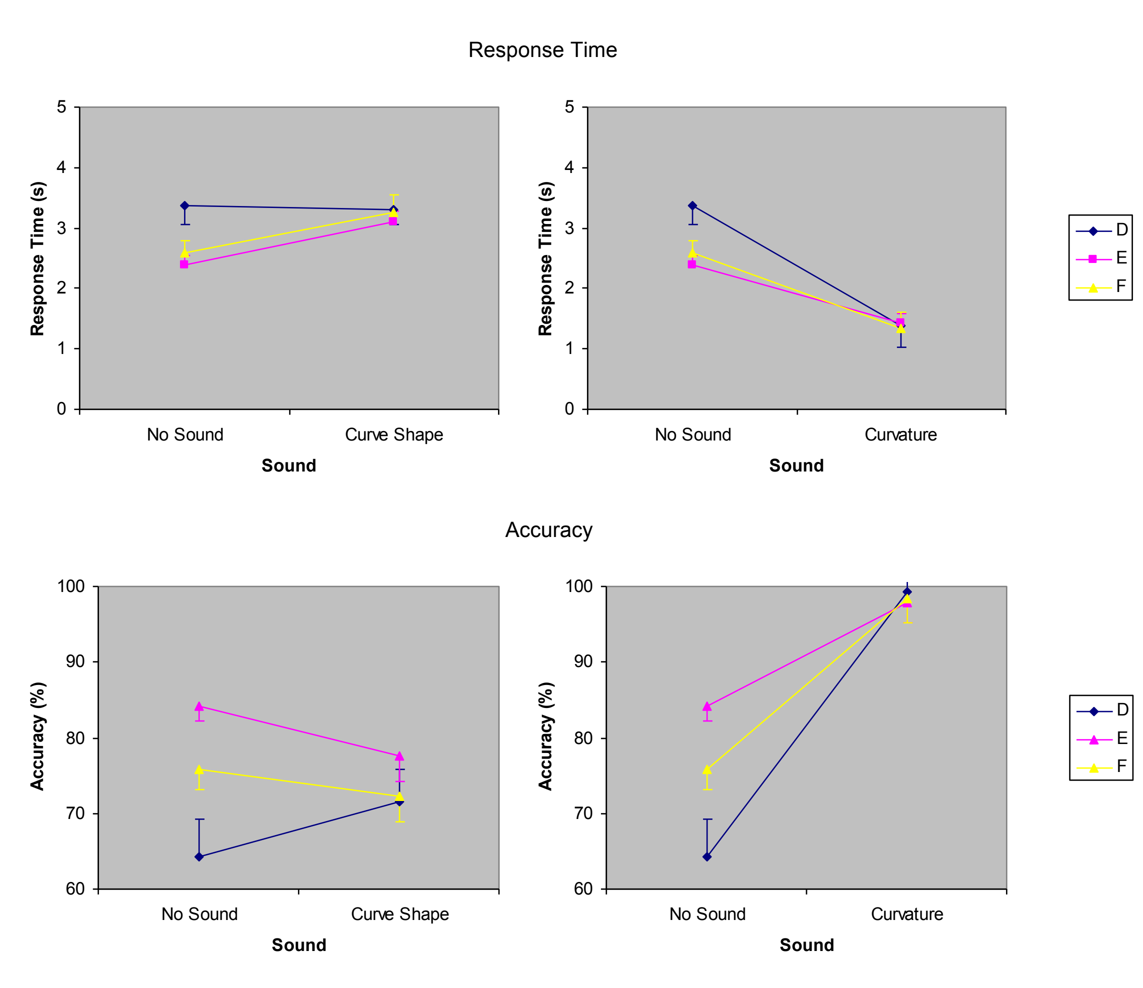


Figure 7-35: The interaction effect of sound is to reduce differences between stimuli

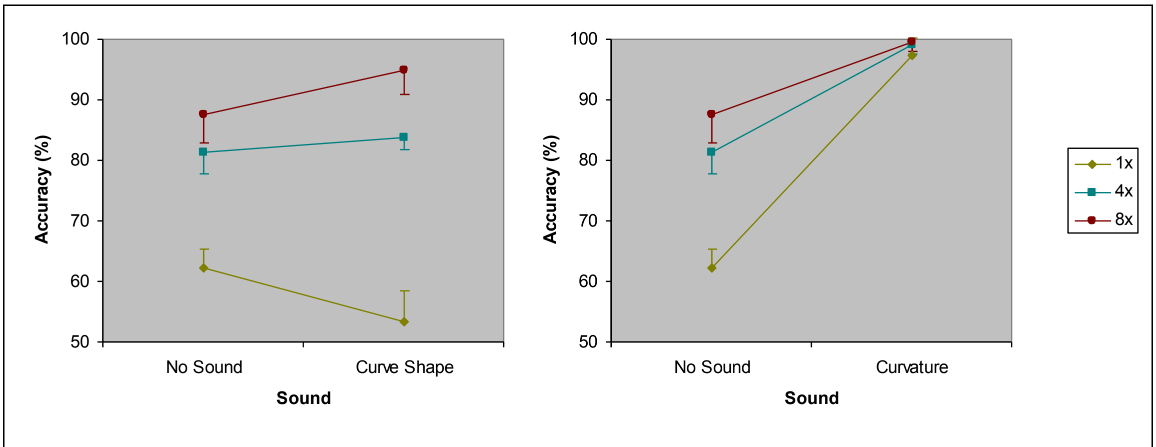


Figure 7-36: The interaction effect of sound with magnitude difference is mixed

This would point to there being some aspect of the sonification that decreased performance for small differences and increased performance for large differences. The likely explanation for this was that for small magnitude differences the rate of change of sound was similar at the standard and comparison ends of the stimulus (see Figure

7-37). Conversely, for large magnitude differences, the rate of change was more pronounced and so easier to judge. For small differences in magnitude the addition of the 'Curve Shape' sound seems to have compounded an already difficult judgment.

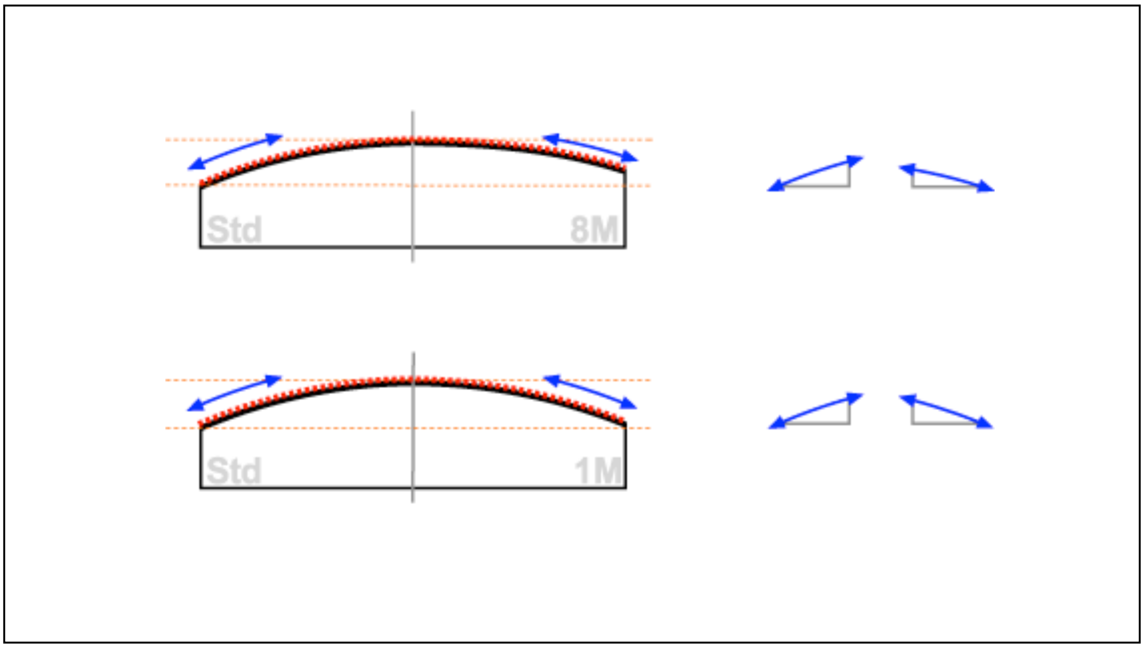


Figure 7-37: Comparison of Sonification differences for '1x' and '8x' Magnitude Difference

Although there were some significant three-way interactions between sound, stimulus, and magnitude difference the effects were very similar to those seen in the two-way interactions discussed above. That is, whatever difference appeared in the 'No Sound' condition was nullified in the 'Curve Shape' and 'Curvature' condition. In reviewing the results for main and interaction effects it has been seen that 'Curvature' sonification consistently outperformed 'No Sound', and that whilst 'Curve Shape' did this to some extent the results are far more mixed. Given this, it follows that some sonification methods are better than others at enhancing performance beyond that were there is no sound, and in the case of this study the 'Curvature' sonification was the most optimum.

A further consideration was whether performance characteristics were affected by the addition of sound (see Table 7-21 and Table 7-22). It was found that with 'Curve Shape' the characteristics were the same as 'No Sound'; that is, response times for correct judgments were faster than incorrect, high confidence judgements were faster than low confidence, and there were no correlations between response time and either accuracy or confidence. When comparing response times for HCC judgments to LCC, HCI, and LCI, it was found that each was significantly different for 'No Sound', and that the only variation for 'Curve Shape' from this was for HCI, which showed no significant difference.

Also, neither ‘No Sound’ nor ‘Curve Shape’ showed a correlation between accuracy and confidence. It can therefore be concluded that there was minimal variation in the characteristics of visual-haptic interaction with the addition of the ‘Curve Shape’ sonification. Conversely, ‘Curvature’ varied in practically every respect from ‘No Sound’. There was no difference in response time between correct and incorrect or high and low confidence judgments, response time for HCC was similar to LCC, HCI, and LCI, and there was a correlation between accuracy and confidence. The only similarity with ‘No Sound’ was the absence of correlation between response time and accuracy or confidence.

The addition of sound therefore may or may not change the underlying performance characteristics of visual-haptic interaction dependent upon the type of sonification used. These results would further suggest that the more successful the sonification method was in terms of performance the more similar the response times were between different types of judgment (e.g. correct versus incorrect). So that it was only with performances that were less optimal that differences in response time were observed between types of judgment. However, since only two sonification methods were used, and of these only one was optimal (at the differences judged), then it would be difficult to justify a wider inference for this result. Even so, this may prove a useful indicator of the point at which performance is starting to be stretched.

Table 7-21: Comparison of significant differences

Study	Modality	Correct RT < Incorrect RT	High Confidence RT > Low Confidence RT	LCC	HCC RT< HCI	LCI
1	Visual-Haptic	✓	✓✓	✓✓	x	✓✓
2	Sound (Cello Curve Shape)	✓	✓✓	✓✓	x	✓✓
3	Visual-Haptic	✓✓	✓✓	✓✓	✓✓	✓✓
3	All Modalities (Curve Shape)	✓	✓	✓	x	✓
3	All Modalities (Curvature)	x	x	x	x	x

Note: significance is shown with ✓ (<0.05) and ✓✓ (<0.01), a ‘x’ means no significant result.

Table 7-22: Comparison of correlations

Study	Modality	Accuracy (%) and RT	Confidence (%) and RT	Accuracy (%) and Confidence (%)
1	Visual-Haptic	x	x	✓
2	Sound (Cello Curve Shape)	✓✓	✓	✓
3	Visual-Haptic	x	x	x
3	All Modalities (Curve Shape)	x	x	x
3	All Modalities (Curvature)	x	x	✓

Note: significance is shown with ✓ (<0.05) and ✓✓ (<0.01), a ‘x’ means no significant result.

Whilst with ‘Curvature’ there is an effect of sound on visual-haptic characteristics, it would also appear that sound characteristics themselves are affected when combined with visual-haptic modalities. In comparing the results of Study 2 (see Chapter 5) with those here, it can be seen that for the ‘Curve Shape’ sonification response time was correlated with both accuracy and confidence. There was also a correlation between accuracy and confidence. However, when the ‘Curve Shape’ sonification was combined with the visual-haptic modality these correlations are no longer apparent. This may be an indication that the visual-haptic modality was dominant in making judgments. Equally this may be a difference inherent in the types of tasks undertaken, and suggests that particular characteristics are not immutable but may be task dependent.

7.4.3 Interaction

The experiment sought to address the following questions in relation to interaction:

- Are particular interaction styles adopted dependent upon the sound used?
- Is the whole of the stimulus explored equally or is exploration focused in some areas more than others?
- Is the extent of interaction (exploration duration) effected by sound and location?

It had been observed in Study 2 that a number of interaction styles, ‘Point’ and ‘Sweep’, had been adopted when exploring the curved shapes with sound. The results from this study showed that there was an association between sound and the type of interaction style adopted. When there was no sound the interaction style was exclusively ‘Sweep’ (100%). This indicates that with no sound feedback the preferred way to haptically assess the difference between the curves was to explore the whole length of the stimuli in a sweeping motion.

It was observed that this interaction changed when sound was introduced. For ‘Curve Shape’ the amount of ‘Sweep’ interaction reduced by about half to 52% and there was now the presence of the ‘Point’ style (5%). In addition to these two styles there was a third type of interaction ‘Both’ (43%) which saw the use of both ‘Sweep’ and ‘Point’ in combination. For ‘Curvature’ the amount of ‘Sweep’ interaction reduced still further to 26%, a quarter of what it had been with no sound and was now significantly less than expected. The use of ‘Both’, which had been equally prominent with ‘Sweep’ for ‘Curve Shape’, reduced to 3% and was also significantly less than expected. The main style adopted for interaction with ‘Curvature’ was ‘Point’ being used for 71% of judgements, and was significantly more than expected.

The results clearly show a change in interaction style when sound is used. They also show a difference in interaction dependent upon the type of sonification used. There were two clear preferences shown; ‘Sweep’ for ‘No Sound’ and “Point’ for ‘Curvature’. These choices of interaction seem optimised for the context. For ‘No Sound’ the only way to judge the stimulus is to sweep along the curve (either visually or haptically) in order to determine the level of curvature. For ‘Curvature’ the task is simplified in that the sound gives feedback as to the curvature value each side of the stimulus, therefore minimal interaction is required as touching a single point each side of the stimulus will give the necessary feedback. Most participants quickly realised this hence the dominance of the ‘Point’ style. Even so around 26% of participants still preferred to use the ‘Sweep’ style, which may indicate a continued desire for some haptic feedback[22].

The interaction observed for ‘Curve Shape’ shape seems to be something of a half-way house as ‘Sweep’ and ‘Both’ were equally dominant. The use of ‘Sweep’ reflects the fact that the ‘Curve Shape’ sonification mimicked the shape of the curve being explored, and therefore provided similar feedback as the haptic and visual channels and so was explored in a similar way. However, the equal dominance of the ‘Both’ style, which used ‘Sweep’ and ‘Point’ in tandem, indicates that the presence of sound was affording other ways to judge differences in the stimulus. It may also indicate that there was a desire to reduce the more complex continual feedback inherent in the ‘Curve Shape’ sonification and copy the simpler feedback found with the ‘Curvature’ feedback by using a ‘Point’ style of interaction. This would reduce the feedback from a multiplicity of tones to a single tone which would have been far easier to make a judgment on (as indicated by the performance results discussed above). However, the fact that this was not adopted on its own (only 5% of interaction was ‘Point’ only) would suggest that whilst there were benefits in using a ‘Point’ style it was not entirely appropriate to this type of sonification method.

It had been observed in Study 1 (see Chapter 3 section 3.4.2) that, when adjusted for length, the response times for the longer stimulus were quicker than those for the short.

[22] Equally this may be due to experiment conditions

It was suggested this may be due to exploration not being equally concentrated across the whole of the stimulus. Observations in this study showed that exploration at the ends (1.25s) of the block were about 53% more than in the middle (0.81s). It was also observed that the exploration duration in the middle was greater than that of the slopes (0.62s). This means that exploration across the stimulus was not evenly distributed but was significantly focused at the ends and middle. This would explain the discrepancy found in study 1, and would suggest that stimulus width at the scales examined does not affect response time. This is further supported by the lack of effect on exploration duration of stimulus.

There was some interaction between sound and location. This manifested as a significant decrease in exploration of the middle (compared to slopes) for ‘Curvature’ compared to ‘No Sound’. This is likely to be an artefact of the change from a ‘Sweep’ style in the ‘No Sound’ condition to a ‘Point’ style in the ‘Curvature’ condition. For ‘No Sound’ the ‘Sweep’ would pass through all locations of the block, with an increase in duration for the ends, then middle, and least for the slopes. However, for the ‘Curvature’ sonification the participant’s concern is sounding out a point on either side of the stimulus, with no need to touch the middle. Therefore the increase in duration for the slopes and decrease in the middle of the stimulus can be seen as characteristic of the ‘Point’ style and is likely to explain the interaction effect.

Exploration duration was also effected by sound. The results showed that exploration for ‘Curvature’ (0.32s) was significantly less than that for ‘No Sound’ (1.18s). It was also found that ‘Curve Shape’ (1.18s) was not significantly different from ‘No Sound’. However, this latter result is problematic as it contradicts an earlier finding from the performance results that response time for ‘Curve Shape’ (3.23s) was significantly slower than ‘No Sound’ (2.78s). Given that both of these results are timed from the same trials there is a need to account for this discrepancy. To understand what may have happened here we need to understand what was timed in each case. Response time was taken from the moment the participant was asked to commence the trial until they answered ‘more’ or ‘less’. Exploration duration was automatically logged, but it only logged the time that the participant was in contact with the stimulus. So what exploration duration describes is purely contact time, whilst response time accounts for this plus time not spent in contact with the stimulus. Table 7-23 shows the response times and exploration durations, it also shows the difference between these two. This difference in time could be accounted for by the time it took to reach the stimuli, or to speak out the response, or both. However, if this was the case then it is unlikely that the differences would vary. In fact the differences for the sound conditions ranged from 1.07s to 2.04s, whereas the differences for the stimuli were more similar ranging from 1.51s to 1.63s. This suggests that for some sound conditions there was something else happening other than the time required to reach for the stimuli and give a response. What this delay may partially represent was the time required to process the feedback. For 'Curvature' this was minimal as one second was just sufficient to reach to the block and state a response, whereas for 'Curve Shape' there appears to be a whole extra second of 'process' time. This may account for the significant difference between 'Curve Shape' and 'No Sound', and helps to explain the discrepancy between response time and exploration duration. The delay time suggests that for the 'Curve Shape' sonification additional process time was required and that it may not have been as intuitively understood by the participants as the 'Curvature' sonification.

Table 7-23: Comparison of response time (RT) and exploration duration (ED)

Variable	RT	ED	Delay
No Sound	2.78	1.18	1.60
Curve Shape	3.23	1.18	2.04
Curvature	1.39	0.32	1.07
Stimulus D	2.69	1.06	1.63
Stimulus E	2.31	0.80	1.51
Stimulus F	2.40	0.83	1.57
All	2.47	0.89	1.57

Note: The delay is the difference between RT and ED

7.4.4 User Experience

The experiment sought to address the following questions in relation to user experience:

- What are participant's subjective experiences of using sound to judge differences?
- Do the senses used change in the presence of sound? Are the senses perceived as providing conflicting or similar information?
- Is their mental workload increased by the addition of sound, and is this dependent upon the particular sonification used?

The results showed that 92% of participants found it easier to make judgements with 'Curvature', and they found most difficulty making them with 'No Sound' (67%). Although almost a third (29%) stated that 'Curve Shape' was most difficult to use. This latter finding concurs with the even split in view as to whether 'Curve Shape' was helpful or not (46% to 54%). There was no such split in view for 'Curvature' where 100% of participants believed it to be helpful. The views expressed towards ease, difficulty, and helpfulness would suggest that there was a positive user experience in relation to 'Curvature', and that opinion towards 'Curve Shape' was much more mixed. It is also clear that whilst 'No Sound' was considered by the majority of participants to be most difficult, a high proportion of participants also felt this about 'Curve Shape'. The mixed view of 'Curve Shape' is also reflected in participant's feelings about their levels of concentration in comparison to 'No Sound'. Approximately two-thirds of the participants felt that they concentrated more (62%), and about a third less (38%) with 'Curve Shape'. Conversely, for 'Curvature' all participants felt that they concentrated less than for 'No Sound'. Overall, it would appear that participants had a more beneficial experience using 'Curvature' as opposed to 'Curve Shape', but that both sounds were largely preferable to using no sound at all.

The use of senses varied depending on whether sound was present or absent, and what type of sonification method was used. For 'No Sound' participants predominantly judged differences using both vision and touch (71%), although a quarter used vision only. When sound was present the combination of vision and touch dramatically decreased, as in 'Curve Shape' (13%) or disappeared altogether as with 'Curvature'. The dominant sense for 'Curvature' was hearing, with three-quarters of participants using this only to the exclusion of all other senses or combinations of senses. This shift from one dominant sense (or combination) was not apparent for 'Curve Shape' as no sense was used significantly more than the others. The most commonly used senses were vision with hearing (29%), and all combined (33%). Notably the only sense combination not used was touch only, otherwise all other senses or combinations were evident for 'Curve Shape'. This may indicate that participants experienced some difficulty in relation to 'Curve Shape' and so no one sense or combination presented itself as an optimum choice. Equally 'Curve Shape' may have afforded more opportunity to use combinations of senses than other conditions. However, in the light of performance results and the comments in relation to difficulty it is more likely that participants experienced 'Curve Shape' as problematic and this has manifested itself in a lack of clear choice between senses or combination of senses. What is clear is that the choice of sense changed when sound was present, and this choice may have been dependent upon how easily the sound was understood in communicating difference. There was indication that in the case of 'Curve Shape' this was not clear enough to rely solely on the sense of hearing, and this was supplemented with reference to other senses.

The quality of the user experience was also dependent upon whether the information provided by each of the senses provided coherent feedback. The participants were asked if the senses that they had used were the same or conflicted in the information they provided. The results showed that there was no significant difference between those thinking that the senses were the same, and those reporting conflict. However, it is worth noting that for 'Curve Shape' nearly two-thirds of participants felt that the senses provided conflicting information, whereas for 'Curvature' this figure was about a third. Whilst these observations are not statistically significant, given the other observations that we have, this would cohere with a sense that 'Curve Shape' was problematic to use.

Despite some of the short-comings outlined above, the addition of sound did not add to the mental workload of the participants. The NASA-TLX scores for 'Curve Shape' (48) and 'Curvature' (41) were not significantly different to that for 'No Sound' (47). This equivalence is a little surprising given that participants felt that 'No Sound' was the most difficult to use and so one might expect to see this reflected in the workload score. However, this score is made up of a number of sub-factors that represent facets of the overall workload; Mental Demand, Physical Demand, Temporal Demand, Performance, Effort, and Frustration. For all sub-factors there was found to be no significant difference between 'No Sound' and 'Curve Shape'. There was however a significant difference between 'No Sound' and 'Curvature'. It was found that 'Curvature' had a lower score for Mental Demand (127) and Effort (81), and an increased rating for Performance (344) compared to 'No Sound' (225,180, and 171 respectively). So whilst the NASA-TLX score indicates that the addition of sound did not increase overall workload, it also appears that in some areas 'Curvature' did show some alleviation of individual workload factors (compared to 'No Sound'). The results also showed that 'Curvature' had a significantly better NASA-TLX score than 'Curve Shape' with a 7-point difference between the two. This fits with other subjective results which show 'Curvature' as the easiest to use.

7.5 Chapter Summary

The study detailed in this chapter set out to understand the effects of sound on perception, performance, interaction, and the user's experience of this. The results indicate that sound can enhance both perception and performance, although this is dependent upon the sonification method used. It was also found that interaction was affected by not only the presence of sound, but also the sonification used. It was seen that users adapt their interaction based upon this. They also show a preference for interacting with the ends and middle of the stimuli, rather than the slopes. The users' experience showed that the addition of sound was largely advantageous, although this view is stronger for 'Curvature' than for 'Curve Shape' which had more mixed responses.

Overall the addition of sound saw increased perception of differences, performance, and a better user experience; however this was heavily dependent upon the sonification method used. The use of ‘Curvature’ produced far better results across all areas investigated than ‘No Sound’, whilst for the most part ‘Curve Shape’ showed little if any improvement over this.

7.6 Related Chapters

This was the final Study undertaken as part of this research programme. The findings of this study and those of the previous two studies were used to inform a framework for the design and evaluation of haptic/sound interfaces (see Chapter 9). Further analysis of the ‘No Sound’ data was undertaken to test the predictive model proposed in Chapter 4 and to further understand the predictability of performance (see Chapter 8). The findings here also contribute to the general discussion (see Chapter 10).

Chapter 8: Prediction of Perception and Performance

8.1 Introduction and Rationale

This chapter complements the work discussed in Chapter 4. In that chapter it was suggested that the development of a model for predicting threshold gradient would be helpful in providing:

- the correct amount of feedback for visual-haptic interfaces
- a quantitative evaluation framework
- a cost-effective way to determine difference thresholds (compared to psychophysical method)

Through examination of the literature it was proposed that there was a relationship between the stimulus gradient and the threshold gradient. This was investigated through regression analysis and it was found that stimulus gradient had a highly significant relationship to threshold gradient for haptic, visual, and visual-haptic modalities. This chapter outlines analyses which were undertaken in order to validate the model proposed in Chapter 4. It also explores the relationship (if any) between performance and gradient (or other stimulus dimension) in order to assess the feasibility of predicting performance.

8.1.1 Experimental Construct

The model outlined in Chapter 4 identified gradient as a predictor of threshold for curve shape differences (through analysis of Study 1 data). With this in mind, Study 3 was partly designed to confirm this relationship and to assess the accuracy of predictions made from the model. The study used three stimuli (‘D’, ‘E’, and ‘F’) which had some dimensions in common whilst others varied (see Chapter 7 section 7.2.4 for description of these stimuli). For the purposes of confirming the effect of gradient on difference threshold two of the stimuli (‘E’ and ‘F’) were designed to have the same gradient whilst differing in other respects (curvature, and width). So, if the gradient thresholds for stimulus ‘D’ were found to be significantly different from those of ‘E’ and ‘F’, and those for ‘E’ and ‘F’ had no significant difference from each other, then it was possible to infer that gradient was the active dimension in difference perception (see Chapter 7 section 7.1.2 for an explanation of the logic behind this and other inferences). In addition, the thresholds produced in Study 3 could be compared to the predictions made prior to the study in order to assess the level of accuracy for the model.

This experimental construct was also designed to be helpful in disambiguating the effect of stimuli dimensions on performance. In Study 1, because of the lack of control over stimuli dimensions, it was difficult to know whether the decrease in response time between stimuli ‘B’ and ‘C’ was due to curvature, width, a combination of both or some unknown difference (see Chapter 3 section 3.2.4 for a description of these stimuli). By ensuring that some dimensions were held in common and others differed, a series of inferences could be made as to the dimension of difference from the ANOVA contrasts (see Chapter 7 section 7.1.2 for details).

8.1.2 Aims

Prior to the start of Study 3 three research questions were posed for the further analysis that would be undertaken on the visual-haptic data, these were:

1. Do stimuli with different gradients produce different JNDs? Do stimuli with the same gradient have the same JND? Do Weber Fractions vary with gradient? How accurately did the model predict the actual threshold gradients observed in Study 3? Does this support the theory that stimulus gradient is a predictor of threshold gradient?
2. Do different dimensions of the stimulus (gradient, curvature, width) have an effect on performance (response time, accuracy, confidence)?
3. Is there a predictable relationship between performance and stimulus for varying magnitude differences? Is this relationship similar for visual-haptic data from Study 1? Can the result from Study 1 and Study 3 be combined to improve the generalisation of the model?

8.2 Further Analysis of Study 3 Data

8.2.1 Predicted Threshold Gradients

The predictive model developed in Chapter 4, was used to calculate the threshold gradients for the stimuli used in Study 3. The accuracy and viability of the model could then be assessed through comparison with the actual gradients observed in Study 3. For each stimulus the height was divided by the width to give the ‘Stimulus Gradient’. The predictions were then calculated using the Visual-Haptic coefficient (see Equation 1) and are shown in Table 8-1.

Equation 1: Visual Haptic Model Coefficient

$$y = 1.285x - 0.004$$

Where y=threshold gradient and x=stimulus gradient

Table 8-1: Predicted threshold gradients for study 3

Stimulus	Stimulus Gradient	Predicted						
		Threshold Gradient	95% CI		JND	WF	95% CI	
			Upper	Lower			Upper	Lower
D	0.037	0.043	0.041	0.045	0.006	16.91	12.15	21.69
E	0.111	0.139	0.136	0.141	0.028	25.20	22.99	27.41
F	0.111	0.139	0.136	0.141	0.028	25.20	22.99	27.41

8.2.2 Design and Hypothesis

Study 3 was a repeated-measures design, and applied the psychophysical method of constant stimuli to establish the difference threshold for gradient in No Sound, Curve Shape, and Curvature conditions. However, the data that were used for further analysis were the visual-haptic or 'No Sound' data.

The further analysis sought to examine the effects of two independent variables; Stimuli (D, E and F), and magnitude difference (3.25, 6.5, 13, and 26 percent) on perception (gradient threshold, JND, and Weber Fractions) and performance (response time, accuracy, and confidence). In order to answer the research questions posed (see 8.1.2) the following hypotheses were examined:

8.2.2.1 Perception (H1-H2)

H1: There is an effect of stimulus on JND.
H2: There is an effect of stimulus on Weber Fractions.

8.2.2.2 Performance (H3-H11)

H3-H5: There is an effect of stimulus on response time; accuracy; confidence.
H6-H8: There is an effect of magnitude difference on response time; accuracy; confidence.
H9-H11: There is an interaction effect between stimulus and magnitude difference on response time; accuracy; confidence.

8.2.2.3 Predictability (H12-H19)

H12-H14: Stimulus dimensions (for stimuli DEF) and magnitude difference account for a significant proportion of the variance in response time; accuracy; confidence.
H14-H16: Stimulus dimensions (for stimuli BC) and magnitude difference account for a significant proportion of the variance in response time; accuracy; confidence.
H17-H19: Stimulus dimensions (for stimuli BCDEF) and magnitude difference account for a significant proportion of the variance in response time; accuracy; confidence.

8.2.3 Data Analysis

The performance data (response time, accuracy, and confidence) were found to be non-normally distributed for some variables; either positive or negative skew, which is usual for these types of data. This was not a problem as ANOVAs are considered to be a robust method of statistical analysis (Davies 1956; Field 2009)(see Appendix F for discussion of this). A number of pre-planned contrasts were used in the analysis of these data. For stimuli, all were contrasted with ‘D’ and ‘F’ to ensure that all were contrasted with each other. For magnitude difference all were contrasted with ‘1x’ which was likely to show the lowest performance and so improvement from this could be gauged. For other statistical tests the appropriate parametric analysis methods were used for non-normal variables (the tests used are noted within the results).

8.3 Results

This section details all results in detail, a summary of the results in relation to the hypotheses can be found in Appendix D.

8.3.1 Perception

For each stimulus a mean threshold gradient was calculated from the individual gradients of each participant, as shown in Table 8-2 (see Appendix M for calculation of these individual thresholds and model statistics). In addition, mean just noticeable difference (JND) and Weber Fractions (WF) were calculated (see Table 8-2).

Table 8-2: Threshold gradient, JND and Weber Fractions for each stimulus

P-ID	Stimulus D			Stimulus E			Stimulus F		
	Threshold	JND	WF	Threshold	JND	WF	Threshold	JND	WF
P1	0.042	0.005	14.065	0.124	0.013	11.354	0.125	0.014	12.388
P2				0.124	0.013	12.113	0.124	0.013	11.974
P3	0.041	0.004	11.922	0.124	0.013	11.985	0.123	0.012	11.033
P4	0.042	0.005	12.592	0.123	0.012	11.211	0.124	0.013	11.722
P5	0.041	0.004	11.732	0.123	0.012	11.202	0.124	0.013	11.456
P6	0.041	0.004	11.054	0.124	0.013	11.541	0.124	0.013	11.521
P7	0.043	0.006	15.719	0.124	0.013	11.671	0.125	0.014	12.374
P8	0.044	0.007	17.630	0.123	0.012	11.134	0.124	0.013	11.832
M	0.042	0.005	13.530	0.124	0.013	11.526	0.124	0.013	11.788
SD	0.001	0.001	2.406	0.000	0.000	0.371	0.001	0.001	0.462

Note: For stimulus D, P2 thresholds were omitted due to extreme lack of fit (R^2=0.01)

A repeated-measures ANOVA revealed that there was a significant effect of stimulus on JND, F(2,12)=233.61, p<0.001, partial η^2=0.97 (see Figure 8-1). Pre-planned contrasts showed that the JND for stimulus 'D' was significantly smaller than that of 'E' and 'F', F(1,6)=205.54, p<0.001, partial η^2=0.97, and F(1,6)=406.12, p<0.001, partial η^2=0.98, respectively. There was no significant difference between the JNDs for stimulus 'E' and 'F', F(1,6)=3.69, p=0.103, partial η^2=0.38. A further repeated-measures ANOVA compared the effect of stimulus on Weber Fractions (see Figure 8-1). This revealed that there was no significant effect of stimulus on Weber Fractions, F(1.08,6.47)=4.69, p=0.069, partial η^2=0.44.

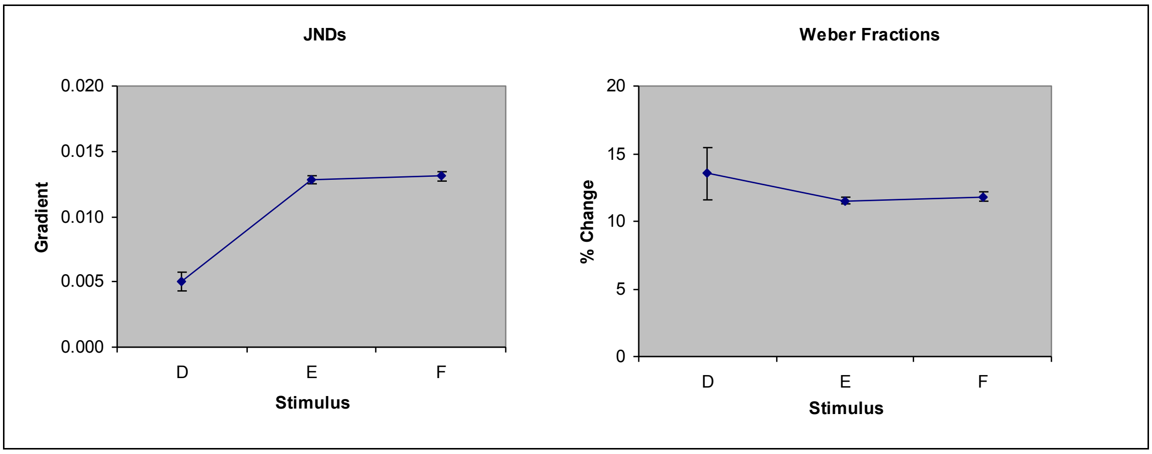


Figure 8-1: Effect of stimulus on JNDs and Weber Fractions

Note: Error bars represent the 95% Confidence Interval of the mean, and do so throughout

Prior to Study 3, predictions were made for the threshold gradients of each stimulus (see Table 8-1). In comparing the predicted and actual gradients (see Table 8-3) it was found that there was a small discrepancy for stimulus 'D' (2.97%), and slightly larger differences for stimuli 'E' (12.04%) and 'F' (11.78%). For stimulus 'D' the 95% CI for the means overlapped which would suggest that this difference is not significant, however for ‘E’ and ‘F’ they did not overlap which would suggest a significant difference between the predicted and actual thresholds. These differences and similarities can be more easily seen when comparing the Weber Fractions (see Table 8-4). An independent t-test was used to compare the mean predicted and actual Weber Fractions for ‘E’ and ‘F’ combined, and for all stimuli combined. In both cases it was revealed that the difference between the actual and predicted Weber Fractions was significant ($p<0.05$).

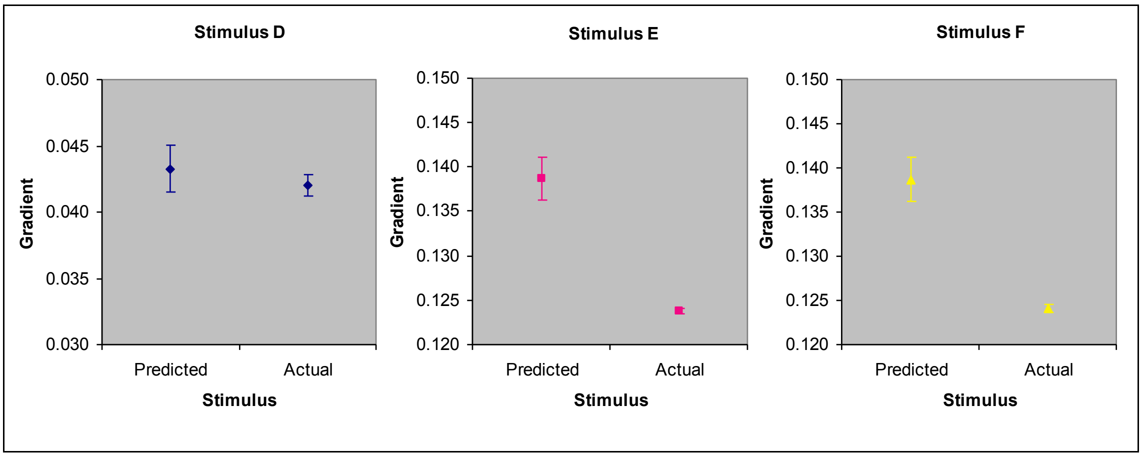


Figure 8-2: Comparison of predicted and actual threshold gradients

Table 8-3: Comparison of predicted and actual threshold gradients

Stimulus	Predicted			Actual			Error (%)
	Threshold Gradient	95% CI		Threshold Gradient	95% CI		
		Lower	Upper		Lower	Upper	
D	0.043	0.041	0.045	0.042	0.041	0.043	2.97
E	0.139	0.136	0.141	0.124	0.124	0.124	12.04
F	0.139	0.136	0.141	0.124	0.123	0.125	11.78

Table 8-4: Comparison of predicted and actual Weber Fractions with significance test

Stimulus	Predicted		Actual		Independent t-test			
	Mean	SD	Mean	SD	t	df	r	sig. (2-tailed)
D	16.91	-	13.53	2.41	-	-	-	-
E	25.20	-	11.74	0.37	98.70	2	1.00	0.00
F	25.20	-	12.01	0.46				
All	22.44	4.78	12.43	0.96	3.55	4	0.87	0.24

Note: The first t-test compares the predicted against the actual Weber Fractions for stimuli ‘E’ and ‘F’. The second t-test (All) compares the predicted against the actual for all stimuli.

As the results here were unexpected, a post-hoc regression analysis was made in order to understand the discrepancy between the predicted and actual scores. This found that ‘DEF’ had different model characteristics to ‘BC’ (see Table 8-5). This is illustrated in Figure 8-3 where the regression lines can be seen to be diverging.

Table 8-5: Regression analysis of Visual-Haptic thresholds (studies 1 and 3)

	B	SE B	βeta	t	Sig.	R^2	R^2_{adj}
Stimuli 'DEF' Visual-Haptic Threshold Gradient							
Constant	0.001	0.000		2.863	0.01	0.999	0.999
Stimulus Gradient	1.107	0.004	1.000	290.497	0.00		
Stimuli 'BC' Visual-Haptic Threshold Gradient							
Constant	-0.004	0.001		-5.541	0.00	0.999	0.999
Stimulus Gradient	1.285	0.004	1.000	300.467	0.00		
Stimuli 'BCDEF' Visual-Haptic Threshold Gradient							
Constant	-0.011	0.003		-4.180	0.000	0.990	0.990
Stimulus Gradient	1.291	0.21	0.995	61.793	0.000		

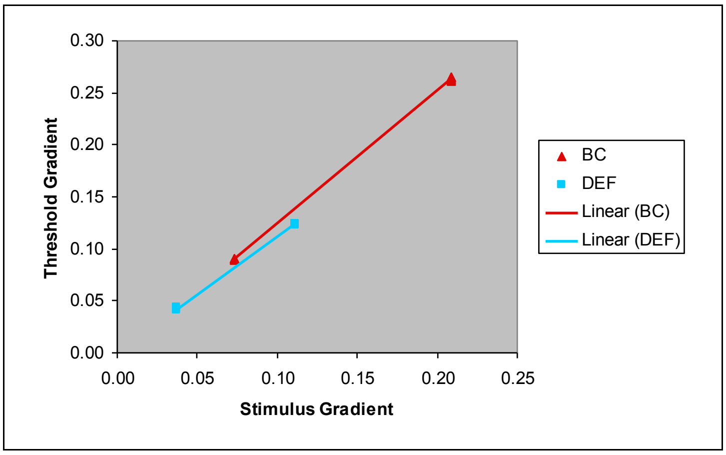


Figure 8-3: Comparison of linear regressions with study 1

8.3.2 Performance

8.3.2.1 Response Time

A repeated-measures ANOVA compared the effect of stimulus and sound on response time (see Figure 8-4). This revealed that there was a main effect of stimulus, $F(2,14)=6.83$, $p<0.01$, partial $\eta^2=0.49$. Pre-planned contrasts showed that 'D' (3.38s) was significantly slower than either 'E' (2.39) or 'F' (2.58s), $F(1,7)=8.58$, $p=0.022$, partial $\eta^2=0.55$, and $F(1,7)=8.92$, $p=0.020$, partial $\eta^2=0.56$, respectively.

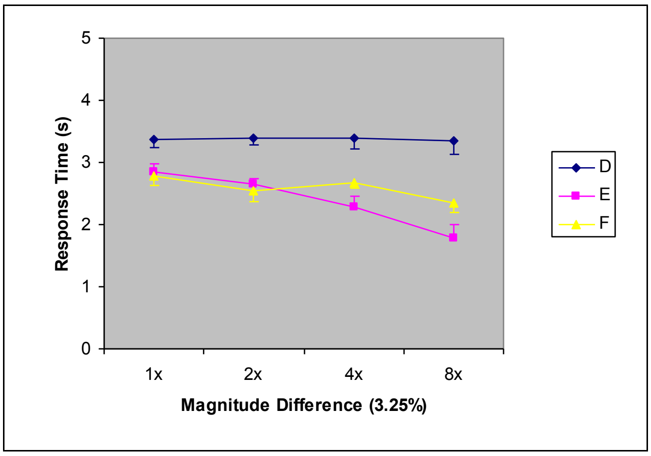


Figure 8-4: Effect of stimulus and magnitude difference on response time

There was also a main effect of magnitude difference, $F(3,21)=19.44$, $p<0.001$, partial $\eta^2=0.73$. Pre-planned contrasts revealed that '1x' magnitude difference (3s) was significantly slower than either '2x' (2.86s), '4x' (2.78s), or '8x' (2.49s), $F(1,7)=10.36$, $p=0.015$, partial $\eta^2=0.60$, $F(1,7)=14.60$, $p<0.01$, partial $\eta^2=0.68$, and $F(1,7)=45.54$, $p<0.001$, partial $\eta^2=0.87$, respectively.

Finally, there was an interaction effect between Stimulus and Magnitude Difference. Pre-planned contrasts revealed that the effect of '4x' magnitude difference (compared to '1x') was greater for 'E' compared to either 'D' or 'F' (see Figure 8-5 (i) and (ii) respectively), $F(1,7)=6.82$, $p=0.035$, partial $\eta^2=0.49$, and $F(1,7)=19.66$, $p<0.01$, partial $\eta^2=0.74$, respectively. It was also found that the effect of '8x' magnitude difference (compared to '1x') was greater for 'E' compared to 'D' (the effect looked similar to that shown for '4x' in Figure 8-5 (i)), $F(1,7)=17.76$, $p<0.01$, partial $\eta^2=0.72$.

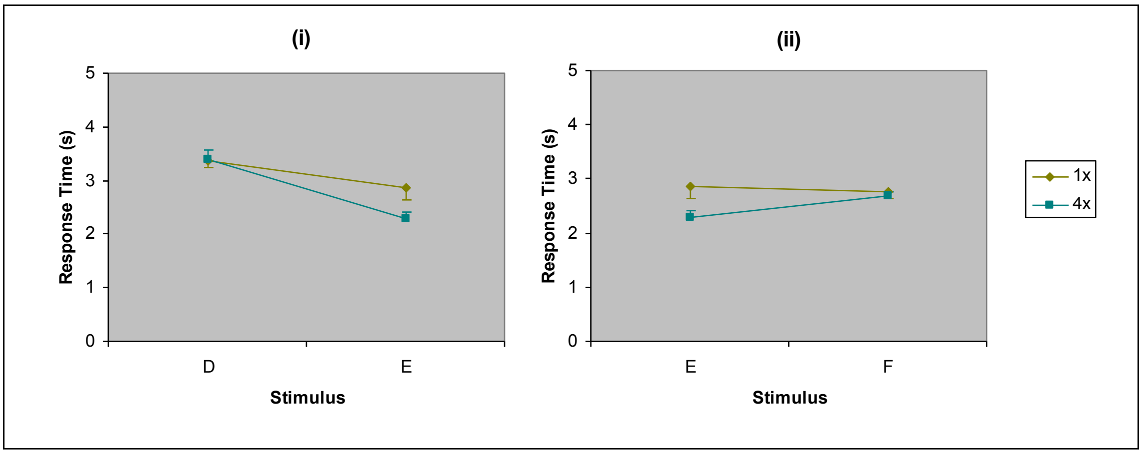


Figure 8-5: Interaction effects between stimulus and magnitude difference

8.3.2.2 Accuracy

A repeated-measures ANOVA compared the effect of stimulus and sound on accuracy (see Figure 8-6). This revealed that there was a main effect of stimulus on accuracy, F(2,14)=36.31, p<0.001, partial η^2=0.84. Pre-planned contrasts showed that 'E' (84%) was significantly more accurate than either 'D' (64%) or 'F' (76%), and that "F' was significantly more accurate than 'D', F(1,7)=48.58, p<0.001, partial η^2=0.87, F(1,7)=18.66, p<0.01, partial η^2=0.73, and F(1,7)=28.96, p<0.001, partial η^2=0.80, respectively. There was also a main effect of magnitude difference, F(3,21)=37.19, p<0.001, partial η^2=0.84. Pre-planned contrasts revealed that '1x' magnitude difference (62%) was significantly less accurate than either '2x' (68%), '4x' (81%), or '8x' (87%), F(1,7)=9.26, p=0.019, partial η^2=0.57, F(1,7)=44.38, p<0.001, partial η^2=0.86, and F(1,7)=77.90, p<0.001, partial η^2=0.92, respectively.

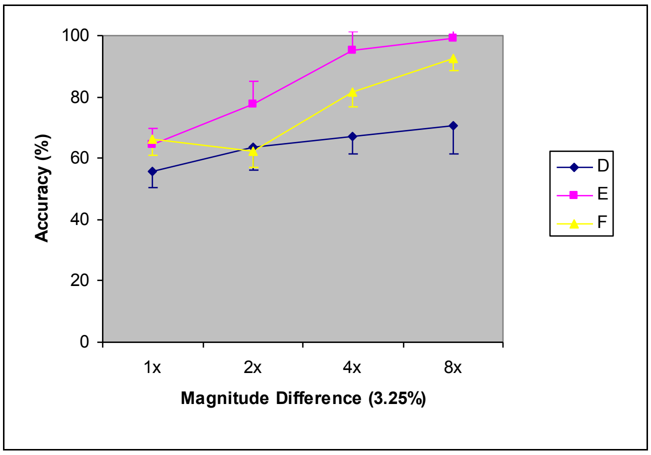


Figure 8-6: Effect of stimulus and magnitude difference on accuracy

Finally, there was an interaction effect between Stimulus and Magnitude Difference, F96,42)=5.01, p<0.01, partial η^2=0.42. Pre-planned contrasts revealed that the effect of '4x' magnitude difference (compared to '1x') was greater for 'E' compared to either 'D' or 'F' (see Figure 8-7 (i) and (ii) respectively), F(1,7)=24.01, p<0.01, partial η^2=0.77, and F(1,7)=20.52, p<0.01, partial η^2=0.75, respectively. It was found that the effect of '8x' magnitude difference (compared to '1x') was greater for 'E' compared to 'D' (see Figure 8-7 (iii)), F(1,7)=8.94, p=0.020, partial η^2=0.56, and that the effect of '2x" magnitude difference (compared to '1x') was greater for 'E' compared to 'F' (see Figure 8-7 (iv)), F(1,7)=7.22, p=0.31, partial η^2=0.51.

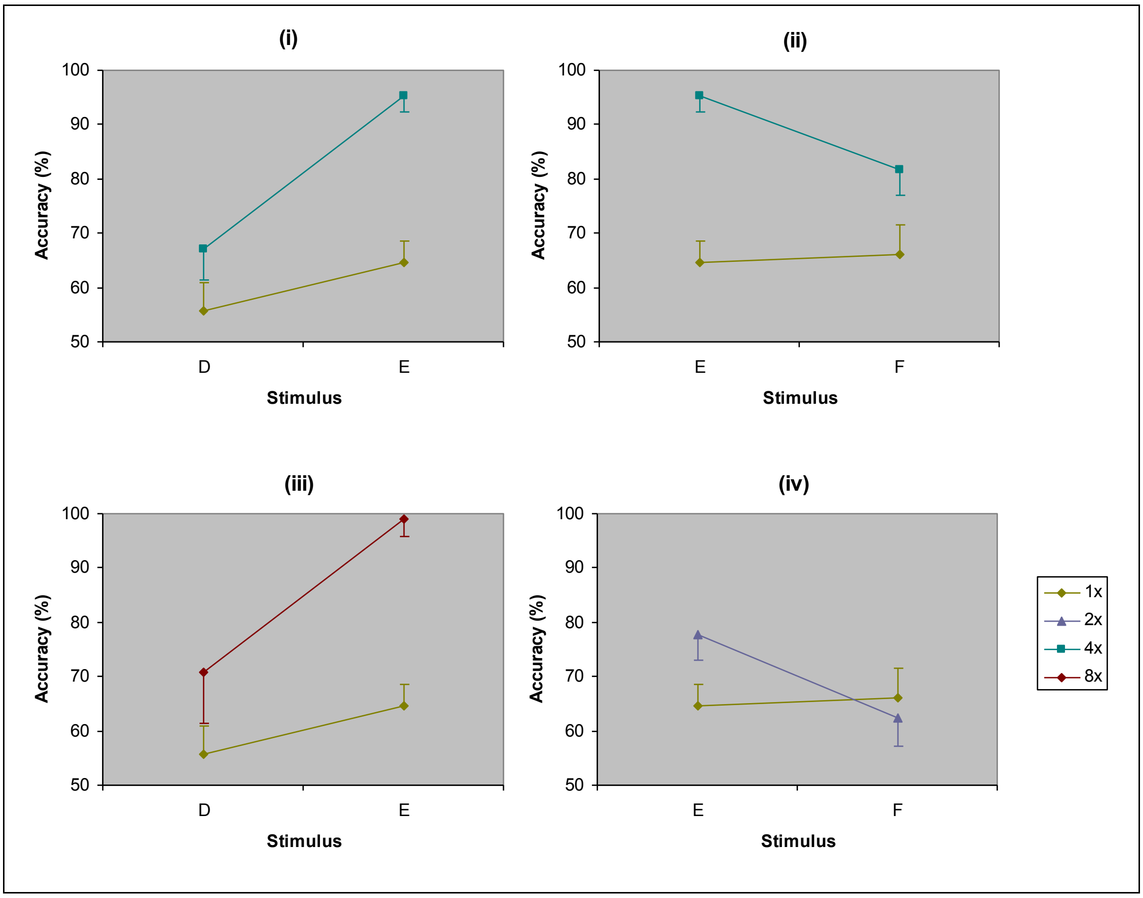


Figure 8-7: Interaction effects between stimulus and magnitude difference on accuracy

8.3.2.3 Confidence

A repeated-measures ANOVA compared the effect of stimulus and sound on confidence (see Figure 8-8). This revealed that there was a main effect of stimulus on confidence, $F(2,14)=4.55$, $p=0.030$, partial $\eta^2=0.39$. Pre-planned contrasts showed that 'D' (49%) was significantly less confident than 'E' (70%), but not significantly less than 'F' (69%), $F(1,7)=7.75$, $p=0.027$, partial $\eta^2=0.53$, and $F(1,7)=5.43$, $p=0.053$, partial $\eta^2=0.44$, respectively. There was also a main effect of magnitude difference, $F(3,21)=15.84$, $p<0.01$, partial $\eta^2=0.69$. Pre-planned contrasts revealed that '1x' magnitude difference (62%) was significantly less confident than either '4x' (63%), or '8x' (77%) but not '2x' (56.42%), $F(1,7)=7.59$, $p=0.028$, partial $\eta^2=0.52$, and $F(1,7)=25.97$, $p<0.01$, partial $\eta^2=0.79$.

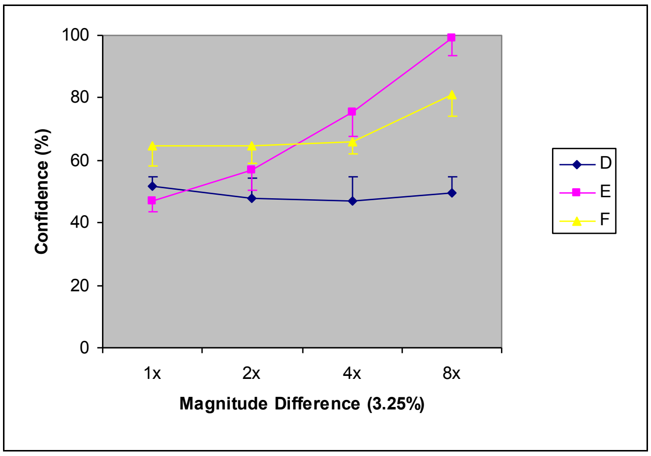


Figure 8-8: Effect of stimulus and magnitude difference on confidence

Finally, there was an interaction effect between stimulus and magnitude difference, $F(6,42)=10.72$, $p<0.001$, partial $\eta^2=0.60$. Pre-planned contrasts revealed that the effect of '4x' magnitude difference (compared to '1x') was greater for 'E' compared to either 'D' or 'F' (see Figure 8-9 (i) and (ii) respectively), $F(1,7)=20.36$, $p<0.01$, partial $\eta^2=0.74$, and $F(1,7)=9.77$, $p=0.017$, partial $\eta^2=0.58$, respectively. It was also found that the effect of '8x' magnitude difference (compared to '1x') was greater for 'E' compared to either 'D' or 'F' (see Figure 8-9 (iii) and (iv)), $F(1,7)=30.93$, $p<0.01$, partial $\eta^2=0.81$, and $F(1,7)=14.35$, $p<0.01$, partial $\eta^2=0.67$, respectively. In addition, it was found that the effect of '8x' magnitude difference (compared to '1x') was greater for 'F' compared to 'D', $F(1,7)=9.45$, $p=0.018$, partial $\eta^2=0.57$.

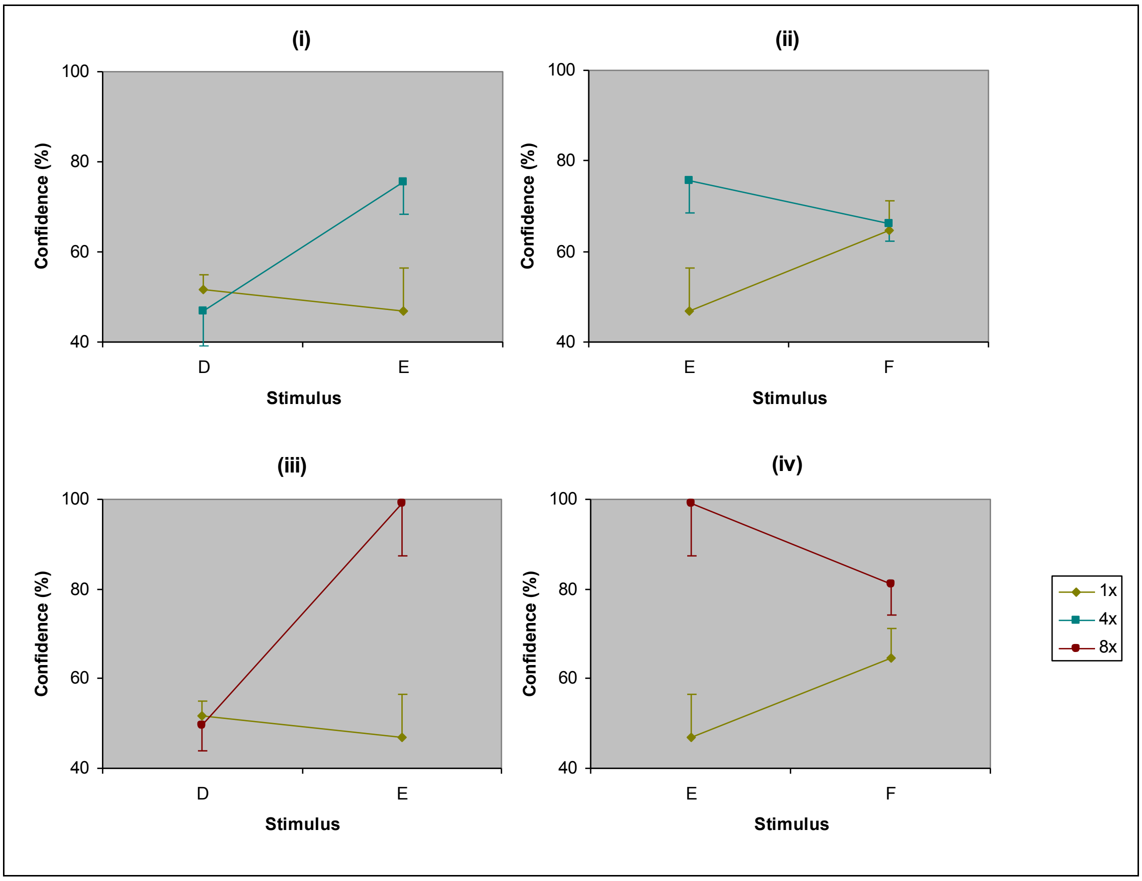


Figure 8-9: Interaction effects of stimulus and magnitude differences on confidence

8.3.3 Performance Predictability – Multiple Regression Analysis

The ANOVA results had revealed several main effects and interactions which meant that the independent variables did not have a straightforward effect upon the various performance measures. For this reason, multiple regression analysis was chosen to explore the data in order to derive a model, since it was felt that more than one independent variable was likely to be influential and so several predictors of performance were expected.

The independent variables were magnitude difference and stimulus. The approach taken was to analyse changes in performance against magnitude difference and stimulus dimensions. There were three dimensions of the stimulus that could be considered; gradient, curvature, and width. However, curvature was ruled out of the analysis as previous results had indicated this had had no effect on performance (see 8.4.2 for discussion of this finding). The predictors used in the regression analysis were therefore magnitude difference, stimulus gradient, and stimulus width[23]. A multiple regression analysis was run for each performance measure; response time, accuracy, and confidence. In addition, multiple regression analysis was undertaken on the visual-haptic data from stimuli 'B' and 'C' from Study 1 (see Chapter 3) for comparison with the model generated from stimuli 'D', 'E' and 'F'. Finally visual-haptic data from all five stimuli were analysed to determine if this improved the models derived from the different groups of stimuli data. In each case the method used was 'forced-entry' in the order; stimulus gradient, magnitude difference, and stimulus width. The results of this analysis are summarised in Table 8-6 to Table 8-14.

8.3.3.1 Response Time

For response time the regression analysis showed a good model fit to the data for stimuli 'DEF' and 'BE', R^2=0.80 and R^2=0.90 respectively. This level of fit was maintained when generalised to the population, albeit with a reduction of 2-3%, R^2_{adj}=0.77 and 0.88. With 'DEF' it was found that stimulus gradient accounted for 65% of the variance in response time, whereas for 'BE' it was 44%. The level of variance accounted for by magnitude difference also varied, for 'DEF' it was found to be 13% whereas for 'BC' it was 46%. For 'DEF', stimulus width[24] accounted for 2% of response time variance, and was not a significant predictor of response time, $t(20)=-1.51$, $p>0.05$. Overall the models (3 and 2, respectively) produced for stimuli 'DEF' and 'BC' provided a better estimate of response time than using the sample mean, F=26.14 and F=56.15, $p<0.001$. Finally, the regression of all stimuli ('BCDEF') did not provide a good fit to the data and therefore did not provide a viable model, R^2=0.01.

[23] Contrasts in the ANOVA results indicated that there was significantly more effect on the longer stimulus ('E') than either of the shorter ones, given this it was considered that contribution of width should be explored.

[24] There was insufficient data for analysis of stimulus width for stimuli 'BC'.

Table 8-6: Response time multiple regression (stimuli DEF)

	B	SE B	βeta	t	Sig.	R^2	R^2_{adj}	ΔR^2
Step 1 (model 1)								
Constant	3.88	0.19		20.97	0.00	0.65	0.63	0.65
Stimulus Gradient	-12.71	2.00	-0.80	-6.36	0.00			
Step 2 (model 2)								
Constant	4.14	0.17		24.41	0.00	0.77	0.75	0.13
Stimulus Gradient	-12.71	1.64	-0.80	-7.75	0.00			
Magnitude Difference	-0.02	0.01	-0.36	-3.42	0.00			
Step 3 (model 3)								
Constant	4.18	0.17		25.04	0.00	0.80	0.77	0.02
Stimulus Gradient	-11.32	1.84	-0.72	-6.16	0.00			
Magnitude Difference	-0.02	0.01	-0.36	-3.52	0.00			
Stimulus Width	-0.01	0.01	-0.18	-1.51	0.15			

Table 8-7: Response time multiple regression (stimuli BC)

	B	SE B	βeta	t	Sig.	R^2	R^2_{adj}	ΔR^2
Step 1 (model 1)								
Constant	8.03	0.83		9.69	0.00	0.44	0.40	0.44
Stimulus Gradient	-17.66	5.31	-0.66	-3.32	0.01			
Step 2 (model 2)								
Constant	9.73	0.43		22.44	0.00	0.90	0.88	0.46
Stimulus Gradient	-16.21	2.38	-0.61	-6.80	0.00			
Magnitude Difference	-0.07	0.01	-0.68	-7.55	0.00			

Table 8-8: Response time multiple regression (stimuli BCDEF)

	B	SE B	βeta	t	Sig.	R^2	R^2_{adj}	ΔR^2
Step 1 (model 1)								
Constant	4.05	0.64		6.37	0.00	0.00	-0.02	0.00
Stimulus Gradient	-1.51	5.21	-0.05	-0.29	0.77			
Step 2 (model 2)								
Constant	4.03	0.68		5.94	0.00	0.00	-0.05	0.00
Stimulus Gradient	-1.66	5.50	-0.05	-0.30	0.76			
Magnitude Difference	0.00	0.02	0.02	0.10	0.92			
Step 3 (model 3)								
Constant	4.32	1.01		4.26	0.00	0.01	-0.08	0.00
Stimulus Gradient	-2.34	5.83	-0.07	-0.40	0.69			
Magnitude Difference	0.00	0.02	0.01	0.08	0.94			
Stimulus Width	-0.01	0.04	-0.07	-0.39	0.70			

8.3.3.2 Accuracy

For accuracy the regression analysis showed a good model fit to the data for stimuli 'DEF' and less so for 'BE', R^2=0.79 and R^2=0.55 respectively. This level of fit was maintained when generalised to the population albeit with a reduction of 3% and 7% (R^2_{adj}=0.76 and 0.48) respectively. With 'DEF' it was found that stimulus gradient accounted for 27% of the variance in accuracy, whereas for 'BE' it was 7%. The level of variance accounted for by magnitude difference also varied, for 'DEF' it was found to be 46% with a similar level for 'BC' at 48%. For 'DEF', stimulus width[25] accounted for 6% of

[25] There was insufficient data for analysis of stimulus width for stimuli 'BC'.

accuracy variance. Overall the models (3 and 2, respectively) produced for stimuli 'DEF' and 'BC' provided a better estimate of accuracy than using the sample mean, F=25.50 and F=7.90, p<0.001. Finally, the regression of all stimuli ('BCDEF') provided a good fit to the sample data, and accounted for 2% less variance in accuracy when generalised to the population, R^2=0.72 and R^2_{adj} =0.70 respectively. This improved on the model fit over 'BE' by 22% but accounted for 7% less variance in accuracy compared to 'DEF'.

Table 8-9: Accuracy multiple regression (stimuli DEF)

	B	SE B	βeta	t	Sig.	R^2	R^2_{adj}	ΔR^2
Step 1								
Constant	55.40	7.19		7.70	0.00	0.27	0.24	0.27
Stimulus Gradient	223.21	77.56	0.52	2.88	0.01			
Step 2								
Constant	42.02	4.97		8.45	0.00	0.73	0.71	0.46
Stimulus Gradient	223.21	48.00	0.52	4.65	0.00			
Magnitude Difference	1.10	0.18	0.68	6.04	0.00			
Step 3								
Constant	40.24	4.56		8.82	0.00	0.79	0.76	0.06
Stimulus Gradient	163.68	50.17	0.38	3.26	0.00			
Magnitude Difference	1.10	0.16	0.68	6.67	0.00			
Stimulus Width	0.42	0.18	0.28	2.37	0.03			

Table 8-10: Accuracy multiple regression (stimuli BC)

	B	SE B	βeta	t	Sig.	R^2	R^2_{adj}	ΔR^2
Step 1								
Constant	84.70	7.22		11.73	0.00	0.07	0.00	0.07
Stimulus Gradient	47.26	46.31	0.26	1.02	0.32			
Step 2								
Constant	72.87	6.11		11.92	0.00	0.55	0.48	0.48
Stimulus Gradient	37.26	33.58	0.21	1.11	0.29			
Magnitude Difference	0.48	0.13	0.69	3.72	0.00			

Table 8-11: Accuracy multiple regression (stimuli BCDEF)

	B	SE B	βeta	t	Sig.	R^2	R^2_{adj}	ΔR^2
Step 1								
Constant	66.15	4.73		13.99	0.00	0.26	0.24	0.26
Stimulus Gradient	140.77	38.77	0.51	3.63	0.00			
Step 2								
Constant	59.28	3.53		16.78	0.00	0.64	0.62	0.38
Stimulus Gradient	91.11	28.65	0.33	3.18	0.00			
Magnitude Difference	0.67	0.11	0.64	6.20	0.00			
Step 3								
Constant	47.78	4.60		10.38	0.00	0.72	0.70	0.09
Stimulus Gradient	117.90	26.47	0.43	4.45	0.00			
Magnitude Difference	0.69	0.10	0.66	7.23	0.00			
Stimulus Width	0.55	0.16	0.31	3.40	0.00			

8.3.3.3 Confidence

For confidence, the regression analysis showed a good model fit to the data for stimuli 'BE' but less so for 'DEF', R^2=0.76 and R^2=0.60 respectively. These relative levels were maintained when generalised to the population with a reduction of 6% and 4% (R^2_{adj}=0.54 and 0.72) respectively. With 'DEF' it was found that stimulus gradient accounted for 33% of the variance in confidence, whereas for 'BE' it was 10%. The level of variance accounted for by magnitude difference also varied, for 'DEF' it was found to be 27% whereas for 'BC' it was 66%. For 'DEF', stimulus width[26] accounted for none of the variance in confidence. Overall the models (3 and 2, respectively) produced for stimuli 'DEF' and 'BC' provided a better estimate of response time than using the sample mean, F=9.90 and F=20.74, p<0.001. Finally, the regression of all stimuli ('BCDEF') provided a good fit to the sample data, and accounted for 3% less variance when generalised to the population, R^2=0.71 and R^2_{adj} =0.68 respectively. This improved on the model fit over 'DEF' by 11% but accounted for 5% less variance in confidence compared to 'BE'.

Table 8-12: Confidence multiple regression (stimuli DEF)

	B	SE B	βeta	t	Sig.	R^2	R^2_{adj}	ΔR^2
Step 1								
Constant	37.32	8.21		4.54	0.00	0.33	0.30	0.33
Stimulus Gradient	291.10	88.57	0.57	3.29	0.00			
Step 2								
Constant	25.20	7.28		3.46	0.00	0.60	0.56	0.27
Stimulus Gradient	291.10	70.24	0.57	4.14	0.00			
Magnitude Difference	0.99	0.27	0.52	3.74	0.00			
Step 3								
Constant	25.11	7.56		3.32	0.00	0.60	0.54	0.00
Stimulus Gradient	288.30	83.10	0.57	3.47	0.00			
Magnitude Difference	0.99	0.27	0.52	3.65	0.00			
Stimulus Width	0.02	0.29	0.01	0.07	0.95			

Table 8-13: Confidence multiple regression (stimuli BC)

	B	SE B	βeta	t	Sig.	R^2	R^2_{adj}	ΔR^2
Step 1								
Constant	74.22	9.04		8.21	0.00	0.10	0.03	0.10
Stimulus Gradient	71.36	57.98	0.31	1.23	0.24			
Step 2								
Constant	56.52	5.65		10.00	0.00	0.76	0.72	0.66
Stimulus Gradient	56.40	31.04	0.25	1.82	0.09			
Magnitude Difference	0.72	0.12	0.82	6.01	0.00			

Table 8-14: Confidence multiple regression (stimuli BCDEF)

	B	SE B	βeta	t	Sig.	R^2	R^2_{adj}	ΔR^2
Step 1								
Constant	50.73	5.69		8.91	0.00	0.30	0.28	0.30
Stimulus Gradient	189.39	46.67	0.55	4.06	0.00			
Step 2								
Constant	42.25	4.13		10.22	0.00	0.68	0.66	0.37
Stimulus Gradient	128.02	33.51	0.37	3.82	0.00			
Magnitude Difference	0.83	0.13	0.64	6.55	0.00			
Step 3								
Constant	33.81	5.88		5.75	0.00	0.71	0.68	0.03
Stimulus Gradient	147.68	33.84	0.43	4.36	0.00			
Magnitude Difference	0.84	0.12	0.65	6.90	0.00			
Stimulus Width	0.40	0.21	0.19	1.95	0.06			

[26] There was insufficient data for analysis of stimulus width for stimuli 'BC'.

8.4 Discussion

8.4.1 Perception

The analysis sought to address the following questions in relation to perception:

- Do stimuli with different gradients produce different JNDs?
- Do stimuli with the same gradient have the same JND?
- Do Weber Fractions vary with gradient?
- How accurately did the model predict the actual threshold gradients observed in Study 3?
- Does this support the theory that stimulus gradient is a predictor of threshold gradient?

Analysis of Study 3 data showed that stimulus 'D' had a JND of 0.004 which was significantly different from the JNDs of stimuli 'E' (0.012) and 'F' (0.013). Additionally there was found to be no significant difference between the JNDs for stimuli 'E' and 'F'. The results showed that stimuli with different gradients produce different JNDs, whilst those with the same gradient produce JNDs with no significant difference. Furthermore, this holds true regardless of other stimulus dimensions such as width or curvature i.e. stimulus shape. This fits with the work of Louw (2002) which showed that shapes with a similar slope, regardless of object shape, were practically indistinguishable from each other; *that is because the gradients were the same the curve shape felt the same*. Here we see that, regardless of curve shape, when the gradient is the same the threshold will also be the same; that is *where different curve shapes share a common gradient it takes the same increase in gradient to feel the curve shape as different*. This finding confirmed the basis of using gradient in the model described in Chapter 4. However the question remains of the level of accuracy provided by this and is discussed later.

It was found that there was no significant difference between the Weber Fractions for Stimulus 'D' (11.82%), 'E' (11.48%) or 'F' (11.43%). This means that the amount of change required to detect a difference in the curve shape is a constant proportion of the stimulus. This is surprising since it means that the difference thresholds found in this study conform to Weber's Law, which is counter to the findings of Study 1 (see Chapter 3 section 3.4.1) and the literature (Kappers and Koenderink 1996). However, for Study 1 (and in the literature) Weber Fractions were calculated based on curvature, whereas for Study 2 these were based on gradient. This may have accounted for the discrepancy in findings. Because of importance of this discrepancy the data for Study 1 were recalculated in terms of gradient and a further ANOVA calculated (see Appendix N for

ANOVA results). This found that, for Study 1, there was a significant effect of stimulus gradient on Weber Fractions, meaning that there was no conformance to Weber's Law. This confirms the discrepancy between findings for Study 1 and Study 3.

There are a number of possible explanations for this difference in results; error in the model, perceptual sensitivity of participants, or differences in experimental conditions. The first two of these were dealt with by re-analysing the data based only on participants who had undertaken both studies (see Appendix N for ANOVA results). The re-analysis confirmed the findings of the original; Study 1 did not conform to Weber's Law whereas Study 3 did. Therefore the most likely explanation of the discrepancy in the findings was to be found in differences between the experimental conditions. The main difference[27], and most likely cause of the discrepancy, was the nature of the standard and comparison stimuli. For Study 1, the standard and comparison stimuli were situated on separate blocks (see Chapter 3 Figure 3-2), whereas for Study 3, the standard and comparison were situated within the same block (see Chapter 7 Figure 7-6). Effectively, this meant that for Study 1, inter-object differences were judged and for Study 3 intra-object differences. The findings for Study 3 therefore show that for intra-object differences there is a conformance with Weber's Law and that threshold difference is about 12%. This differs from Study 1 where inter-object thresholds were higher (23%) and there was no conformance to Weber's Law. This difference in thresholds may have implications for the prediction of threshold and is discussed next in relation to the accuracy of the predictions for Study 3. This also had consequences for the prediction of performance which is discussed in 8.4.3.

Given the above, it is now of little surprise that the predictions made from the model had varying degrees of accuracy. It was found that the prediction for stimulus 'D' was not significantly different from the observed threshold (2% difference), but that the predictions for stimuli 'E' and 'F' were significantly different (12% difference). Further to this, whilst the prediction for 'D' fell within the 95% CI for the mean both 'E' and 'F' were outside of the mean CI, which further suggests the failure of prediction for those stimuli. Prior to Study 3 there was no reason to believe that there would be a differential effect

[27] There was another difference that may have had a bearing on this; the use of powder to cut down on friction in study 3. The curvature of a high-friction surface is often over-estimated, where the curvature of a low-friction surfaces is often underestimated (Christou and Wing 2001)

between inter- and intra-object comparisons. However, the results of this study suggest that there were differences between these two modes of comparison and so would account for the failure in prediction accuracy. The post-hoc regression analysis showed that as the gradient increased ‘BC’ and ‘DEF’ regression lines diverged (as the model coefficients are different). This would have accounted for the small error seen in ‘D’ and the larger errors for ‘E’ and ‘F’. It could also be seen that when a regression was made for all the data, the model fit, whilst high, was not as good as the models which describe each of the data separately. Therefore, the failure of the model to predict the gradients for ‘E’ and ‘F’ did not so much undermine the underlying theory that stimulus gradient was the predictor of threshold gradient, but highlighted that this was dependent upon whether judgements were made for inter- or intra-object differences. So whilst the model based on stimuli ‘BC’ did not prove an accurate fit for stimuli ‘DEF’, the underlying theory was confirmed through producing a model with an extremely high fit with the same predictor variable.

8.4.2 Performance

The analysis sought to address the following question in relation to performance:

- Do different dimensions of the stimulus (gradient, curvature, width) have an effect on performance (response time, accuracy, confidence)?

The results showed that, for response time, ‘D’ (3.4s) was significantly slower than either ‘E’ or ‘F’ (it differed from both of these in respect of gradient). As there was no significant difference between the response times of ‘E’ (2.4s) and ‘F’ (2.6s) then this suggests that the dimension of effect is likely to be gradient (see Chapter 7 section 7.1.2 for explanation of the inferences used here). However, this was complicated by the fact that there were interactions between stimuli and magnitude difference. When these were examined it was seen that, at higher magnitudes, there was more of an effect on response time by stimulus ‘E’ as opposed to ‘D’ or ‘F’. So whilst there was seen to be a main effect on response time of gradient (both ‘E’ and ‘F’ are the same), there was also some interaction effect of width at larger magnitudes (‘E’ was wider than both ‘D’ and ’F’ which had the same width).

For accuracy it was found that ‘E’ (84%) was more accurate than either ‘D’ or ‘F’ (it differed from both of these in respect of width). However, as ‘F’ (76%) was more accurate than ‘D’ (64%), which was the same width, it is difficult to infer a sole effect of width on accuracy. Given that both ‘E’ and ‘F’ have higher accuracy than ‘D’, it is likely that gradient also has an effect, and it is the combination of gradient and width that has increased ‘E’s accuracy over ‘F’s and ‘F’s over ‘D’s. As with response time, this was complicated by the fact that there were interactions between stimuli and magnitude difference. Again it was seen that, at higher magnitudes, there was more of an effect on performance by stimulus ‘E’ as opposed to ‘D’ or ‘F’. So whilst there was seen to be a main effect on accuracy of gradient and/or width, there was also some interaction effect of width at larger magnitudes (‘E’ was wider than both ‘D’ and ’F’ which had the same width). Interestingly the effect on ‘E’ appears more pronounced for accuracy than it did with response time, this could be due to the main effect being a mix of width and gradient; so the effect of width is further accentuated.

To conclude, it was seen that performance was affected by gradient (response time and confidence), and a mixture of gradient and width (accuracy). It was also observed that at larger magnitude differences there was a subtle influence of width for an improved performance (this was greater for accuracy). However, there is no evidence to suggest that there is any effect of curvature on performance in this study.

8.4.3 Predictability

The analysis sought to address the following questions in relation to prediction of performance:

- Is there a predictable relationship between performance and stimulus for varying magnitude differences?
- Is this relationship similar for visual-haptic data from Study 1?
- Can the result from Study 1 and Study 3 be combined to improve the generalisation of the model?

The results showed that it was possible to fit a model to each of the performance data. The models for response time and accuracy produced a very good fit (R^2=0.80 and R^2=0.79, respectively) whilst that for confidence accounted for only just over half the variance observed (R^2=0.60). The main predictor for each performance measure was different, for response time it was gradient (ΔR^2=0.65), for accuracy it was magnitude difference (ΔR^2=0.46), and for confidence there was a roughly equal split between gradient (ΔR^2=0.33) and magnitude difference (ΔR^2=0.27). Of the predictors, stimulus width had little impact on the model for any of the performance measures (ΔR^2=0.02, ΔR^2=0.06, and ΔR^2=0.00). However it should be noted that to some extent width was accounted for within the gradient term as this expressed a ratio of height to width. There is no suggestion of any collinearity between gradient and width, so the variation described is distinct. Interestingly the largest variance accounted for by width (6%) is found in the model for accuracy. This aligns with the finding for the interaction effect between stimulus and magnitude differences, which showed a greater effect for larger widths (stimulus 'E') in accuracy than the other performance measures. So whilst it would be tempting to remove this term from the model it may be useful at larger magnitude differences[28]. Further investigation would be needed in order to be definitive about the exact usefulness of this predictor.

Table 8-15: Summary of model statistics for all performance measures

Performance Measure	Study	Stimuli	ΔR^2			All Entered	
			Gradient	Mag. Diff.	Width	R^2	R^2adj
RT	3	DEF	0.65	0.13	0.02	0.80	0.77
	1	BC	0.44	0.46	-	0.90	0.88
	1&3	ALL	0.00	0.00	0.00	0.01	-0.08
ACC	3	DEF	0.27	0.46	0.06	0.79	0.76
	1	BC	0.07	0.48	-	0.55	0.48
	1&3	ALL	0.26	0.38	0.09	0.72	0.70
CON	3	DEF	0.33	0.27	0.00	0.60	0.54
	1	BC	0.10	0.66	-	0.76	0.72
	1&3	ALL	0.30	0.37	0.03	0.71	0.68

In comparing the models produced from Study 1 ('BE') and Study 3 ('DEF') data it was seen that there were some differences in the variance accounted for by each predictor. It is difficult to assign any meaning to these differences and it would be necessary to provide more samples in order to detect any significant patterns. However, if we recall that the difference threshold[29] was higher in Study 1 than in Study 3, it is interesting to note that for accuracy the models are similar in the levels of variance accounted for by magnitude difference, and vary for stimulus, with 'DEF' having 27% and 'BE' having 7%. It may be that gradient information was more easily processed in intra-object judgments and so contributed to the higher values seen in 'DEF'. However, this is highly speculative and would need further research.

Overall the individual models produced describe the data well. They showed the main predictors to be gradient and magnitude difference. Width did play some part in describing the variance, although this is thought to be more relevant at greater

[28] It increases to 8% of the variance in the Model 'BCDEF' which includes magnitude differences up to 60%

[29] Difference threshold is the magnitude difference at which there is 75% accuracy

magnitude differences and for accuracy, and should remain within the models. The strongest models were for response time and accuracy for Study 1 (R^2_{adj}=0.88 and R^2_{adj}=0.72), and response time and confidence for Study 3 (R^2_{adj}=0.77 and R^2_{adj}=0.76), as both reduce little when generalised to the population. In this respect the models that describe response time are the most robust since they consistently account for high proportions of variance across the two studies. The models for accuracy and confidence were more variable, although in both cases the use of the model as a predictor is better than the mean. One way to overcome the low r-squared values would be to substitute for the combined model ('BCDEF') in these cases. However, some caution should be exercised with this, since the two models varied and it was difficult to gauge the importance of this variation. Although for response times, the models proved totally incompatible and this was reflected in the combined model, so substitution in the case of accuracy and confidence would probably be acceptable. None of the models accounted for 100% of the variance seen in the performance measures, and so further improvement could be made[30]. Whilst the models here demonstrate the principle of performance predictability and would be a good starting point for guidance (see Figure 8-10), further research and analysis would be needed in order to fully understand the contribution of predictors, and in particular differences between intra- and inter-object differences.

Inter-Object Judgements (Study 1):

Response Time = (0.44 x Gradient) + (0.46 x Magnitude Difference)

Accuracy = (0.07 x Gradient) + (0.48 x Magnitude Difference)

Confidence = (0.10 x Gradient) + (0.66 x Magnitude Difference)

Intra-Object Judgements (Study 3)

Response Time = (0.65 x Gradient) + (0.13 x Magnitude Difference) + (0.02 x Width)

Accuracy = (0.27 x Gradient) + (0.46 x Magnitude Difference) + (0.06 x Width)

Confidence = (0.33 x Gradient) + (0.27 x Magnitude Difference) + (0.00 x Width)

Figure 8-10: Predictive models for inter- and intra-object judgements

[30] Some exploratory analysis was undertaken using other predictors such as curvature and direction of change, however nothing significant could be established and further work would be needed. It may also be the case that some of these factors, like width, may only be of significance at higher levels of magnitude or require more cases in order to achieve significance.

8.5 Summary of Chapter

The work reported in this chapter showed that stimulus gradient is an effective predictor of threshold gradient, and the theory outlined in Chapter 4 is supported. It was observed that different curve shapes with the same gradient require the same increase in gradient for a change in curve shape to be perceived, whilst curve shapes with different gradients require different changes in gradient in order to be perceived. However, it was also found that Weber Fractions for different gradients were the same, which means that for Study 3 conformance with Weber's Law was found. This differs from Study 1 where there was no conformance. It was thought that this was due to the nature of the comparisons undertaken; that is in the first study inter-object differences were judged whilst in the second study intra-object difference were compared. As a result of this the model developed from the Study 1 data was poor at predicting the thresholds observed in Study 3. However, when the visual-haptic data from Study 3 was regressed, stimulus gradient was again found to be a highly significant predictor of threshold gradient. This means that whilst different models need to be used for inter- and intra-object comparisons, thresholds can be accurately predicted. In relation to predicting performance it was found that magnitude difference, stimulus gradient, and stimulus width were effective predictors of response time, accuracy, and confidence. However, the level of fit varied (R^2=55-90%), and was not as good as that achieved for predicting threshold (R^2=99%), although all models were significantly better than using the mean. Again it was found that different models were needed for inter- and intra-object predictions.

8.6 Related Chapters

Further work that could be considered in relation to developing these predictive models is discussed in Chapter 10.

Chapter 9: Evaluation of Multimodal Interfaces for Shape Exploration – An Emergent Framework

9.1 About this Chapter

This chapter provides a synthesis of the experience gained whilst conducting this research in order to suggest a framework for the evaluation of multimodal interfaces for shape exploration. This framework is summarised in Figure 9-1, and can be used to access the relevant information within this chapter.

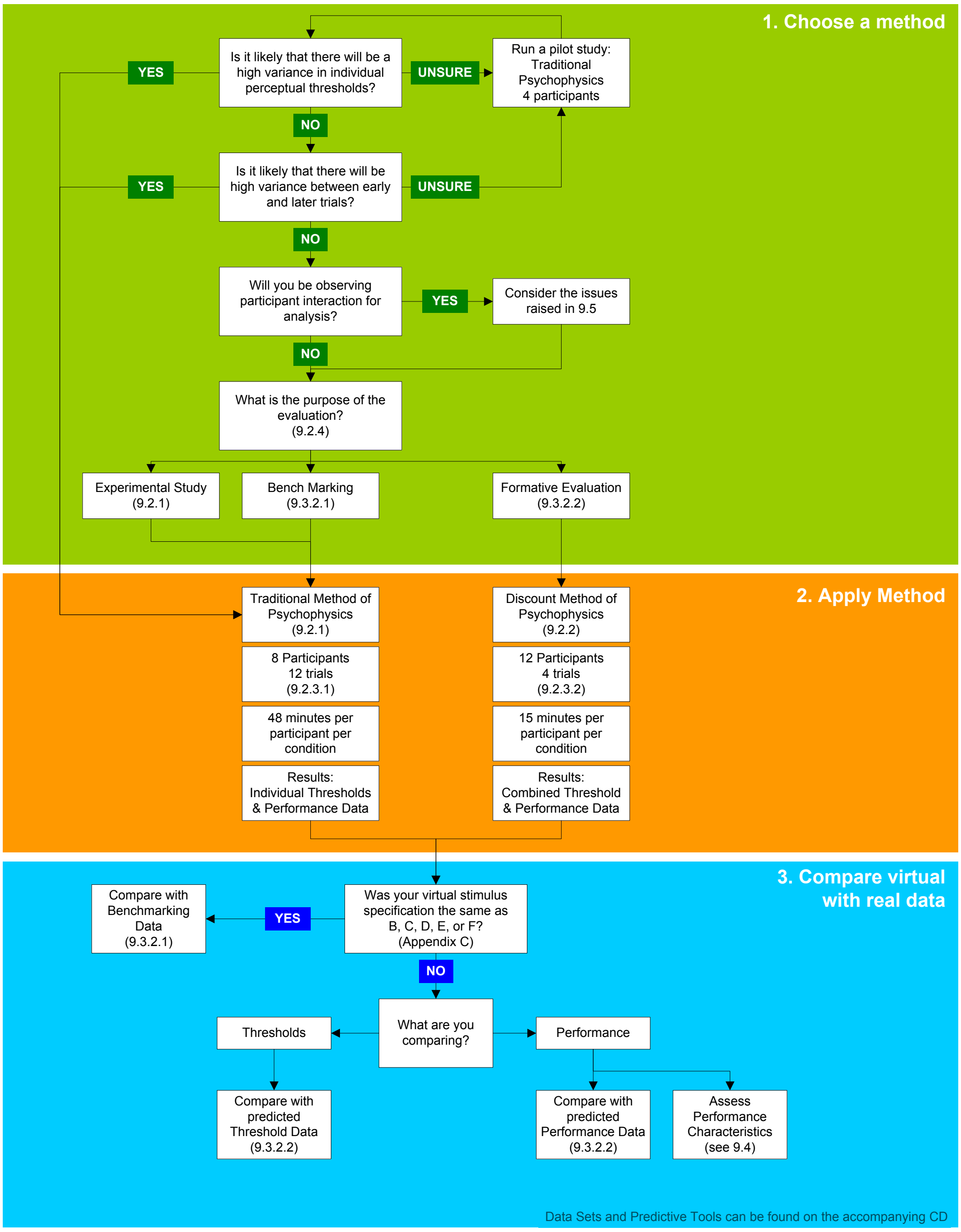


Figure 9-1: Evaluation Framework

9.2 Evaluation Method

9.2.1 The Use of Traditional Psychophysics

The traditional psychophysics method was selected as it is the de facto standard for perceptual threshold investigation. The background and use of this method were discussed in the literature review and will not be discussed further here (see Chapter 2). The particular psychophysics method selected for use in these studies was that of 'constant stimuli'. The exact method used is detailed in the experimental chapters (see Chapter 3 and Chapter 7). However, a brief review would probably be helpful at this point. The method uses a standard stimulus and compares this against comparison stimuli (3 to 4 stimuli above and below the standard). A number of trials are conducted which compare the standard to the comparison stimuli. Participants are required to judge whether the stimuli are more or less curved (for example) than the standard. From these answers the difference threshold was calculated, and the point of just noticeable difference (JND) identified. This was the point at which, for 75% of the time, participants answered correctly; that is they perceived a difference.

This method proved useful not only in calculating difference thresholds and providing the level of perceptual acuity, but also gave a robust context in which to judge performance. One of the difficulties of assessing performance is that it is highly dependent upon the difficulty of the task. That is to say, hard tasks may take longer or have a lower chance of success. Therefore in order to assess performance in a meaningful way the level of task difficulty needs to be known. For perceptual experiments the nearer the comparison stimulus is to the standard the less successful judgments about differences between the two become. So the level of difference between the stimuli to be judged is a good indicator of task difficulty. In the studies undertaken here this difference was quantified as 'magnitude difference' and so performance at differing levels of difficulty could be assessed. This was also a good way to relate the threshold results to those of performance. As once a JND has been calculated, by use of its associated Weber Fraction (percentage change from standard stimuli required to perceive a difference) the level of associated performance could be estimated (through the predictive model described in Chapter 8).

The drawback of this method is that it is time consuming and therefore expensive to apply, and so may not be suitable for all evaluative contexts. However, the possibility of a more streamlined approach has undergone preliminary investigation through exploration of the data from Study 1, and this is discussed next.

9.2.2 Is 'Discount' Psychophysics Possible?

The term 'Discount Usability' was coined by Jacob Nielson (1993) to describe a cost-effective way of conducting usability evaluations as an alternative to user-testing. The basis of his 'discount' method was that statistically about 3-6 users will identify 80% of usability issues, and so would be as effective as user testing, requiring 20-plus users for statistically valid results, for identifying the majority of problems. The cost benefit here is that it is far cheaper to identify the majority of issues for practical purposes, whilst recognising that not all problems will be captured and that the issues identified may not have statistical significance.

Traditional psychophysics methods require a high level of resources in order to complete. They are designed to provide a high level of accuracy in determining either detection or difference thresholds in sense perception. The main reason for the expense of such experimental evaluation is the number of trials needed in order to determine the threshold value. There are two reasons for the high number of trials. Firstly, the threshold is based upon a probability, and as such requires sufficient trials in order to capture the true mean. Secondly, the threshold is based on the aggregation of a number of individual thresholds. This is necessary since individual perception may vary, and so in order to get a good estimate of the population threshold a number of individual thresholds need to be considered. It can be seen that because of this the number of trials can quickly exceed what is practically feasible (in terms of time and cost).

A 'discount' method would require that the number of trials were drastically reduced, so that the overall process becomes less resource intensive. There are possibly two ways in which this could be achieved, one would be to reduce the number of trials needed to capture an individual threshold and another would be to reduce the number of individual thresholds used to calculate the overall threshold. Reductions in either of these areas are problematic. Reducing the number of trials to calculate an individual threshold would mean a reduction in trials per comparison stimuli. The effect of this would be that the mean for each comparison stimulus would have a high standard error (see 9.2.3.1 for detailed discussion of this). Reducing the number of individual thresholds may be possible. The studies here have used thresholds generated from eight individuals; however within the literature (see Table 9-1) these have been calculated from as few as three. The problem however is not so much in the generation of the overall threshold, but in terms of the overall experimental design. What is of consequence is how these thresholds are used. For example, the intention within each of the studies detailed in Chapter 3 (Study 1) and Chapter 7 (Study 3) was to assess the effect of a number of conditions upon difference threshold. As such, they would be analysed using an ANOVA which would require a reasonable sample size. So by reducing the number of thresholds the ability to use inferential statistics is diminished.

If however there is little need for individual thresholds then there still may be a route to a ‘discount’ method. This would be by using a number of individuals to calculate a combined threshold rather than using the mean of individual thresholds. This would do away with the need for a significant number of trials, and so alleviate the resource demands of psychophysical testing. However there are a number of prerequisites that would need to be satisfied in order for this to be valid; individual thresholds should not be highly variable, and early trials should not vary significantly from late trials.

The first of these prerequisite may be difficult to prove statistically, and is certainly so from the limited data provided by these studies. This is because only one threshold per person has been generated and so differences between individual means cannot be calculated. What would be required is that a single threshold is repeatedly tested by a number of people. This would then generate sufficient individual means of the same threshold to be tested statistically against those of other individuals. As has already been stated, psychophysical testing is resource intensive, and this type of study would probably be prohibitive to run. However, if looked at in a different way, there may be sufficient evidence already from which we may draw some conclusion as to the closeness of individual thresholds. For Study 1, comparison was made between high and low curvatures. For example, the Weber Fraction for low curvature in the haptic condition was 22.43 (SD=0.59), and for high curvature it was 23.10 (SD=0.63). These Weber Fractions are very close to each other, however they were found to be significantly different from each other in an analysis of variance. This meant that there was more variation between means for different curvatures than between the means for the same curvature. So whilst there is little apparent difference between the Weber Fractions for different curvatures, the amount of variation between participants was sufficiently low for this to be significant. Similar results were also found for the visual and visual-haptic conditions. This should be sufficient to conclude that there is little variation between individual thresholds, and that the first prerequisite has been fulfilled.

The second prerequisite was that early trials should not vary from late trials. In other words, there should be no significant learning effect. Otherwise early trials would not be counterbalanced by later trials and an inflated threshold may be generated. So if just a few trials were used they would not necessarily be a good reflection of actual perceptual ability. This was explored by examining haptic data from Study 1 (see Appendix O for results), and it was found that there was no significant difference between early, mid, or late trials. This means that there is no apparent learning effect of earlier trials on later trials. Therefore the second prerequisite was fulfilled.

Given these results it is possible to propose a ‘discount’ psychophysical method. This would apply the same protocol as the traditional method, but with reduced trials per person and would generate the threshold from the accumulated data of a number of participants, rather than using the mean of individual thresholds.

9.2.3 Number of Participants and Trials

The number of participants and trials will depend upon whether a traditional or ‘discount’ method of psychophysics is being used, and so each will be considered separately.

9.2.3.1 Traditional Method

Within the psychophysics literature it is difficult to ascertain what is considered to be the appropriate number of participants or the number of trials necessary in order to achieve statistical significance for the results. It is recommended by Rose (2006) that about 50 trials per stimulus intensity are required in order to achieve significant results. However, it should be noted that there are no hard rules, and it will very much depend upon the sense that is being investigated, number of conditions, and the strength of the effect that is being observed. In addition to this there are the practical constraints that operate for any experimental context i.e. time and money. Therefore, the approach taken here was to survey the reported figures within the haptic literature and formulate a number based on that. As can be seen from Table 9-1, the number of participants and trials varies. The average number of trials per standard stimulus per person was approximately 84 (excluding Gordon and Morrison (1982) which is atypical). The time taken to run this number of trials would be in the region of 45 minutes.

Table 9-1: Comparison of number of participants, comparison stimuli, and trials

Studies		Stimuli				Trials per Threshold	
Author	Date	Participants	Standard	Comparison	Trials	per person	Total
Davidson	1972	16	1	4	9	36	576
Gordon & Morrison	1982	6	2	24	20	480	2880
Goodwin et al	1991	6	2	6	20	120	720
Goodwin & Wheat	1992	5	2	6	20	120	600
Kappers & Koenderink	1996	4	1	18	8	144	576
Pont et al	1997	3	3	8	8	64	192
Pont et al	1999	4	1	7	8	56	224
Louw et al	2000	3	3	8	16	128	384
Louw et al	2002	4	2	8	12	96	384
Study 1*		8	2	8	12	96	768
Study 3*		8	3	8	12	96	768

*Study 1 and Study 3 data is included here for information purposes

In itself this does not seem prohibitive, however it must be remembered that this would be multiplied by the number of thresholds to be calculated and conditions. For Study 1 this meant a total of 2 thresholds x 3 conditions x 45 minutes, giving a total time of 4.5 hours per participant. This was beginning to look a little more resource intensive, particularly as the planned 8 participants would take a total of 36 hours to complete. It was therefore decided that an initial pilot study (using 4 participants) should examine the statistical validity of using 48 trials per person (8 comparison stimuli x 6 trials). This was judged by examining the reduction in standard error of the mean for successive trials. The aim was to identify at what point the reduction in standard error started to flatten out and the increase in trials began to produce less benefit. It was found that with 6 trials that the standard error had not started to flatten out and was still reducing at a substantial rate (see Figure 9-2 (i)). It was therefore decided to run a second pilot (using a different 4 participants) doubling the number of trials to 96 per person (8 comparison stimuli x 12 trials). It was found that between 6 and 12 trials the rate of decrease in standard error lessened and appeared to be flattening out (see Figure 9-2 (ii)). On this basis it was decided that Study 1 should be run using 96 trials per person per threshold. As some of the participants in this study also took part in the pilots it meant that additional data were available, and that the decrease in standard error could be calculated up to a total of 18 trials. As can be seen in Figure 9-2 (ii), the standard error continues to reduce between 13 and 18 trials; however the rate of this reduction is comparatively small (see Table 9-2). It is likely that this reduction will continue as the number of trials increase, however the rate of this reduction is likely to become increasingly smaller. From the data presented it is clear that 18 trials produce a slightly lower standard error than 12 trials, however the cost/benefit of this increase is debateable. The range of trials (8 to 20) shown in the literature fall within a zone of flattening standard error as shown in Figure

9-2 (ii) and so represent a range of viable choices. At the lower end of this range there are still substantial changes in standard error and so should be avoided if resources allow. However, given the small increase obtained versus the cost of running additional trials, a level of around 12 trials is recommended.

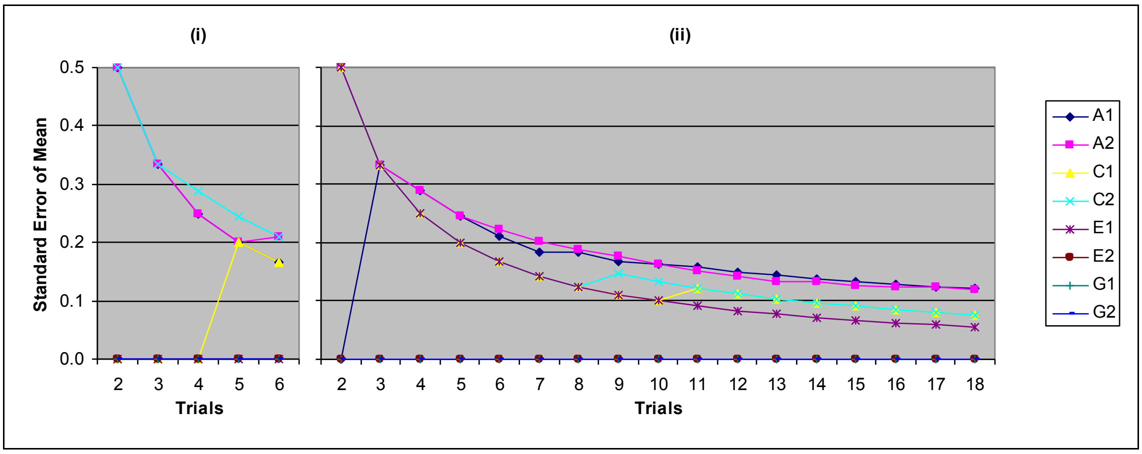


Figure 9-2: Standard Error with increase in number of trials

(i) shows the data for participant 5 (ii) shows the data for participant 8

Table 9-2: Rate of change in Standard Error for each comparison stimulus

Trials	Rate of Change per Trial								
	A1	A2	C1	C2	E1	E2	G1	G2	All
2 to 6	0.042	0.045	0.033	0.033	0.033	0.000	0.000	0.000	0.023
7 to 12	0.006	0.010	0.005	0.005	0.010	0.000	0.000	0.000	0.005
13 to 18	0.004	0.002	0.005	0.005	0.004	0.000	0.000	0.000	0.002

9.2.3.2 Discount Method

For this method individual thresholds are not being considered as the threshold is derived from participants' combined trials. In order to decide on the number of participants required, it is necessary to identify the number of trials required to produce a low and flattening standard error (the relevance of this is explained in 9.2.3.1), so what is of interest is the number of trials necessary in order to achieve a reasonable degree of standard error. Study 1 haptic data were used to generate a plot of accumulated standard error across 84 trials (seven participants undertaking 12 trials each). The trials were plotted successively in order i.e. trial 1 for each participant then trial 2 for each participant until all trials had been plotted (see Figure 9-3). The aim was to identify at what point the reduction in standard error started to flatten out and the increase in trials began to produce less benefit. It can be seen that between 2 and about 20 trials that the standard error reduces at a substantial rate. This declined further to about 40 trials, at which point the reduction in standard error became increasingly small. From 70 trials onward there was little appreciable reduction for any of the comparison stimuli and standard error showed signs of flattening out. The mean standard error at this point was 0.03. From this it was possible to conclude that the amount of trials required would be in the region of 40 to 70. In order to decide on the exact figure, the rate of change per trial was examined for groupings of 12 trials across the 84 trial range (see Table 9-3). This showed that the largest reduction in the rate of change was between '2 to 13' and '14 to 25', and that from '38 to 49' trials onwards the rate of reduction per trial was minimal (taking the mean of all stimuli). It would therefore be optimum (in terms of accuracy and economy) to undertake somewhere in the region of 40 to 50 trials. The minimum number of trials per person should be four. This would ensure that trials of stimuli could be counter-balanced for right and left placement, and that more than one trial took place in each position. Given this recommendation, around 12 participants would be needed to undertake 4 trials each for every comparison stimuli. This means that each person would undertake 32 trials to produce a combined threshold. The time to run such a set of trials would be about 15 minutes per participant, which makes it more feasible to run alongside other formative tests. Using the traditional method would take approximately 48 minutes per participant.

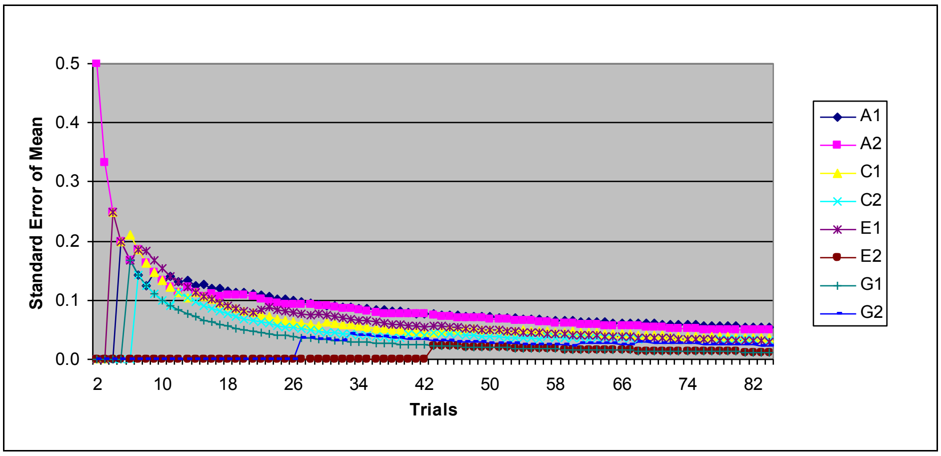


Figure 9-3: Standard Error with increase in number of trials

Table 9-3: Rate of change in Standard Error for each comparison stimulus

Trials	Rate of Change per Trial								
	A1	A2	C1	C2	E1	E2	G1	G2	All
2 to 13	0.0111	0.0101	0.0087	0.0087	0.0101	0.0000	0.0064	0.0000	0.007
14 to 25	0.0020	0.0018	0.0040	0.0035	0.0027	0.0000	0.0026	0.0000	0.002
26 to 37	0.0013	0.0012	0.0010	0.0013	0.0014	0.0000	0.0010	-0.0031	0.001
38 to 49	0.0008	0.0006	0.0000	-0.0002	0.0008	-0.0017	0.0005	0.0007	0.000
50 to 61	0.0006	0.0007	0.0003	0.0003	0.0007	0.0003	0.0003	0.0000	0.000
62 to 73	0.0004	0.0005	0.0005	0.0000	0.0005	0.0002	0.0002	0.0001	0.000
74 to 84	0.0004	0.0003	0.0002	0.0002	0.0004	0.0001	0.0001	-0.0006	0.000

Note: the figures show the rate of change per trial averaged across the group of trials. Some figures show a negative sign, and this means that there was actually an increase in SE. These reversals are expected and usually occur for high accuracy comparison stimuli i.e. those that would usually score a 100% but the odd incorrect answer is given. What matters is not that these reversals occur, but that there are sufficient trials that they do not unduly skew the data.

9.1.1 Choosing between Traditional and Discount Methods

Traditional psychophysics or the discounted method may be used within this framework, and both have their particular benefits. Traditional psychophysics is a proven method and is the de facto standard for assessment of perceptual thresholds. It will generate data that can be analysed to produce inferential statistics so is appropriate to experimental designs. However, in order to do this requires a high level of resources (as explained in 9.2.3.1). Because of this it may not be suitable to all contexts of enquiry, and is best reserved for investigating issues of human perceptual acuity, or where the experiment is tackling issues of a fundamental nature where precision and generalisability are required. This was the choice made when undertaking Study 1 and Study 3, since these were fundamental investigations into levels of perceptual acuity and associated performance. It was envisaged that the results would be used to make wider inferences and so needed to have statistical validity.

The discount method is far more economical in relation to resources, and so provides a fast method suited to formative evaluation (for the purposes of informing development). The weakness of this method is that it does not lend itself to experimental designs as it will not produce data that can be analysed to produce inferential statistics. However, it is envisaged that this type of method would be used for formative assessment of interfaces and would therefore provide sufficiently indicative results. There would also be some scope for comparative analysis of data as confidence intervals can be produced for each threshold generated by regression analysis. This would allow, for example, some sense of whether the thresholds produced by different prototype interfaces were similar or different (see Figure 8-2 for an example of this type of assessment).

To summarise, if precision and inference are required then the traditional method should be selected. If economy and indicative results are required the discount method should be considered. The traditional method is best suited to fundamental issues of perceptual acuity, whereas the discount method is perhaps best reserved for formative assessments of interface usability.

9.3 The Evaluation of Multimodal Interfaces using Difference Threshold and Performance Metrics

9.3.1 Choice of Variables

9.3.1.1 Performance Measures (dependent variables)

The metrics produced relate to three measures of performance; response time, accuracy, and confidence. The first of these are standard measures used for usability assessment and are usually referred to as efficiency and effectiveness (ANSI/NCITS 354-2001 2001; ISO/IEC 25062 2006). These help to identify where an interface may be performing too slowly or is not achieving an appropriate level of accuracy. The third measure, confidence, is not well known as a usability requirement. It was selected as it was felt to be a useful indicator in a number of respects. Firstly, it is vital that a user feels secure in their interpretation of information presented by a system, particularly in a high-cost context such as design. This measure would give an overall sense of judgment confidence for a particular interface, so can differentiate between similar interfaces. For example, an interface may be highly rated for efficiency and effectiveness but may not instil a sense of confidence in the user about that performance. Secondly, by using it to breakdown other performance measures, categorised by high and low confidence, it can be used to determine where feedback might be improved or where feedback is giving a false sense of performance (see 9.4.2). In a similar way its positive correlation with accuracy will also indicate whether a user had a realistic perception of their efforts e.g. a highly confident performance correlates to a highly accurate one.

9.3.1.2 Predictors (Independent variables)

The predictor for difference threshold was found to be gradient, and the reasoning behind this was discussed in Chapter 4. Briefly, gradient was shown to be the active dimension in human perception of changes in curve shape (Gordon and Morison 1982; Pont, Kappers et al. 1997). As such it was proposed as a predictor of difference threshold. Through the experimental investigation undertaken in Study 3 it was shown this was the case and through regression analysis a model was derived to predict that change. It was also found that the model for difference threshold varied depending on whether comparisons were made between (inter) or within (intra) objects, and so two predictive models were necessary.

Predictors for performance were explored experimentally in Study 3, based on insights gained in Study 1. Dimensions of the stimulus that were investigated were curvature, gradient, and width. In addition, it was thought that magnitude difference also had a role to play. Through multiple regression analysis it was found that between 60-90% of variance in performance could be accounted for by a combination of gradient, magnitude difference, and width. These were therefore used as predictors in the model for calculating performance metrics. Again it was found that there was a difference between intra and inter-object performance and so two predictive models were necessary.

9.3.2 Type of Evaluation, Comparative Data, and Predictive Models

There are two scenarios of evaluation envisaged within this framework. The first is in benchmarking the performance of multimodal interfaces. The second is in the usability assessment of multimodal interfaces. It should be noted that these are not mutually exclusive scenarios and are more a convenient categorisation in order to discuss the applications of comparative data and predictive models produced by the studies undertaken.

9.3.2.1 Benchmarking

There are a range of haptic and visual devices available that may be used within multimodal interfaces. However, the suitability of this equipment can be difficult to discern from manufacturers' specifications. In such cases it may be desirable to benchmark such equipment prior to incorporation into a prototype interface. The data produced during the course of this research can be used to provide comparative results against which devices can be assessed (for comprehensive data sets see CD /benchmarking-data/). It is important to note that some of these are for inter-object and some are for intra-object evaluation, since thresholds and performance were found to be different for each. The dimensions of the stimuli are given in Appendix C and can be used to produce the virtual stimuli required for testing. A traditional method of psychophysics should be used for benchmarking as this will provide the necessary statistical validity.

9.3.2.2 Formative Evaluation

Where no pre-existing interface exists it can be difficult to assess the level of performance achieved by a prototype. The threshold and performance predictors produced as a result of this research are aimed at providing comparative data to help in this type of situation (for threshold and performance prediction tools see accompanying CD ‘Prediction-Tools’). Unlike the fixed metrics used in benchmarking, the predictors can produce thresholds across a range of stimuli dimensions i.e. for different heights, widths, curvatures, and gradients (for stimuli in a range of 0^{o} to 27^{o}). They are also able to provide a series of performance metrics dependent on the magnitude difference specified i.e. for small or large differences between stimuli. It is important to note that some of these are for inter-object and some are for intra-object evaluation since thresholds and performance were found to be different for each. Given the formative nature of this testing, and that this is likely to form a small part of a testing protocol, the discount method of psychophysics should be used.

9.4 Characterising Performance and its use in Evaluating Multimodal Interfaces

9.4.1 Performance Characteristics

The use of performance characteristics was another way in which the effect of combining modalities could be evaluated. The premise was that different types of judgment could be characterised in different ways. For instance, it was observed that correct judgements were faster than incorrect judgements, and that high confidence judgements were faster than low confidence ones (see Chapter 3 Section 3.3.3). However these characteristics were seen to change in the presence of the ‘curvature’ sound. Under the influence of this sound there was no difference in response time across these different categories. It is therefore thought that when sounds are highly efficient there is no difference in response time between categories. This was also seen when comparing High Confidence Correct (HCC) answers to each of Low Confidence Correct (LCC), High Confidence Incorrect (HCI), and Low Confidence Incorrect (LCI). It is therefore thought that to breakdown results in this way is useful for discerning optimal and sub-optimal performance.

It is important to note that whilst these differences were significant, this is not to imply a causal relationship between either confidence and response time or accuracy and response time. However, whilst there was no significant association between response time and either level of accuracy or confidence i.e. percentage correct or percentage confidence, there was an association between the levels of confidence and accuracy themselves. This characteristic broke down when combining certain modalities, visual-haptic and curve shape sound, in Study 3. Again this type of characteristic of performance may be useful in identifying where combinations of modalities are problematic.

Whilst there were some interesting patterns amongst the performance characteristics, it is acknowledged that these come from a limited range of studies and may not be apparent elsewhere. However, it would be worthwhile monitoring these characteristics across further studies to discern their use as an indicator of performance issues.

9.4.2 Feedback & User Error

The performance characteristics described above are in need of further validation in order to prove their usefulness. However a more substantive use for the classification of judgements by confidence (High or Low) and accuracy (Correct or Incorrect) was found when exploring the data from Study 1 and Study 3. Whilst these were not reported in the main study results (because of their explorative nature), they are nonetheless useful in describing facets of interface usability and so are discussed here. There were two ways in which data were classified that proved useful.

The first was by categorising accuracy data by confidence level in order to differentiate between good and poor feedback. The term ‘Good Feedback’ refers to where participants had high confidence in their judgement and gave a correct answer. Whereas, ‘Poor Feedback’ was where the participant had given a correct answer, but their confidence in this judgement was low. It is assumed that a partial reason for this low confidence is insufficient or ‘poor’ feedback from the interface with regard to the target of the given judgement. This was explored through an area graph as shown in Figure 9-4. The data shown are those for accuracy level and the coloured areas correspond to the total accuracy level for a given condition. This is categorised into high confidence correct (dark green) and low confidence correct (light green) judgements. Thus, whilst the participant may have scored highly overall for accuracy, they may only have confidence in a correct response for part of this score. For the remainder of their judgements there is uncertainty. This uncertainty level may be reduced by improving feedback to the participant so that they change from a position of low confidence in their judgement (uncertainty about their perceptions) to a position of high confidence (certainty in their perceptions). In other words, the feedback they are receiving from the interface (or object) has been improved. This is seen in Figure 9-4 with an increase in ‘Good Feedback’ (HCC) and a decrease in ‘Poor Feedback’ (LCC). These areas of good and poor feedback (HCC and LCC) can be compared statistically through the use of analysis of variance. In this way different mechanisms of feedback can be evaluated. In the case shown, the use of a 'curvature' sonification to supplement visual-haptic perception has improved feedback substantially and virtually eliminated uncertainty (poor feedback).

The second method was to categorise confidence data by accuracy in order to differentiate between 'Good Judgement' and 'Over-Confidence'. The term 'Good Judgement' refers to where participants considered that they had high confidence in making the judgement, and the judgement was correct. Whereas, 'Over-Confidence' is considered to be those judgements where the participant had high confidence in their judgement, but the judgement was found to be incorrect. This was explored through an area graph as shown in Figure 9-5. The data shown are those for high confidence and the coloured areas correspond to the total high confidence level for a given condition. This is categorised into high confidence correct (dark green) and high confidence incorrect (red) judgements. Thus, whilst the participant may have been highly confident in their judgements, only some of this confidence may have been justified by providing a correct judgement. The remainder may be termed 'over-confidence' as the judgement is incorrect. This is useful in understanding whether the feedback given in a particular interface is misleading; that is, it causes the user to be confident in their judgement where it is incorrect. To an extent there will always be an area of over-confidence as we do not always exercise good judgement even though we may believe it to be. However, through effective feedback this area of over-confidence should be diminished. This can be seen in Figure 9-5 where the use of sound feedback ('curvature' sonification) increases 'Good Judgement' (HCC) and practically eliminates 'Over-Confidence' (HCI). Conversely, when a 'Curve Shape' sonification is used to provide sound feedback the area of 'Over-Confidence' is increased, and indicates that the user has been misled by the feedback or it is poorly understood.

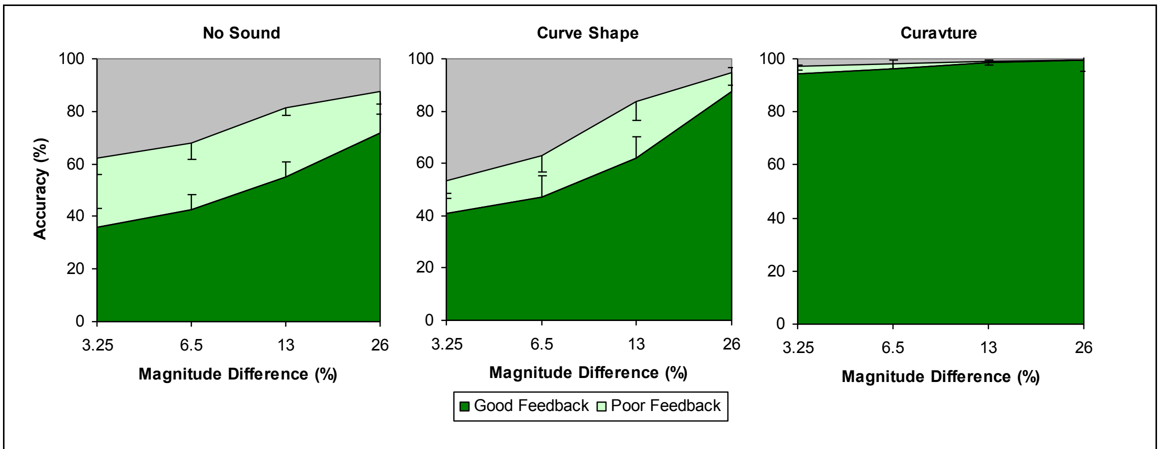


Figure 9-4: Using confidence level to assess improvement in feedback

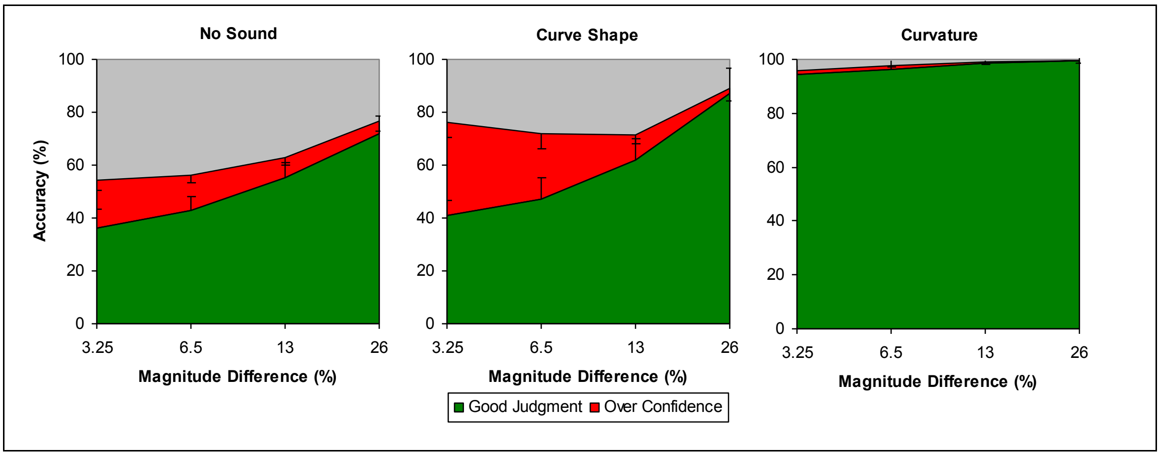


Figure 9-5: Using accuracy level to assess misleading feedback

The use of these two analyses in tandem helps to identify where feedback has been improved, but will also warn of where it may be misleading. So in the example shown in Figure 9-4, it can be seen that the addition of sound feedback in the form of a 'Curve Shape' sonification makes no appreciable difference to the level of accuracy. However, there is indication from Figure 9-5 that it may have misled the user into being 'Over-Confident' in their judgements which is an undesired effect of the feedback. In this way it is possible to gain a more complete understanding of the effects of different forms of feedback.

9.5 Observing Interaction

A number of methods were used in order to assess various aspects of interaction. These ranged from a simple observation of comparisons made between stimuli in Study 1, to a more complex automated tracking of interaction developed for Study 3. Whilst the use of observational techniques was limited within this research, they did prove useful for understanding multimodal interaction and so are included as part of this framework.

9.5.1 Direct Observation

The use of direct observation methods was found to have had varying degrees of success. In Study 1 the number of comparisons made between stimuli was quantified. In some respects this was a preliminary exercise towards understanding what was feasible to observe, and how well it might be captured. The method was quite straight forward and involved counting hand or eye movement between the stimuli blocks. However, when it came to the visual-haptic condition, the limitations of having only one observer were realised; either hand or eye movement could be counted but not both. Because of this, potentially important data were not captured, and the effect of combining modalities on interaction could not be fully analysed. Additionally, it was felt that whilst interactions were captured the level of accuracy could not be guaranteed. In retrospect, because of these limitations, it may have been better to have undertaken post-experiment analysis of video footage in a software package such as Observer XT (Noldus 2009).This would have allowed for accurate capture of the observed interactions, and also allowed more complex analysis of them.

Direct observation was also used to categorise and quantify the types of interaction used by participants during Study 2 (see Chapter 6) and Study 3 (see Chapter 7). The initial study was used to identify and propose different categories of interaction. The second study used these categories to quantify and monitor the effect of sound on interaction. These tasks were easy to perform and were well suited to human observation (the need assign a category). The data obtained from 'in situ' observation were accurately recorded and proved suitable to answer the research questions. Again post-evaluation analysis may have been able to probe more complex associations between interaction, stimuli, and sound but this would have been at a significant time cost. In some senses this highlights the dilemma when undertaking these types of observational studies; quick and simple, or time-consuming and complex. Given the initial nature of this experimental work, the quick and simple approach proved useful. This was able to identify issues that could be tackled by further analysis of data (using Observer XT for example) or used in defining further studies.

9.5.2 Mediated Observation

It was apparent from undertaking observations in Study 1 that it was difficult for the human observer to note all that was happening for a given interaction. As suggested above, this may be remedied through post-evaluation analysis using specialist software. However, the nature of the interaction would still need to be viewable by a human observer. In Study 3 there was a requirement to observe the duration of interaction across the stimulus and identify the duration of exploration in various zones. The problem presented in Study 3 was that the nature of this interaction made it difficult to observe directly. Therefore a software tool was developed that tracked the participant's movement across the stimuli and collected data as to the position and time spent in each zone (see Chapter 7 Section 7.2.2). In this way the interaction was mediated and made observable.

The strength of mediated observation is that it can reveal the nature of the actual interaction rather than the apparent interaction that would hitherto have been observed by a person. For example, whilst observing participant interaction with the stimuli there was no apparent bias in exploration duration between zones. However, the tool revealed that participants had spent a significantly longer time exploring the middle and end zones of the stimulus as opposed to the slopes. Through using technology to mediate the observation, important details were captured and insights gained that would have been lost through human observation alone.

Unfortunately one aspect of interaction was not able to be captured during Study 3. This was the role of vision during multimodal interaction and how this was affected by the addition of sound. This is mentioned here as it is believed that the use of eye-tracking may have provided useful insights in relation to this, and future studies may be helped by including this type of technology.

9.5.3 Recommendations for Observing Multimodal Interaction

The observation techniques that have been used within this research have been categorised as direct and mediated. This differentiation is felt to be important as it recognises the limitations of the human observer. In this sense they also provide a convenient way to apportion use. Direct methods are ideally suited to situations where the interactions are observable either in situ or through post-evaluation analysis. The former of these circumstances is ideal for formative or exploratory work, whereas the latter lends itself to more defined experimental conditions. In particular, post-evaluation analysis is useful for ensuring accuracy of observations and analysing data that may occur across dispersed events (for instance relationship of interaction to particular conditions or sets of conditions). It is also useful for events that happen in rapid succession or where multiple events are happening simultaneously; these are both particular features of multimodal interaction.

Leading on from this there are techniques that use technology to mediate observations. That is they reveal data that could not be observed directly, and present it in a form that is observable. Within this research, mediated observation was used to discover the differential focus of haptic exploration on certain areas of the stimulus. This could not have been revealed by direct observation methods, and so underlines the value of this type of technique. However, other than eye-tracking technology which is now widely available, using technology to mediate observations is likely to involve bespoke development which may be a barrier to use. It also requires prior investigation to properly identify areas of research interest, as was the case for Study 3 which followed up questions raised about interaction from Study 1.

Thus, the recommendations for observing multimodal interaction may be summarised as follows: observe what happens (in situ), record and explore what happened (post-evaluation), and finally, investigate with the aid of technology (mediated observation). Finally, it should be stressed that there are subtleties of interaction that are too quick, too complex, or too dispersed to observe without mediation or post-evaluation analysis.

9.6 Guidelines for the Development of Multimodal Interfaces

During the course of the research undertaken here (Fundamental Studies – S1, S2, S3) and the evaluations of the SATIN prototype (SATIN Evaluations – E1, E2, E3, E4), a number of observations have been made that would be helpful in guiding the development of multimodal interfaces.

Table 9-4: Guidelines for the development of multimodal interfaces

Guideline	Source
Auditory Feedback	
Where sounds are used in a symbolic way it is important that they are distinct from other such sounds in order to aid memorability.	E2
When made against a background of continuous sound, discrete auditory feedback needs to be of a frequency such that it is perceptible against this.	E2
Simplified sound types, sine wave and harmonic sounds, are preferred by users over complex sounds.	E2, S2
Simplified sound types are easier to perceive resulting in quicker response times and greater accuracy than more complex sounds.	S2
Sounds with harmonic frequencies ('Cello') may improve participant confidence in their judgements.	S2
There is no user preference for use of headphones or speakers when receiving auditory feedback. The choice between these may therefore largely depend on other factors such as, the extent of environmental noise, user-comfort, and collaborative requirements.	E1
Differences between two discrete tones or short chunks of continuous tones are easier to perceive than prolonged auditory feedback.	S2, S3
In order to accurately judge the orientation of a curve, the sonification method should provide a substantial rate of change in the information displayed (within the first second) in order to facilitate perception.	S1
Multimodal Interaction	
Users should have freedom of interaction with the haptic interface (in order that they may adopt an optimum style of interaction to suite their perceptual needs).	S1, S3
Where the hand cannot be seen it is useful to provide a visual surrogate (e.g. pointer) to inform of hand position (absence of such feedback was observed to be disconcerting to users).	E2

Sound may cause distraction or conflict with other modalities, it is important that users have the option to interact without auditory feedback.	E2, S3
Haptic devices should be appropriate to the task and interaction required e.g. they should not require excessive force, should be responsive, and consistent in the feedback produced.	E2, E3
Identification of discontinuities in curvature should (through visual, haptic, and visual-haptic modalities) be made within 30 seconds. A response time of less than 10 seconds would be optimum. Response times in excess of 30 seconds may indicate a poor haptic or visual interface.	E2, S1, S2, S3
For exploratory tasks, user workload should not be above the mid-range of the NASA-TLX scale (45-55).Ratings above this may suggest excessive workload for this type of task, and therefore a poorly performing interface.	E4, S2, S3
Without the aid of auditory feedback, there needs to be a curvature change of about 23% in order for a difference in curve shape to be detected between objects.	S1, S3
For curvature differences below 23 % some form of additional feedback will be required in order to aid judgement of discontinuity.	S1, S3
The haptic interface should present the whole stimulus in order to allow examination of the end and middle regions.	S3
The interface should allow users to switch between unimodal and multimodal exploration. However, excessive swapping between different modalities may be an indication of a poorly performing interface.	S3

9.7 Chapter Summary

This chapter has suggested a framework that can be used for the evaluation of multimodal interfaces for shape exploration. As well as detailing the traditional psychophysical method it suggests adjustments that can be made to this in order to provide a ‘discount’ method. By adoption of the latter method, evaluation times could be reduced up to 70%. However, it is suggested that this method is confined to more formative evaluation activities and that, for benchmarking interfaces, a more traditional method is used. Within the framework a number of performance metrics are provided. These are either developed from the data or generated by predictive tools (see the accompanying CD). In addition to metrics, it is suggested that certain performance characteristics can be used to assess the impact of interface changes. Perhaps the most useful of these is to use; High Confidence Correct (HCC) and Low Confidence Correct responses to judge ‘Good’ and ‘Poor’ feedback, and HCC with High Confidence Incorrect response to monitor ‘Good Judgment’ and ‘Over Confidence’. It was suggested that these measures may be particularly good at checking that an increase in ‘Good’ feedback is not accompanied by an increase in ‘Over Confidence’, which is not considered desirable. Finally a number of recommendations are made for observing user interaction. In particular it is noted that there are subtleties of interaction that are too quick, too complex, or too dispersed to observe without mediation or post-evaluation analysis, and that direct-observation may be of limited use in observing multimodal interactions.

9.8 Related Chapters

Further work that could be considered in relation to developing these predictive models is discussed in Chapter 10.

Chapter 10: General Discussion

The general discussion is centred on the themes of perception, performance, interaction, and sound, as identified in the introduction, and upon which the objectives were based. It does not seek to replicate the detailed discussion provided within the studies, although there will be some summarisation in order that studies can be compared and general issues identified. This will enable a synthesis of results and will lead towards the presentation of conclusions in Chapter 11.

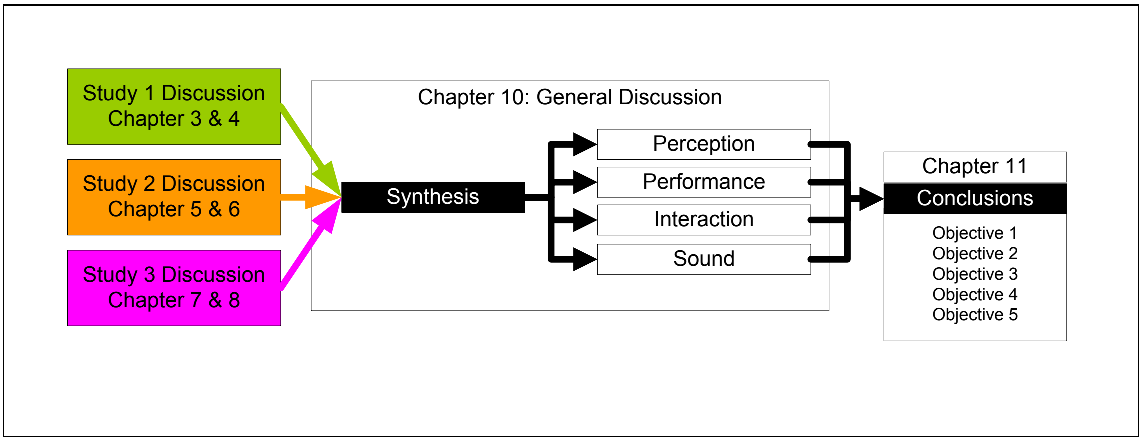


Figure 10-1: Structure of the Thesis - Discussion and Conclusions

10.1 Perception & Multimodality

During the course of studies 1 and 2 a number of unimodal, bimodal, and multimodal conditions were explored. These studies gave a good understanding of the level of perceptual acuity for curvature exploration, and the results are summarised in Table 10-1 for ease of discussion.

Table 10-1: Summary of Weber Fractions from all studies

Modality	Type	Study	WF
Visual	unimodal	1	23.00
Haptic	unimodal	1	22.77
Visual-Haptic	bimodal	1	22.97
Visual-Haptic	bimodal	3	12.28
Visual-Haptic-Sound (Curve Shape)	multimodal	3	12.47
Visual-Haptic-Sound (Curvature)	multimodal	3	2.48

10.1.1 Unimodal versus Bimodal Exploration of Curvature Differences

In Study 1 there was found to be no significant difference between the Weber Fractions for the unimodal or bimodal conditions (see Table 10-1). This is contrary to observations of Ittyerah and Marks (2008) who found that vision had the best acuity, followed by visual-haptic, with haptic being the least acute. It was felt that the design of Study 1 was more reflective of applied conditions (sub-optimal viewpoint) and so the results here are more salient to the practical context of interface design. These show that there are no appreciable differences between the modalities used for curvature exploration (see Table 10-1). However, given the results of Ittyerah and Marks (2008), it is acknowledged that there may be conditions under which the visual modality provides a greater level of perceptual acuity than either unimodal haptic or bimodal visual-haptic.

In considering naturalistic[31] exploration of curvature differences, it may be tempting to consider the use of purely visual inspection and optimise the interface accordingly. Certainly the evidence from Ittyerah and Marks (2008) would be suggestive of such a strategy. However, this would be to ignore the user's desire to touch the object in order to judge its quality (SATIN Consortium 2007). Evidence suggests that this was an important facet of exploration with a significantly increased confidence for visual-haptic judgements (compared to visual or haptic only). Given this, it is considered important that interfaces are developed to allow for this type of naturalistic interaction.

10.1.2 Bimodal versus Multimodal Exploration of Curvature Differences

Study 3 investigated the effect of augmenting visual-haptic exploration with a harmonic sound ('Cello' from Study 2, see 5.1.3). Participants were asked to judge curve shape differences in three conditions; visual-haptic ('No Sound'), visual-haptic plus a sonified curve shape sound ('Curve Shape'), and visual-haptic plus a sonified curvature sound ('Curvature'). It was found that there was no significant difference between Weber Fractions for 'No sound' and 'Curve Shape', but that 'Curvature' was significantly more acute.

The similarity between acuity for the 'No Sound' condition and the 'Curve Shape' condition indicates that there was no real advantage gained by adding this type of sound mapping. In essence, this sonification traces the shape of the object by providing continuous feedback in a similar fashion to haptic exploration. The similarity between

[31] The term 'naturalistic' has been used here to differentiate this from exploration that is augmented with the use of sound.

touch and 'Curve Shape' is also their temporal revelation of the shape over time. At best this type of mapping may be confirmatory, leading to an increase in confidence, and at worst may be confusing with an increase in response time (see 10.2.4). On balance this type of sonification offers little in terms of improvement over visual-haptic exploration and certainly is unable to deliver increased acuity.

The curvature sonification provided a different type of mapping. By mapping curvature to frequency it meant that participants effectively had to judge the difference between two tones. This was a much more simplified task than interpreting the rising and falling tones associated with the curve shape sonification. It could also be judged instantly rather than being revealed over time; in this respect it behaved much more like vision. The multimodal condition was a factor of five times more acute than the bimodal condition. It is clear that the addition of sound with this type of mapping was extremely successful and provided a strong improvement over visual-haptic perception.

The success of the curvature sonification is almost certainly due to the simplification of the task into judging the difference between two tones. The simplification in this sense is twofold; firstly that the comparison is between two distinct tones (as opposed to a continuous stream of changing tones), and secondly that these are of a degree of difference that makes judgement easy. This ties into behaviour observed in Study 2 Experiment 3 where a simplification of the feedback was sought through the adoption of particular interaction strategies (see 10.4 for further discussion of interaction). It is therefore evident that the success of the curvature sound lies in the dichotomous nature of its signal, the clear contrast of which dominates perceptual attention.

However, the question has to be raised as to whether this was truly multimodal (or indeed what we mean by this). This is because three-quarters of participants stated that they did not attend to the other modalities and relied solely on auditory information when making their judgements. In essence, participants made their judgments unimodally. In the 'No Sound' and 'Curve Shape' conditions the same level of participants reported acting bimodally or multimodally. It would appear that participants in these studies chose in a multimodal environment to make a unimodal judgement, since this presented a clearly optimal choice. In the other conditions, bimodal or multimodal judgements were made since no optimal information was considered to be available. This is particularly seen in the 'Curve Shape' condition where practically every combination of modality is used.

It would seem that it is necessary to be broad in our interpretation of multimodal interaction, in that this is not necessarily an equal adoption of all modality inputs but a combination of conscious attention to one or several perceptual modes or the unconscious dominance of one (usually vision). The defining characteristic of multimodality is the availability or choice between, and not necessarily the use of, different modalities (as this may be load and/or task dependent, which may be difficult to predetermine). Across these studies it is evident that, from the range of modalities, the best performing modality has been consciously chosen or subconsciously dominates.

10.1.3 Differences between Intra- and Inter-Object Judgements

A serendipitous finding of this research was a difference in visual-haptic acuity between Study 1 and Study 3. In the first study the Weber Fractions were found to be approximately 23% whereas in the third study they were 12%. In addition to this it was found that there was a conformance to Weber's Law in Study 3. This finding was counter to the results of Study 1, where no conformance was found in the bimodal or separate unimodal conditions.

After exploration of a number of factors that may have biased these results (see 8.4.1), it was considered that it was the nature of the stimuli that had led to the difference in Weber Fractions. In Study 1 the standard and comparison stimuli were separate objects, whereas for Study 2 the standard and comparison were opposite halves of the same object. Effectively, this meant that for Study 1 inter-object differences were being judged and for Study 3 intra-object differences were judged.

These results indicate that people are twice as good at perceiving differences in curve shape within an object as differences that exist between objects. It is difficult to account for this difference other than in the way that participants interacted with the stimuli. In Study 1 whilst the contact with each stimulus was fairly continuous it was necessary to break contact in order to compare the other stimulus. Conversely, for Study 3 the participant was able to remain in constant contact with the stimuli whilst making comparisons between the curve shapes. This would suggest that it may be this momentary break in contact that was responsible for the differences in the levels of acuity. A possible reason for this is that sense information stored in the perceptual working memory system is prone to decay. The length of time that a percept can be held before this decay impairs recall varies, and can also be dependent upon intervening tasks (Ittyerah and Marks 2007; Shih, Dubrowski et al. 2009; Craddock and Lawson

2010). However none of this previous research examines the effect of delay on difference threshold, so further research would be needed in order to understand and quantify the extent of the effect observed within these studies. However, regardless of the mechanism of this effect it is possible to assert that intra-object judgements were far more acute than inter-object judgements.

These findings present issues for experimental design. There is orthodoxy within haptic psychophysical experiments for the presentation of stimuli either in parallel or one after the other. The experimental design in Study 1 followed this approach; however this was varied in Study 3 in a desire to replicate a more applied context. Doing this led to a serendipitous discovery that intra-object perception conformed to Weber's Law and acuity was finer. This may act as a cautionary note to following established methods without question. It may also prompt a re-examination of what is known about haptic[32] perception, which is currently considered not to conform to Weber's Law based on inter-object comparisons (Kappers and Koenderink 1996).

10.1.4 Effective Stimulus for Difference Perception and Prediction of JNDs

In the preceding discussion we have talked about curvature perception. This is because it relates to the context of the SATIN project. However, in the strictest sense it is not curvature that is detected, but the change in gradient. Whilst curvature and gradient have a geometric relationship and so are interrelated, the effective stimulus was the change in gradient. This was demonstrated through an analysis of the Study 1 data, and a further analysis which included data from the haptic literature. Both of these analyses showed a strong relationship between the stimulus gradient and the threshold gradient (r^2=0.99)

The research presented in Study 3 confirmed gradient as the effective stimulus of differentiation, through demonstrating that objects with the same gradient (but different curve shape) require the same level of JND to detect change. This is supported by the work of Louw (2002) who found that shapes with similar gradient were difficult to distinguish from each other. This finding means that gradient, as well as being the

[32] Whilst the disparity was found in relation to visual-haptic JNDs, because there was no significant difference between modalities, then it is thought that this might also apply to haptic JNDs.

effective stimulus for detection of flat from curved (as shown by Gordon and Morrison, 1982), is also the effective stimulus for perceiving difference between curves. However, what still remains unclear is whether gradient or 'attitude difference' (Pont, Kappers et al. 1997) are both effective stimuli, or if one of these is the effective stimulus. The difficulty arises as these are difficult to disambiguate from each other because of their geometric interrelationship (see 2.4.5). Given that they are both related it may not be as important to know which it is since they both produce a similar effect. This means that they are both likely to stimulate the underlying haptic perceptual mechanism in the same way, so to this extent it may be possible to infer that the underlying haptic structures rely on slope differences to detect presence and changes in curvature (see Johnson (2001) for a neuroscience perspective on this). It is also interesting to note that, for the proprioceptive sense, the position of joint angles has been shown to be known with a precision of 0.6-1.1^{o} (van Beers, Sittig et al. 1998). This is in line with the 0.5^{o} detection threshold demonstrated by Gordon and Morrison (1982). Therefore any haptic interface should seek to exploit the underlying biological mechanism. An example of such a device is the haptic interface developed by Wijntjes et al. (2009). This comprised a flat plate that pivoted to display tangent gradient (attitude difference) as it traversed a virtual curve.

10.1.5 Implications for interface design

This research was guided with reference to concerns arising from the SATIN project. This sought to develop a novel interface for the exploration of object shape using haptic and sound feedback. The presence of haptic feedback can be seen as a positive contribution since evidence suggests that it boosts confidence in judgements (see Chapter 3 Section 3.4.2). However, at best, differences of 11% can be perceived within an object, and 23% between objects. To ensure a Class-A surface, designers would need to be able to perceive a difference between curvatures of as little as 0.01/m. It is therefore possible to conclude that visual-haptic judgements alone are not sufficient for this purpose.

A further consideration was that augmentation of visual-haptic exploration with sound may increase acuity. This was shown to be the case with a Weber Fraction of approximately 2% for judgments made with the curvature sonification. It is thought that this may increase further to about 0.3% based on human perceptual limits for differentiation of pitch. Even so, this is still not sufficient for judging discontinuities in Class-A surfaces. However, there is no reason why scaling could not be used to overcome such limitations. This is not appropriate to either visual or haptic conditions as it would change the nature of the object's properties. There are however no such constraints on auditory presentation of information due to its abstract nature. However, careful consideration would be needed of the scaling factor, and user-centred studies would be needed in order to ensure a meaningful fit with users, as indicated by Walker and Nees (2005).

The fact that haptic thresholds have been found to be at best around 11% (Goodwin, John et al. 1991) presents an opportunity for interface design. It means that there can be a fairly high degree of tolerance between the actual and displayed curvature. This is of benefit because the mathematically described curvatures that exist within CAD models cannot be replicated precisely by the currently available haptic technologies (SATIN Consortium 2008). However, it would seem that there is little need to pursue precision when human perceptual ability would be incapable of realising it. To this extent the approach used, for example in the SATIN prototype, presents the benefits of free hand movement at the expense of a precision that is not required and so would seem to be an appropriate trade-off. In addition, given that the effective stimulus is gradient, there may be no need to display curvature at all and an approach such as that demonstrated by Wijntjes (2009), which presented tangent gradient, would be sufficient. However, this type of approach with interaction restricted to one finger may not suit the more naturalistic interaction favoured by end-users (within the product design domain).

10.2 Performance, Multimodality and other Affective Factors

Across the three studies that contribute to this research a number of unimodal, bimodal, and multimodal conditions were explored. The focus of these studies was to discover the effect of different variables (modality, stimulus properties, and magnitude difference) on response time, accuracy, and confidence. The mean results for these studies are summarised in Table 10-2 and Table 10-3.

Table 10-2: Summary of performance from all studies (modality)

Modality	Type	Study	Response Time (s)	Accuracy (%)	Confidence (%)
Haptic	unimodal	1	12.88	84.81	61.46
Visual	unimodal	1	3.63	87.93	74.48
Visual-Haptic	bimodal	1	5.69	89.93	82.90
Sound (Curve Shape)	unimodal	2	5	94	82
Sound (Curvature)	unimodal	2	4.7	93	89
Visual-Sound (Curve Shape)	bimodal	2	28	79	80
Visual-Sound (Curvature)	bimodal	2	28	77	79
Visual-Haptic	bimodal	3	2.78	74.74	62.54
Visual-Haptic-Sound (Curve Shape)	multimodal	3	3.23	73.78	77.08
Visual-Haptic-Sound (Curvature)	multimodal	3	1.39	98.52	98.00

Table 10-3: Summary of Visual-Haptic performance from all studies (gradient)

Stimulus	Gradient	Angle	Study	Response Time (s)	Accuracy (%)	Confidence (%)
A	0.022	1.26	1	-	-	-
B	0.073	4.17	1	6.82	86.81	78.30
C	0.208	11.75	1	4.56	93.06	87.50
D	0.037	2.12	3	3.38	64.32	48.96
E	0.111	6.33	3	2.39	84.11	69.53
F	0.111	6.33	3	2.58	75.78	69.14

10.2.1 Modality

The effect of modality was one of the main considerations of this research. In particular there was a desire to know how the unimodal conditions compared to the levels of performance observed when modalities were combined (see Figure 10-2 for combinations explored). A further aspect of this enquiry was the appropriateness of sound, and findings in this respect are discussed later (see 10.2.4 and 10.4).

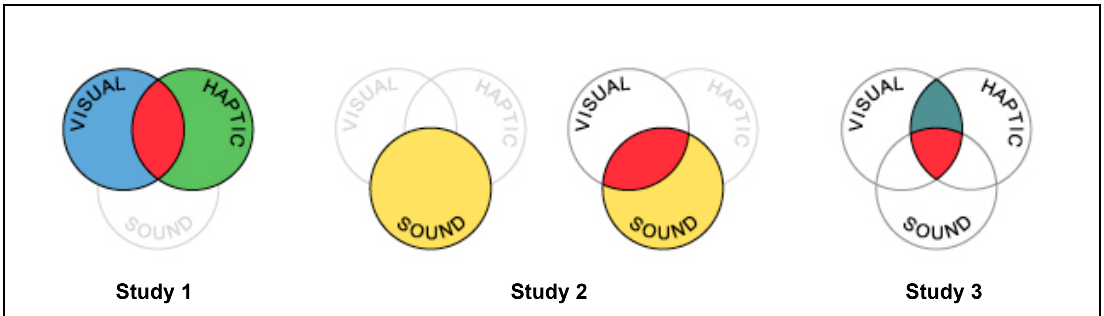


Figure 10-2: Modality combinations investigated through this research

In Study 1 it was found that unimodal haptic performance was much slower, less accurate, and less confident than visual-haptic performance. Visual performance was found to be not significantly different to visual-haptic performance in terms of response time and accuracy. Given this similarity between the visual and visual-haptic condition, it is thought likely that visual perception is dominant. However, the combination of visual and haptic senses enabled a significantly higher confidence level amongst participants than either of the senses produced independently. The fact that in combination haptic feedback did not impact on response time or accuracy, and helped to improve confidence would suggest that bimodal feedback would be an optimal choice (when considering naturalistic modes). This finding is supportive of interfaces, such as that developed in the SATIN project, which seek to integrate haptic as well as visual feedback for shape exploration. This meets with end-user needs (SATIN Consortium 2007), and has a measurable effect on user confidence. It should be noted that this confidence is not unfounded since there was a positive correlation between confidence level and accuracy.

Prior to investigating multimodal interaction in Study 3, a number of experiments were undertaken to investigate the appropriateness of sound. The findings for these are discussed in more detail later (see 10.2.4 and 10.4). However, it is important to note here that unimodal auditory performance was similar to bimodal visual-haptic performance (see Table 10-2). In the case of response times, this may not give an accurate reflection of the time needed to make a judgement. This is because participants were constrained and could not freely interact with the sound but had it played to them (once only). This constraint also reflects well on the levels of accuracy and confidence achieved given that there was only one comparison. Although haptic and sound explorations are both temporal in nature, the level of accuracy was higher for sound exploration. This may be a result of sound enabling a higher level of acuity than touch.

The effect of combining sound (curve shape and curvature sonifications) with the visual-haptic modality was investigated in Study 3. This found that participants' performance with the curve shape sonification was slower, but as accurate as and more confident than with visual-haptic alone. These findings contradict the similarities found between unimodal audition (Study 2 Experiment 2) and visual-haptic (Study 1) performances. It would be reasonable to assume that when combining these modalities the subsequent multimodal performance remains similar rather than diminishing. The fact that this has happened, points to a conflict between the haptic and auditory perceptions of curve shape (remembering that both of these explore the curve in a temporal fashion). This sense of conflict was reported by participants who undertook this study. Surprisingly, whilst there was some evident confusion caused by the addition of sound, it also enhanced confidence. The enhancement of confidence may be considered beneficial; however where this gives a false sense of accuracy this may not be a desirable outcome (see Chapter 9 Section 9.4.2 for a more detailed discussion of this).

Participants' performance with the curvature sonification was faster, more accurate and more confident than the visual-haptic performance. The scale of this was quite marked, with response time being approximately half, and accuracy and confidence increasing by at least 20%. This difference is all the more remarkable since that in Study 2 (Experiment 2) the unimodal sonifications ('Curve Shape' and 'Curvature') were seen to have a similar level of performance. This was also the case in Experiment 3 of the same study which had a combination of visual and auditory feedback. It is tempting to conclude that the conflict between auditory and haptic senses accounts for the differential seen. However, this seems too large to be accounted for by a single factor. So it may also be due to an improvement in performance of the curvature sonification in a multimodal context. That is, the added visual perception may have acted in a confirmatory way that boosted performance. This is speculative, and so further research would be required in order to understand these discrepancies.

10.2.2 Stimulus Properties

During the course of the research a number of stimulus properties have been investigated. In Study 1 it was found that participants were faster, more accurate, and more confident when judging differences between high compared to low curvatures. However, it was difficult to disambiguate other stimulus properties from these findings and so Study 3 was designed to investigate the relative contributions of curvature, gradient, and width. It was found that curvature had no effect on performance, and that the active dimension of difference was gradient. Participants were able to judge differences between high gradients more quickly, with greater accuracy and confidence than low gradients. This matches with the results found in Study 1 between high and low curvatures (as these corresponded to high and low gradients).

There was also a small effect of width on accuracy, although this was only in interaction with larger magnitude differences. Interestingly there was no effect of width on response time, which indicates that exploration is not linear. That is to say, participants do not explore the stimuli in a continuous fashion across the whole of the surface. Instead they interact with distinct parts of the stimuli which accounts for the similar response times for varying widths (for further discussion of this see 10.3).

For each study it was observed that performance improved or worsened relative to the gradient (see Table 10-3). However, when the data from both studies are viewed together they do not follow this relationship in a predictable pattern. This may indicate, as with perception, that inter- and intra-object performances are different. However, for response time, the difference that exists may be due to the nature of the task. The requirement for Study 1 was to examine two objects whereas for Study 3 there was only one object to examine. This may have been one of the factors that resulted in the halving of response time in Study 3. For this reason, or through intrinsic differences, it was not possible to produce a model that fits both types of stimuli.

10.2.3 Magnitude Difference

In psychophysical testing the accuracy of detecting a difference between the standard and comparison stimuli is used to identify the level of acuity (Gescheider 1985). Since the magnitude between the standard and comparison becomes greater, the accuracy with which a difference can be detected increases. In effect the perception of difference becomes easier as the magnitude of the comparison stimuli increases. Because of this characteristic of perception the magnitude of differences was thought to offer a good way to quantify task difficulty. In this way it is possible to quantify and understand performance relative to the demands of the judgement being made. Therefore, as the difficulty of the task increases we might expect to find a diminishment in performance. However, whilst this is true of accuracy in relation to discriminating curvature differences, it was unknown as to what effect there would be on measures such as response time and confidence.

It was found that participants were slower, less accurate, and less confident when judging small differences than when discriminating between medium or large magnitude differences. There was also found to be a correlation between magnitude difference and the various performance measures. This effect is not surprising, as this interaction is well known for accuracy and magnitude difference as it forms the basis of psychophysical calculation of thresholds. In identifying a similar relationship between other performance measures, this presented an opportunity to predict performance based on JND (see 10.2.5 for discussion concerning this).

The categorisation of magnitude differences into large, medium, and small did not provide a meaningful quantification and so in Study 3 differences were based on a factor of 3.25%. The results were the same, showing that participants performed less well when judging small differences (3.25%) compared to larger magnitude differences (6.5%, 13%, and 26%). However, there was a greater sense of what these results meant in terms of the decreasing difficulty of the judgement and relative improvement in performance. They were also more easily compared to the level of JND, which again enabled a sense of the relative difficulty to be known.

These findings therefore enable a sense of performance in relation to task difficulty to be formed. By using magnitude difference as a gauge it allows for a well-defined level of task difficulty to be quantified. Whilst such a mechanism may not be readily identifiable (or is some cases possible) outside the current domain, the value of such quantification has been demonstrated and may prompt exploration of something similar for other types of task.

10.2.4 Sound Type and Sonification

As part of assessing the appropriateness of sound, Study 2 compared a number of different types of sound. These were a simple sine wave (‘Sine’), a harmonic sound (‘Cello’), and a complex sound (‘Physical’). Additionally, ‘Cello’ and ‘Physical’ formed two new sounds with the addition of a kinetic module which altered the sound dependent upon the pressure applied to the sensor (Study 2 Experiment 3). It was found that there was little difference between the ‘Sine’ and ‘Cello’ sounds for most performance measures. There were a couple of exceptions; participants were more confident with ‘Cello’ in judging magnitude differences, and were faster in matching auditory feedback to curve shape in the multiple choice task (Experiment 3). These differences are quite minor and suggest that either sine wave or harmonic sounds are well suited to conveying curve shape or curvature information. However, the physical sound showed some serious limitations in places. In judging orientation participants’ accuracy dropped to 51% when using this sound (Experiment 1). When judging magnitude differences participants’ response times were longer, less accurate, and less confident than either ‘Cello’ or ‘Sine’ (Experiment 2). It is thought that this more complex type of sound has too much ‘noise’ in order to convey the necessary level of information (see 10.4.1 for detailed discussion of this).

During the course of this research two sonification strategies were adopted; ‘Curve Shape’ and ‘Curvature’. These varied in the way that sound frequency was mapped to curve properties. For ‘Curve Shape’, frequency was mapped to shape through position, and for ‘Curvature’ it was mapped to curvature value. Effectively this meant that two types of information were conveyed to participants; the first concerned the shape of the curve, the second concerned the mathematically defined curvature of the curve.

In Study 2 participants showed similar response times, accuracy, and confidence in assessing differences in curve shape through both these methods. This was unexpected in the case of Experiment 3 where it had been considered that participants would struggle in understanding the curvature sonification. That they did not is an indication of how well sound can convey these types of property even where this is quite abstract. However, it was also during the course of this experiment that participants were observed in developing behaviours that helped to simplify the auditory information they were receiving. Rather than a continuous stream of information they seemed to prefer discrete auditory bursts. So rather than running along the whole length of a curve which would produce a continuous stream of sound, participants would instead pick out points along it in order to produce a number of discrete tones which could be easily compared.

The value of this type of tonal comparison was fully realised in Study 3. Here the ‘Curvature’ sonification drastically out-performed both ‘No Sound’ and ‘Curve Shape’ conditions in terms of response time, accuracy, and confidence. The similarity of ‘No Sound’ and ‘Curve Shape’ was their characteristically continuous feedback (either haptic or auditory) and their difference from ‘Curvature’ was its discontinuous character (ability to produce two tones that could be compared). It is also telling that some participants attempted to produce discontinuous feedback in the ‘Curve Shape’ condition; that is they selected points to produce tones which they could then compare (whether this strategy was successful is unknown). On this basis it is clear that the ‘Curvature’ sonification was the more appropriate strategy to adopt in conveying this type of information. However, it is probably more accurate to identify simplification and discontinuous presentation as the characteristics that aided success, rather than attribute this to a particular sonification mapping i.e. ‘Curvature’. Additionally it is important to note that the success of the sonification method was also dependent upon the other modalities with which it was combined. This was perhaps more obvious in the case of the ‘Curve Shape’ sonification which seemed to clash with the haptic modality (Study 3).

The use of sound has been shown to be highly effective in conveying curve shape and curvature information. However, there are issues in its use and a methodical approach needs to be taken in its applications. Not least of all because, unlike other modalities, sound demands our attention; we cannot close our ears, or move them away. Where this provides optimal information other senses may be ignored and all is well. Where this is sub-optimal, confusion or conflict may arise with other senses causing a diminishment in perception and performance. This sense of conflict was reported by participants in Study 3. Therefore, whilst sound can improve performance (and acuity, see 10.1.2), this is only the case through careful consideration of the correct mapping; that is where human perceptual ability (e.g. distinguishing pitch) is aligned with an appropriate property of the object to be judged (e.g. curvature). Other considerations may include such things as polarity and scaling factors (Walker and Nees 2010). The choices involved in sonification of data are not trivial and for optimal results should be undertaken with regard to users' needs; perceptual, cognitive, and practical (see 10.4).

10.2.5 Prediction of Performance

In Study 1 it was found that there was a correlation between magnitude difference and performance measures (response time, accuracy, and confidence). However, a single set of metrics were not considered appropriate because performance levels were better for high curvature in relation to low curvature. This led to the proposition that metrics could be defined for various magnitude differences but with separate scales for high and low curvatures. In this way comparative data was provided for benchmarking purposes (see Chapter 9 Section 9.2.3.1).

Having identified that there was a correlation between magnitude difference and performance measures, it was proposed in Study 3 to take the analysis further by using multiple regression. In this way the contributions of a number of predictors could be assessed. It was found that magnitude difference, stimulus gradient, and stimulus width were effective predictors of performance. There was a good fit for response time and accuracy (r^2=0.80 and r^2=0.79), but only just above half of the variance could be accounted for in the case of confidence (r^2=0.60). This latter result may be due to poor linear fit in one of the underlying variables[33], and indicates the difficulty of applying multiple regressions in such cases. Similar levels of fit were found for the Study 1 data with a good fit for response time and confidence (r^2=0.90 and r^2=0.76), but with a poorer fit in this instance for accuracy (r^2=0.55). Whilst there are some reservations about the appropriateness of having applied multiple regression analysis in some cases (because of linearity considerations), all models had significant ANOVA results which indicated that they were more accurate than if the sample mean had been used.

The work here indicates that performance may have a predictable relationship but more work needs to be done in order to identify other factors which may account for at least 20% of the remaining variance[34]. It is also acknowledged that the models would benefit from the provision of more data. This would particularly help in decreasing the confidence intervals which were found to be quite large. The research here is an initial

[33] The confidence data for stimulus 'D' whilst appearing linear had a poor overall fit (r^2=0.003). This was due to low confidence for all magnitude differences at this gradient level.

[34] This is for those variables that were considered to have a good level of linearity.

attempt to understand the factors which may influence performance and assess if this relationship is predictable. By no means should this be considered definitive, but it does present initial models to be developed further. However, this is not to dismiss the multiple regression analysis, but to point out its limitations. The data may still be considered useful in providing a point of comparison. The models were found to be closer to the population mean than the sample means and may be used in preference to them in generating performance metrics.

10.3 Interaction

10.3.1 Styles

With the exception of those experiments specifically aimed at investigating interaction (Davidson 1972; Lederman and Klatzky 1987), most haptic experiments heavily constrain the interaction of participants due to the research aims or equipment. The limitations imposed within the studies reported within this thesis have been minimal (requirement to touch the stimuli with the index finger of the dominant hand), and so this allowed participants a large degree of freedom to interact as they wished. Therefore, whilst the main focus of research within this thesis has been concerned with perception and performance, it has been possible to explore interaction and the effect of sound on it.

Experiment 3 of Study 2 presented the first opportunity to observe the ways in which participants used sound feedback to explore visual curves. It was found that two broad types of interaction were used; sweep and point. These could be further classified into three predominant styles; ‘total’ which applied to sweep only and referred to a total traverse of the curve, ‘tracking’ where a distinct section was explored, and ‘comparison’ where two distinct sections were compared (see Chapter 6 Figure 6-17 for an illustration of these). In Study 3 the same variety of types and styles was not observed, but was confined to ‘sweep total’ and ‘point comparison’. These were not adopted universally but had distinct associations with particular sound or sonification conditions. Therefore it would seem that task conditions and the nature of the sonification both contribute to the interaction type and style adopted by the participant.

The ‘sweep’ interaction type observed in Studies 2 and 3 is very similar to ‘contour following’ (Lederman and Klatzky 1987) and ‘Top Sweep’ (Davidson 1972) identified in the haptic literature. However, a further type (point) and number of styles (total, tracking, and comparison) were identified as a result of the studies conducted within this thesis.

By identifying these interaction types and styles it was possible to monitor participant behaviour and develop an understanding about how interaction is affected by the presence of sound (see 10.3.2 for a discussion this).

10.3.2 The Effect of Sound

In Study 3 the effect of sound on interaction was explored. It was found that particular types of interaction were associated with different conditions. Without sound participants were observed to use the ‘sweep’ interaction type exclusively. Since this has been previously identified within the haptic literature, it may be assumed that this is a preferred or naturalistic interaction for this sort of task (Davidson 1972; Lederman and Klatzky 1987). It may therefore be inferred that deviations away from this are due to the influence of other factors.

The introduction of sound feedback was observed to induce other types of interaction. When exploring with the ‘Curvature’ sonification, participants predominantly used ‘point’ interaction (71%). This enabled them to produce a different tone either side of the mid-point in order to determine which side was ‘more curved’. Even so, some participants still adopted the ‘sweep’ style which indicates that more extensive haptic feedback was desired. This was corroborated in that, whilst 75% said they used sound only for making judgements, 17% said they used sound and touch. Therefore whilst the availability of sound had required minimal interaction to be effective, a sizeable proportion desired extensive haptic feedback as experienced through a sweep along the whole curve shape.

When participants used the ‘Curve Shape’ sonification they were observed to adopt two types of interaction; ‘point’ and ‘sweep’. About half adopted the sweep type exclusively, whilst 43% alternated between the point and sweep interactions. What is of interest is not that the sweep style was adopted, since this is an obvious choice (given the temporal/spatial nature of the feedback), but that participants opted for other types of interaction. In this context the choice of the point style can be seen as a desire to simplify feedback; the interaction rather than naturalistic is adaptive. This drive to simplify was seen more extensively in Experiment 3 of Study 2. Here participants adopted a number of sub-styles in order to control and make sense of the sound feedback. The continuous stream of rising and falling tones was largely abandoned in favour of discrete chunks of feedback. The effect of sound is therefore to alter naturalistic patterns of interaction in favour of interactions that seek to control auditory presentation.

10.3.3 Haptic Duality

What is evident from the observed interactions is the duality of touch; it was used for both perception, and for action (Wolfe, Kluender et al. 2006). This is clear through the identification of different types of interaction behaviour. In Study 3 it was seen that 'sweep' was used exclusively when there was no sound and can be clearly identified as an action to perceive as there is no other outcome available. However, it is not always so certain as to what is being observed, and it is difficult to know if an interaction is for perception or for action; that is, was touching of the stimulus for the purpose of perceiving the nature of the curve or is it to control audible feedback?

In some instances action and perception are one and the same. For example, whilst using the 'Curve Shape' sonification, participants perceive and act in spatial and temporal unison; the control and perception overlap as the participant sweeps their finger along the curve of the stimulus to perceive its form and the sound feedback is generated. This type of duality might be considered beneficial, however there was indication that this led to confusion as there was conflict between what was felt and what was heard. However the adoption of different interaction styles for this particular sonification may indicate an effort to separate out these dual functions by, for example, adopting the 'point' style.

On other occasions action and perception are already distinct. For example, whilst using the 'Curvature' sonification, participants touch a point on the stimulus not to perceive but to activate sound feedback. In this instance any perceptual feedback is purely incidental. In this way the desire of the participant to control is not coupled to intrusive haptic perception. It may be considered that the participant has successfully decoupled sensation from this transaction, however it is interesting to note that confidence is 10% higher for visual-haptic-sound than for sound only or sound-visual feedback (see Table 10-2).

Within this research distinctions between touch for action and touch for perception are fairly clear, but they illustrate a wider issue of how we observe the difference and more importantly design the difference for more complex interactions. It also cautions that careful consideration should be given as to how action and perception are assigned within haptic interfaces in order to avoid unnecessary sensory conflicts.

10.4 Sound and Sonification

The effect of sound has been discussed in relation to its effect on perception and performance (see 10.1.2 and 10.2.4). Here there is a focus on the human factors issues identified in the literature with regard to sonification of data (Walker and Nees 2010), and in particular, the appropriateness of sound for conveying curve shape and curvature information.

10.4.1 Perceptual Capabilities

Perceptual capabilities were directly explored in the first and second experiments conducted as part of Study 2 (see Chapter 5). These experiments investigated participants' ability to perceive the orientation and magnitude of a curve through sound. It was found that participants were able to respond quickly, confidently and with a high level of accuracy (see 10.2.4 for discussion concerning performance). However, consideration here will be given to the areas in which participants experienced error or performed less accurately, since this should provide insights into the perceptual limitations of sonification.

The type of sound used was found to have a significant effect on the type of error. The 'Physical' sound, which was the least accurate, had a significant level of shift error (44%). This meant that participants were unable to accurately distinguish between the orientations of neighbouring curve segments, and so made an inaccurate identification. Whilst these types of errors were found with the sine wave ('Sine') and harmonic sound ('Cello') they were not at significant levels. It is therefore thought that it must be something in the characteristics of the 'Physical' sound that make it more difficult to perceive. If we look again at the sonograms for the different types of sound (see Fig 5-4) it is evident that the 'Physical' sound is more complex, and it may be this that accounts for the perceptual inaccuracy experienced when using this sound to convey curve orientation. It is difficult to be definitive about this as only one complex sound was tested. However, it is clear that both the sine wave and harmonic sound, both of which had simple wave patterns, showed higher levels of accuracy. Therefore it may be reasonable to assert that simple sounds are more easily perceived and should therefore be used in preference to complex sounds when conveying curve information.

A further characteristic that has been found to be important in the perception of curve shape information is the rate of tonal change. This was manifested in two ways; the rate of tonal change needed in order to convey the orientation of a curve, and the rate of tonal change difference needed to distinguish curves of different magnitudes. The first of these was observed in Experiment 1 of Study 2 were it was found that orientation ‘A’ was more accurately identified than either ‘B’ or ‘E’. This difference can be understood by looking at the difference between orientations ‘B’ and ‘E’ and their counter-parts, ‘C’ and ‘F’, which were not significantly less accurate than ‘A’. It was demonstrated (see Figure 5-19) that the rate at which the tone of the sound changes, as it traverses the curve, is less over the first half of orientation ‘B’ (or ‘E’) as opposed to the first half of ‘C’ (or ‘F’). It is therefore suggested that in order to convey orientation information there needs to be a pronounced rate of tonal change at the beginning of the sonified segment in order for it to be perceived accurately.

The second type of tonal rate change dependence was seen in Experiment 2 of Study 2 and was also found in Study 3. In both these cases it was found that, when using the ‘Curve Shape’ sonification, participants were faster, more accurate, and more confident in judging medium to large magnitude differences than small. It is thought that this occurs because, with small magnitude differences, the standard and comparison stimuli have very similar rates of tonal change and that the difference between these is therefore difficult to perceive (see Chapter 7 Figure 7-37 for an illustration of this). This type of perceptual difficulty may have prompted adoption of the ‘point’ style observed in Experiment 3 of Study 2. In essence the use of a ‘point tracking’ interaction style (see Chapter 6 Figure 6-17d) can be seen as an adaptive means of customising the rate of tonal change, in order to make it more pronounced than the default auditory presentation experienced when sweeping the curve in a continual motion. In this way the participant ensures an auditory presentation that they are easily able to perceive. If this is the case, then it is important to ensure that, when presenting continuous sounds such as the ‘Curve Shape’ sonification, an appropriate rate of tonal change is used, and that were this is not certain that interaction can be adapted by the user to counteract any perceptual deficiencies experienced.

10.4.2 Musical and Cognitive Abilities

The experimental studies presented within this thesis had not been designed to explore musical background as a factor, but had rather controlled for this by ensuring a range of backgrounds. However, it became apparent whilst undertaking Study 2 that there were differences between the ‘General’ participants, who had no musical training, and the other categories of participant, who had varying levels of musical training. Therefore it was decided to undertake some post hoc analysis based on participant background. It was found that across all experiments there was no effect of background on response time or confidence. However, there was an effect of background in relation to accuracy. In Experiments 1 and 3 it was found that those with a ‘General’ background (no musical training) were significantly less accurate than those with a ‘Music’ (advanced musical training) or ‘Physics/Maths’ (some musical training). It is tempting to conclude, in line with some of the literature (Neuhoff, Knight et al. 2002), that those with musical training are more accurate than those without. However, Experiment 2 showed that only those with an advanced level of musical training (‘Music’) were significantly more accurate than those without (‘General’); although the differential was only around 10%. Given this low level of difference, and the fact that there was no or limited differences with those with some musical training, it would be more appropriate to conclude that musical training makes little practical difference, which is in keeping with the conclusions of Walker (2010).

These differences were found when using the ‘Curve Shape’ sonification, and it should be noted that there were no significant differences found when using the ‘Curvature’ sonification (Experiments 2 and 3). In relation to Experiment 2 this is of little surprise since this effectively meant making a judgement between two tones; a relatively simple task even for those with no musical training. However, for Experiment 3 the ‘Curvature’ sonification had the same level of complexity as the ‘Curve Shape’ sonification, and so to this extent it was surprising that no difference was found between the different backgrounds. However, whilst not significant, it is interesting to note that those with a ‘Physics/Maths’ background scored more highly (86%) than those with a ‘Music’ background (72%), who on this occasion performed more similarly to the other backgrounds. It is thought that those with a ‘Maths/Physics’ background may have been more easily able to conceptualise curvature and so maintained a similar level of performance. This type of cognitive advantage has been observed in other studies, for example, spatial reasoning ability can predict performance with auditory graphs (Walker and Mauney 2004).

These observations suggest that auditory feedback should be considered in two stages; first is the perception of sound, and second is the conceptualisation of meaning. It is likely that most people will perform highly at this first stage as there seems little need for any particular musical training. However, more specialist knowledge may be required to maintain performance through the second stage of understanding. This pattern of perception and cognition may be evidenced in the difference between response time and exploration duration found in Study 3. The exploration duration may encompass the perception stage, whilst the difference between this and the response is the cognitive stage. So that once perceived, additional time is required in order to conceptualise the meaning of this perception, e.g. high tone means more curved. More research would be needed into this, but potentially this may provide another way to monitor the success of different sonification strategies; that is, to find the ones with minimal lapsed time between exploration duration and response time. In this way it would be possible to identify more intuitive sonifications.

Following observations made in Study 1 and the SATIN evaluations there was a concern that workload might be high and that the addition of auditory feedback would create still further cognitive demands. This was monitored in Study 2 and Study 3 through the use of the NASA-TLX Score (Hart and Staveland 1988). It was found that there was little difference in workload score between any of the experiments in Study 2, or those and the Study 3 scores (see Table 10-4). These were in a range of 41 to 57, and indicate that participants were operating within the centre of scale. However, the scores for Experiment 3 which had a more complex task show a rise over the other scores reported and may indicate that there are additional demands associated with perceiving the fluctuations of variable as opposed to continuous curves.

Table 10-4: Comparison of NASA-TLX Score and sub-factors across all studies

Sub-factors & Overall Score	Study 2					Study 3		
	Exp 1	Exp 2		Exp 3				
	CS	CS	CURV	CS	CURV	NS	CS	CURV
Mental Demand	149	139	112	288	286	255	219	127
Physical Demand	30	16	15	19	13	12	12	7
Temporal Demand	51	43	36	61	71	49	55	40
Performance	255	327	339	224	189	171	203	344
Effort	122	107	97	215	217	180	170	81
Frustration	44	45	42	55	69	70	64	13
NASA-TLX Score	44	47	45	57	56	47	48	41

Whilst the overall NASA-TLX scores are similar, there are differences apparent within the sub-factor ratings. One factor of particular interest, as it is indicative of cognitive load, is Mental Demand. It can be seen that this is highest for Experiment 3 and gives further indication of the higher demands of this type of task as opposed to the sonification of simple curves. It is also interesting to note that when using the ‘Curve Shape’ sonification there is increased demand when combined with the haptic modality (as indicated by the increase in Study 3[35] over that of Experiment 2 in Study 2). This may be further indication of the conflict between particular sonifications and the haptic modality (as discussed in 10.2.4).

10.4.3 Training

The issue of whether or not practice improves performance was considered through the experiments conducted in Study 2. It was found that there was some improvement in response time (-1s), accuracy (5%), and confidence (6%). These are smaller factors of change than indicated by Walker and Nees (2005), where a 50% increase in performance was reported for practice with feedback. This suggests that greater improvement may have been found had the participants been given feedback during testing.

The value of training, however, may be in more than a simple improvement in performance. It was found in Experiment 1 of Study 2 that there was a reduction of 25% in the mental demand experienced by participants after practice. Given that mental demand has been observed to increase for multimodal interaction, it may be useful to ensure that users are allowed practice as a possible counter to this type of increase.

10.5 The value of this research for SATIN and future projects

The main driver for this research was the author's involvement in the SATIN project (as discussed in 1.1). It is therefore appropriate to review the outcomes of this research against that context. There were a number of concerns following the initial evaluations of the SATIN prototype and these gave focus to the author's research. This gave rise to such questions as; does the integration of modalities have an impact on perception and performance? Is sound a suitable medium through which to convey curve shape information? What constitutes a good level of performance? Does multimodality induce high levels of workload?

---

[35] It should be noted that there was no significant difference found between the 'No Sound' and 'Curve Shape' sub-factors in Study 3

In relation to these SATIN concerns, the research has been able to provide valuable insights. It was shown that the integration of modalities had no significant impact upon perception or performance (see 3.4.1 and 3.4.2). Sound was found to be a suitable medium through which to convey curve shape information, although this was highly dependent upon the type of sonification method used (see 6.4.1.1 and 7.4.2). Performances across a range of difference judgments were found to be quick, accurate, and confident (see 3.4.2 and 7.4.2.) Finally, in relation to workload it was seen that, across all modalities, participants operated within the mid-range of their capabilities (see 6.4.1.2 and 7.4.4). The implications of these findings for the SATIN prototype were an indication that performance was not optimal. However, this is not surprising as the evaluations were undertaken at an early stage of development. Even so, what is important here is that at the time of the evaluation this judgement could not have been made, it is only in the light of the research undertaken within this thesis that it is possible to suggest the limitations of such a prototype.

Whilst these particular findings were of direct use within a project such as SATIN, the research outcomes as a whole are applicable more widely. The insights and knowledge gained were developed into a framework that may be used to evaluate a wide range of multimodal interfaces for shape exploration (see Chapter 9). As with the SATIN project it is of value to know, for any given interface, whether the performance achieved by users is of an acceptable level. The difficulty arises in knowing what might be an acceptable level and thus the quantification of usability metrics such as efficiency (response time) and effectiveness (accuracy). It might be expected that such quantification would be available within the relevant haptic and auditory literature. However this type of research is not evident and is perhaps, given the relative newness of the research domains, secondary to research of a more fundamental nature or proof of concept work. The research conducted here therefore contributes to an understanding, and provides a means to measure, human performance and perception in relation to judging curve shape differences. This may be achieved either by using the data sets produced through this research as a means to benchmark interfaces (see 9.3.2.1) or through the use of perception (JNDs) and performance (response time, accuracy, and confidence) models to generate a range of metrics (see 9.3.2.2). These would enable a range of interfaces to be evaluated against objective measures and for an optimum configuration to be derived.

In addition to the provision of performance and perceptual metrics, the framework provides for a ‘discount’ method of evaluation (see 9.2.2). The psychophysical method used within this research was very time consuming; for example Study 1 took 72 hours to find thresholds for two curvatures in three conditions giving a total of 6 outcomes (see 3.2.3). This level of resource is unlikely to be adopted in anything other than experimental circumstances; however most interfaces are evaluated in short formative evaluations (like those conducted on the SATIN prototype, see 1.1.5.1 and 1.1.5.2). Therefore, the discount method, which requires about 12 participants to undertake 15 minutes of trials each, is more easily integrated. This is an important deliverable of this research, as it allows for the integration of a psychophysical type method within a formative evaluation setting.

There are some limitations within this work (see 11.3), and the tools are only applicable to haptic or visual-haptic interfaces. However, the framework (methods, guidelines, and predictive tools) does provide a valuable contribution to the evaluation of such interfaces for shape exploration.

Chapter 11: Conclusions and Further Work

11.1 Conclusions

The aim of this thesis was to explore the value of haptic and sound feedback in the perception of curve shape, and provide a framework for its evaluation.

In order to meet this aim the following objectives were identified, to:

1. Identify the effect of combining haptic, visual, and sound feedback on perception, performance, and interaction.
2. Inform theory and develop a predictive model based on measurement of just noticeable differences.
3. Assess the appropriateness of sound for conveying curve shape and curvature
4. Develop appropriate performance metrics for evaluation of curve shape using unimodal and multimodal feedback.
5. Provide guidance for evaluation of interfaces for shape exploration.

The main conclusions that can be drawn in relation to these objectives will now be discussed.

11.1.1 Identify the effect of combining haptic, visual, and sound feedback on perception, performance, and interaction.

The effect of combining modalities on perception was measured by comparing acuity (expressed as percentage JND or Weber Fraction). It was found that haptic, visual, and visual-haptic modalities had the same level of acuity (23%). However it was later shown that, with appropriate sonification ('Curvature'), sound could improve differentiation of curve shape substantially (2.5%). It is therefore possible to conclude that the combining of haptic and visual modalities does not impact acuity detrimentally, and that with the addition of sound acuity can be increased.

It was found that visual-haptic performance was similar to that achieved by using vision only, and was better than haptic performance. It was therefore thought that the combination of visual and haptic senses was dominated by vision. However, it was also found that visual-haptic judgements were more confident than either haptic or visual modalities in isolation. It can therefore be concluded that, the combination of visual and haptic senses may not improve performance (and importantly does not decrease it), but has the benefit of increased confidence and fulfils the user need of naturalistic interaction. It can be further concluded that sound feedback can substantially improve visual-haptic performance, although this is highly dependent upon the type of sonification method used.

It has been shown that interaction is adapted in order to control sound feedback. There is indication that continuous sound is difficult to perceive or conflicts with perceptions from the haptic sense. This led to user interactions which simplified feedback, so that auditory presentation was made in discrete chucks that could be more easily perceived. It is therefore possible to conclude that discrete sound feedback is more effective than continuous for this type of comparison task.

11.1.2 Inform the Theory and Develop a Predictive Model based on JNDs

It has been demonstrated within the haptic literature that gradient (Gordon and Morison 1982) or attitude difference (Pont, Kappers et al. 1997) is the effective stimulus for curve detection. Implicit within this is that a similar effect might be found for perception of curve difference. Through analysis of Study 1 data and that of the haptic literature, it was found that there is a very strong linear relationship between stimulus gradient and difference threshold gradient (r^2=0.99). Study 3 confirmed this relationship, and so evidenced and made explicit the link between gradient and difference threshold. It can therefore be concluded that gradient is the effective stimulus for perception of curve difference. In identifying gradient as the effective stimulus it was possible to develop a predictive model for generating JNDs for each modality (see Figure 4-6). Prior to this work no such model was available to predict the change necessary in a stimulus for a difference to be perceived. This will be useful for the specification and evaluation of multimodal interfaces for shape exploration.

In addition, a serendipitous finding of this research was that the acuity of curvature difference perception for the visual-haptic modality was not the same for Studies 1 and 3. In the first study a Weber Fraction of 23% was found, and in the third study the acuity was double with a Weber Fraction of 12%. It was also found that conformance to Weber's Law varied, with non-conformance for Study 1 and conformance for Study 2. It can therefore be concluded that the perception of inter- and intra-object curvature differences is not the same, and that these differences should be considered when designing multimodal interfaces for curve shape exploration.

11.1.3 Assess the Appropriateness of Sound for Communicating Curve Shape and Curvature Information

Initial evaluations on the SATIN prototype suggested that the use of sound to convey curve shape or curvature information may be problematic for end-users (SATIN Consortium 2008). It was therefore decided to investigate the appropriateness of sound in conveying this type of information through a series of fundamental experiments (Study 2). The first two experiments examined sound as a means of conveying information about the orientation and magnitude of a curve, with a third involving an auditory to visual matching task. It was found that this performance was fast, accurate, and confident, and was not particularly dependent on the type of sound utilised. However, pure sine or harmonic sounds were found to perform better than more complex sounds. The workload demands were not found to be great for any of the tasks undertaken, with most participants operating within the mid-range of the NASA-TLX scale. It was found that performance was slightly increased after training, and that participants with musical and mathematical backgrounds may have had a slight cognitive advantage in undertaking the tasks. The two sonification methods used, ‘Curve Shape’ and ‘Curvature’, enabled similar levels of performance within these experiments. Given the ease with which participants undertook the tasks, the level of workload, and success using both sonification methods, it can be concluded that sound is appropriate for conveying curve shape and curvature information. However, there were found to be some perceptual barriers to the effectiveness of sound and a desire by participants to simplify the feedback from continuous to discontinuous sound. The contribution of this work has been in identifying the appropriateness of some sound types over others, the applicability of the sonification approaches, and the perceptual and human factors issues encountered when using sound to explore curves.

11.1.4 Develop Appropriate Performance Metrics for Evaluation of Curvature using Unimodal and Multimodal Feedback

This research has taken two approaches to the development of comparative data against which performance can be evaluated; metrics and predictive models. In Study 1 it was found that there was a correlation between performance and magnitude difference. It was also found that this had a linear relationship. However, a complicating factor was that performance for high and low curvatures was different, and that there were also differences across modalities. In addition, following Study 3, it was found that inter- and intra-object performance was different. This led to the production of performance metrics based on gradient (the effective stimulus for curvature perception), magnitude difference (performance increases with increasing difference), and modality (performance with haptic feedback only is poorer than with either visual or visual-haptic feedback).

The analysis undertaken in Study 3 identified an effect of stimulus dimensions on performance. There were found to be some main effects of stimulus gradient, but there were also some small interaction effects of stimulus width. In addition, there was a main effect of magnitude difference on performance. These findings suggested that there was not a single factor that affected performance. Therefore multiple regression analysis was chosen in order to determine if there was a relationship between these factors and performance. It was found that stimulus gradient and magnitude difference were strong predictors of response time, accuracy, and confidence. There was also found to be some effect of stimulus width but at most this accounted for around only 6% of the variance. Similar findings were achieved for the Study 1 data. It was necessary to derive separate models for each of these data sets as there were found to be differences between inter- and intra-object performances (see Figure 8-10).

The models produced were found to be a more accurate estimate of the population mean than the sample mean, and so were recommended for the specification of metrics. However, there were a number of limitations associated with these models. In some cases the r-squared was relatively low, and for all predicted values the 95% confidence intervals were large. This means that whilst the models are better than using the mean, there is a need to refine these with further data. In addition, the models have at best accounted for 80% of the variance and so further work is needed in order to establish other predictors. It is also noted that data were gathered from a limited range of curvatures and, in order to increase the applicability of the model, data from a wider range would need to be integrated. This final point is also a limitation of the metrics produced, and these too would benefit from further research data.

Thus, it is considered that whilst this work has limitations, its contribution is the demonstration that performance has a predictable relationship to magnitude difference and stimulus dimensions. It has provided performance metrics in the form of comparative data. It has also provided models for the prediction of performance. These models have been utilised in a tool for generating bespoke performance metrics for tasks involving the judgement of differences in curvature. Prior to this research, and the development of these models, it was not possible to estimate the level of performance that might be expected when judging if curves where different. The benefit of this research is an indication of the level of performance that may be expected at different degrees of task difficulty.

11.1.5 Provide a Framework for the Evaluation of Virtual Interfaces for Curve Shape Exploration

The literature review had identified that, whilst there was consensus in relation to what characterised a useable interface, there was no specific guidance as to the development or evaluation of interfaces for shape exploration. The knowledge and insights gained while conducting this research have led to the development of a framework for the evaluation of multimodal interfaces for curve shape exploration. This includes predictive tools for the generation of JNDs and performance metrics. In addition, insights have been developed into a set of guidelines for interface development.

11.2 Research Contribution

The research presented in this thesis contributes to knowledge in a number of ways, including:

- The work reported here has taken what was implicit within the haptic research literature, and extended it to become an explicit and evidenced theory of difference perception based on stimulus gradient. Through exploration of this mechanism it has been possible to produce a model of JND prediction. Prior to this work no such model was available to predict the change necessary in a stimulus for a difference to be perceived. This will be useful for the specification and evaluation of multimodal interfaces for shape exploration. In addition, a serendipitous finding of this research has been to show that perception of curve shape differences conforms to Weber's Law when judgments were made within an object (intra) as opposed to when judgements were made between objects (inter). It was also found that perception of intra-object differences were more acute than inter-object judgements.
- This research has demonstrated the appropriateness of sound feedback for the exploration of curve shape. It has shown that performance is high and not particularly dependent on the type of sound utilised, although pure sine or harmonic sounds perform better than more complex sounds. Further, it has identified a number of perceptual barriers to sound's effectiveness.

- This research identified stimulus gradient and magnitude difference as predictors of performance (response time, accuracy, and confidence) and developed a model that expresses this relationship. This provided a tool for generating performance metrics to help in the evaluation of interfaces for curve shape exploration. Prior to this research, and the development of these models, it was not possible to estimate the level of performance that might be expected when judging if curves where different. The benefit of this research is an indication of the level of performance that may be expected at different degrees of task difficulty.
- The knowledge and insights gained while conducting this research have led to the development of a framework for the evaluation of multimodal interfaces. This includes predictive tools for the generation of JNDs and performance metrics, and guidelines for interface development. In addition, a ‘discount’ method for the calculation of difference thresholds has been proposed which gives a 66% time saving over more traditional psychophysical methods; thus making it suitable for formative evaluations.

A diagram of how each of the studies undertaken in this thesis maps onto different types of contribution can be seen in Figure 11-1.

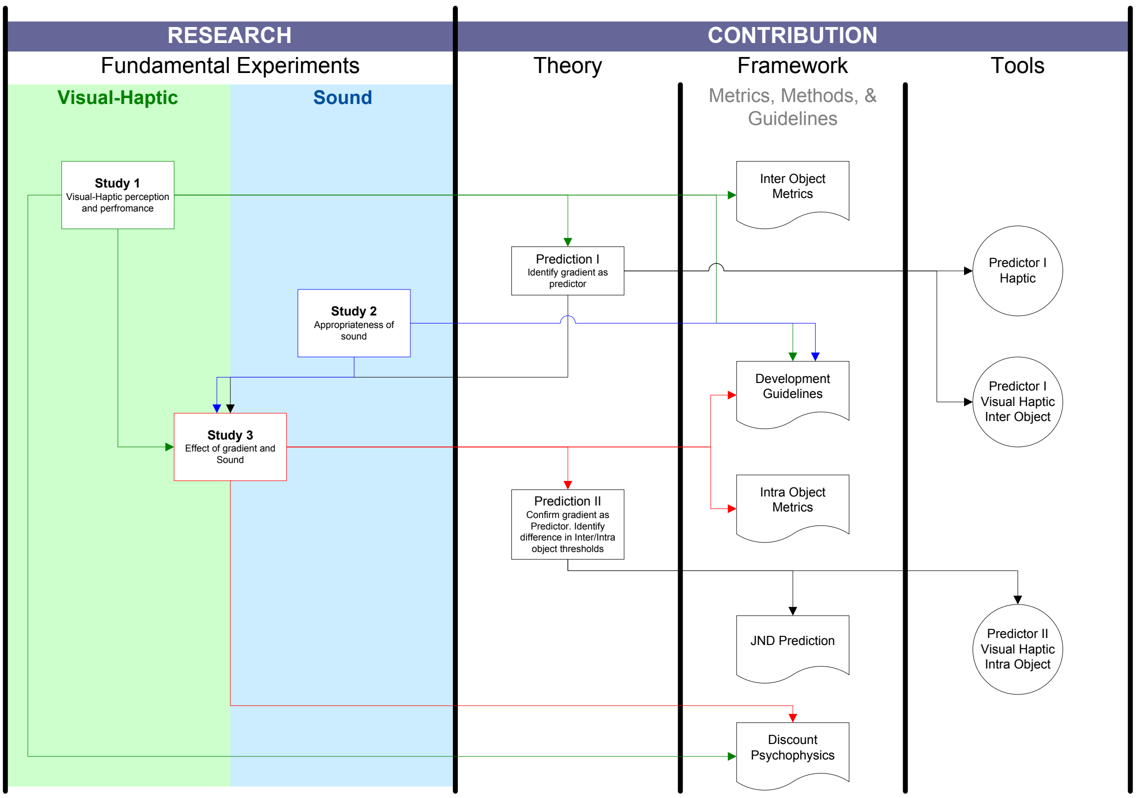


Figure 11-1: Research contribution made by each of the studies

A summary of how this research contributes in relation to the focal psychophysics literature can be found in Table 11-1.

Table 11-1: Contribution of the research in relation to the focal literature

Literature	Contribution of Research
Gordon and Morrison (1982) and Pont et al. (1997) have all shown that the effective stimulus for curvature detection is gradient (or attitude difference). They also indicate that an elevation of about 0.5^{o} is required in order for a curved surface to be detected from flat. A later finding of Wijntjes (2009) also suggests orientation (gradient) to be the effective stimulus for curvature detection for virtual as well as real curvatures.	Analysis of Study 1 data, and comparison with data derived from the literature, suggested that gradient was also the effective stimulus for judging curve shape differences. In Study 3 it was shown that stimuli with the same gradient also had the same difference threshold thus confirming gradient as the effective stimulus for difference perception. This finding extends on the literature (absolute threshold) by demonstrating that gradient was the active dimension in perceiving curve shape difference (difference threshold). In addition, to the author's knowledge, there has been no research that examines performance in relation to judging of curve shape differences. A contribution of the research within this thesis has been to report the level of response time, accuracy, and confidence in judging curve shape differences. Further, it has taken this data and produced a model by which performance may be predicted.
Louw (2000) extended the work above to show that absolute threshold could be calculated and was dependent upon the width of the stimulus. Absolute threshold was shown to be $width^{1.3}$.	Study 3 indicated that stimuli of the same width did not have the same threshold. Therefore a contribution of this research has been to show that Louw's formula, in the form published, was not applicable to the calculation of difference threshold. Further, the research has provided a model through which it is possible to calculate the difference threshold for any given curvature.
Kappers and Koenderink (1996) found that for haptic stimuli there was a non-conformance with Weber's Law in relation to difference threshold.	Study 1 confirmed that there was non-conformance to Weber's law for haptic, visual, and visual-haptic difference perception for inter-object comparisons. However, Study 3 suggested that there was conformance with Weber's Law for visual-haptic difference perception when intra-object comparisons were made. This finding suggests that there may be a difference between intra and inter-object perception of curve shape difference, and as such makes a contribution to thinking in this area.

11.3 Limitations and Future Work

The discussions within this thesis have identified limitations and made suggests for further work. This section will highlight a few of the more important ones in relation to key themes.

11.3.1 Perception

One of the important findings of this research has been to identify differences in inter- and intra-object perception of gradient differences. However, this conclusion was based on the results of one experiment. Within the haptic literature, to the author's knowledge, there is no indication of such a distinction. It would therefore be prudent to see if this result can be replicated. It would also provide for a more robust comparison of this effect if the same gradients were used for inter- and intra-object conditions.

It was found that the ‘Curvature’ sonification produced a difference threshold of 2.5%. It was suggested that this might be improved further up to a theoretical limit of 0.3%. There was also the suggestion that with the use of scaling this might be increased to enable still finer judgements to be made. Both of these are in need of empirical investigation. A further finding in relation to sound perception was that there was conflict between the curve shape sonification and the haptic sense. This may be because the sonification followed the curve, whereas haptically the gradient is sense. Further investigation could be made of this to see if the sonification of the gradient reduced this conflict with the haptic sense.

11.3.2 Performance

The research here has shown that it is possible to model the relationship between a number of predictors and performance measures. However, it was found that there was at best still about 20% of the variance unaccounted for. In addition, the confidence intervals for the predicted performance metrics were relatively wide (in some cases a range of around 40%). This means that whilst the principle has been established, and the predictions are nearer the population mean than the sample mean, there is scope for improvement in the modelling. Therefore, further research needs to focus upon other possible predictors and providing a broader range of cases in order to widen the application of the model and improve the level of confidence in the predictions.

The framework for evaluation is hitherto untested, and requires validation through application with multimodal interfaces such as the SATIN prototype. This would allow for the refinement of methods and also enable the collection of comparative data for different interfaces. In particular, it would be useful to systematically benchmark a range of haptic interfaces to establish their suitability for shape exploration. The SATIN project produced a useful categorisation of explorative interfaces and it would be useful to set against this indicative performance parameters.

11.3.3 Interaction

Through observation of haptic interaction in Study 1 it was possible to identify the effects of combining this modality with vision. It was found that, when used in conjunction with vision, the number of haptic comparisons of stimuli decreased. In Study 3 it was revealed that when judging the difference between two curves haptic exploration is focused upon the ends and middle of the stimulus. However what is lacking from these studies is an understanding of how visual interaction is affected in combination with haptic and auditory feedback. In particular it would be useful to know if visual inspection changes in the presence of sound. There is some anecdotal evidence to suggest that participants have a tendency to close their eyes or look away in an attempt to focus solely on the auditory display. There may be further ways in which the inclusion of auditory feedback impacts upon visual interaction. It would be helpful to firstly identify normal patterns of visual inspection, and then to see how these are complimented or disrupted by the addition of other modalities. In this way it would be possible to more fully understand multimodal exploration of curve shape, and from this ensure that interfaces are optimised for this type of interaction.

11.4 Final Comments

This thesis has three recurring key themes; perception, performance, and interaction. An understanding of the relationship between these has evolved as the objectives of this thesis have been explored. Perception is the bedrock from which all else follows; it helps to define task difficulty which has given insights into task performance and enabled a structure to evolve for the prediction of this. It also defines the nature of interaction as we seek to perceive, and act to understand. It is perhaps this duality of interaction that is of most interest, since this more than anything may influence or constrain the way in which interfaces are developed to be truly user-centred. In evaluating we can only focus upon the external; that is the extent of user performance or their tell-tale interactions. These are reflections of what we perceive and difficulty in these manifests the restrictions of our embodiment. However, technology presents opportunities for us to move beyond this and by harnessing this to the advantage of our senses we are able to be more than we were; interacting and out performing ourselves in ways that were not imaged. The work here provides a few small steps that can guide us down a path to better user-centred interactions.

References

Adelstein, B. D., Lee, T. G., & Ellis, S. R. (2003). Head Tracking Latency in Virtual Environments: Psychophysics and a Model. Human Factors and Ergonomics Society Annual Meeting Proceedings, 47(20), 2083-2087.

ANSI/NCITS 354-2001. (2001). Common Industry Format for Usability Test Reports. Retrieved. from.

Auvray, M. (2007). Learning to Perceive with a Visuo-Auditory Substitution System: Localisation and Object Recognition with `The vOICe'. Perception, 36, 416-430.

Barras, S., & Zehner, B. (2000). Responsive Sonification of Well-logs. Paper presented at the Proceedings of the International Conference on Auditory Display, Atlanta, USA.

Bearne, M., Jones, S., & Sapsford-Francis, J. (1994). Towards usability guidelines for multimedia systems. Paper presented at the Proceedings of the second ACM international conference on Multimedia.

Bierderman, I. (1987). Recognition-by-Components: A Theory of Human Image Understanding. Psychological Review, 94(2), 115-147.

Billger, M., Heldal, I., Stahre, B., & Renstrom, K. (2004). Perception of color and space in virtual reality: a comparison between a real room and virtual reality models. SPIE-5292, 90-98.

Blake, R., & Sekuler, R. (2006). Perception (Fifth ed.). New York: McGraw-Hill.

Blattner, M. M., Sumikawa, D. A., & Greenberg, R. M. (1989). Earcons and Icons: Their Structure and Common Design Principles. SIGCHI Bull., 21(1), 123-124.

Bordegoni, M., & Cugini, U. (2005, May 30 - June 1). Haptic System for Shape Generation. Paper presented at the 6th International Conference on Computer-Aided Industrial Design and Conceptual Design, Delft, The Netherlands.

Bordegoni, M., Espinach, F., & Tresserras, J. (2005). Design Virtual Products Using Hands. Paper presented at the Congreso International XVII Ingegraf - XV ADM.

Bordegoni, M., Ferrise, F., Covarrubias, M., & Antolini, M. (2010). Haptic and Sound Interface for Shape Rendering. Presence, 19(4), 341-363.

Bowman, D., Gabbard, J., L. , & Hix, D. (2002). A survey of usability evaluation in virtual environments: classification and comparison of methods. Presence: Teleoper. Virtual Environ., 11(4), 404-424.

Bowman, D. A., & Hodges, L. F. (1999). Formalizing the Design, Evaluation, and Application of Interaction Techniques for Immersive Virtual Environments. Journal of Visual Languages & Computing, 10, 37-53.

Brewster, S., Wright, P., & Edwards, A. D. N. (1993). An evaluation of earcons for use in auditory human-computer interfaces. Paper presented at the Proceedings of the INTERACT '93 and CHI '93 conference on Human factors in computing systems.

Carlile, S. (2002). An Overview of Auditory Dimensions that can be used to Represent Information. Paper presented at the Proceedings of the 12th International Conference on Auditory Display, Kyoto, Japan.

Challis, B., & Edwards, A. (2001). Design principles for tactile interaction. In Haptic Human-Computer Interaction (pp. 17-24).

Chen, J. Y. C., & Thropp, J. E. (2007). Review of Low Frame Rate Effects on Human Performance. Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on, 37(6), 1063-1076.

Cheshire, D. G., Evans, M. A., & Dean, C. J. (2001). Haptic Modelling - An Alternative Industrial Design Methodology? Paper presented at the EuroHaptics International Conference, Birmingham, UK.

Christou, C., & Wing, A. (2001). Friction and curvature judgement. Retrieved 14th August, 2007, from eurohaptics.vision.ee.ethz.ch/2001/christou.pdf

Ciriello, V. M., Maikala, R. V., Dempsey, P. G., & O'Brien, N. V. (2010). Psychophysically determined forces of dynamic pushing for female industrial workers: Comparison of two apparatuses. Applied Ergonomics, 41(1), 141-145.

Craddock, M., & Lawson, R. (2010). The Effects of Temporal Delay and Orientation on Haptic Object Recognition. Attention, Perception, & Psychophysics, 72(7), 1975-1980.

Craig, J. C. (1985). Attending to two fingers: Two hands are better than one. Perception & Psychophysics, 38, 496-511.

Davidson, P. W. (1972). Haptic judgements of curvature by blind and sighted humans. Journal of Experimental Psychology, 93, 43-55.

Davies, O. L. (Ed.). (1956). The Design and Analysis of Industrial Experiments (2nd ed.): Oliver and Boyd.

Deol, K. K. (1999). A design advice tool presenting usability guidance for virtual environments. Paper presented at the Workshop On User Centered Design And Implementation Of Virtual Environments.

Díaz, I., Hernantes, J., Mansa, I., Lozano, A., Borro, D., Gil, J. J., et al. (2006). Influence of multisensory feedback on haptic accessibility tasks. Virtual Reality, V10(1), 31-40.

Dinh, H. Q., Walker, N., Hodges, L. F., Chang, S., & Kobayashi, A. (1999, 13-17 Mar 1999). Evaluating the importance of multi-sensory input on memory and the sense of presence in virtual environments. Paper presented at the Virtual Reality, 1999. Proceedings., IEEE.

Egeth, H. E., Kamlet, A. S., & Bell, R. A. (1970). The reversal of classical contrast in temperature perception. Psychonomic Science, 19, 96.

Ehrich, J. M., Flanders, M., & Soechting, J. F. (2008). Factors Influencing Haptic Perception of Complex Shapes. IEEE Transactions on Haptics, 1(1), 19-26.

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. Nature, 415(6870), 429-433.

Essock, E. A., Krebs, W. K., & Prather, J. R. (1992). An anisotropy of human tactile sensitivity and its relation to the visual oblique effect. Experimental Brain Research, 91, 520-524.

Field, A. (2009). Discovering Statistics Using SPSS (3rd ed.). London: Sage.

Gabbard, J. L. (1997). A Taxonomy of Usability Characteristics in Virtual Environments. Virginia Polytechnic Institute and State University, Blacksberg.

Gabbard, J. L., Hix, D., & Swan, J. E. (1999). User-Centered Design and Evaluation of Virtual Environments. IEEE Comput. Graph. Appl., 19(6), 51-59.

Gabbard, J. L., & Swan, J. E. (2008). Usability Engineering for Augmented Reality: Employing User-Based Studies to Inform Design. IEEE Transactions on Visualization and Computer Graphics, 14(3), 513-525.

Gescheider, G. A. (1985). Psychophysics: Method, Theory, and Application. London: Lawrence Erlbaum Associates.

Ghez, C., Rikakis, T., Dubois, R. L., & Cook, P. R. (2000). An Auditory Display System for Aiding Interoint Coordination. Paper presented at the Proceedings of the International Conference on Auditory Display, Atlanta, USA.

Gibson, J. J. (1966). The senses considered as perceptual systems. Boston: Houghton-Mifflin.

Goodwin, A., & Wheat, H. (1992). Human tactile discrimination of curvature when contact area with the skin remains constant. Experimental Brain Research, 88(2), 447-450.

Goodwin, A. W., John, K. T., & Marceglia, A. H. (1991). Tactile discrimination of curvature by humans using only cutaneous information from the fingerpads. Experimental Brain Research, 86(3), 663-672.

Gordon, I. E., & Morison, V. (1982). The haptic perception of curvature. Percept Psychophys, 31(5), 446-450.

Green, D. M. (1982). Profile analysis: A different view of auditory intensity discrimination. American Psychologist, 38, 133-142.

Green, D. M., Kidd, G., Jr., & Picardi, M. C. (1983). Successive versus simultaneous comparison in auditory discrimination. Journal of the Acoustical Society of America, 73, 639-643.

Griffiths, G., Sharples, S., & Wilson, J. R. (2006). Performance of new participants in virtual environments: The Nottingham tool for assessment of interaction in virtual environments (NAIVE). International Journal of Human-Computer Studies, 64, 240-250.

Guest, S., Catmur, C., Lloyd, D., & Spence, C. (2002). Audiotactile interactions in roughness perception. Experimental Brain Research, 146(2), 161-171.

Hale, K. S., & Stanney, K. M. (2004). Deriving haptic design guidelines from human physiological, psychophysical, and neurological foundations. Computer Graphics and Applications, IEEE, 24(2), 33-39.

Han, S. H., Jung, E. S., Jung, M., Kwahk, J., & Park, S. (1998). Psychophysical Methods and Passenger Preferences of Interior Designs. Applied Ergonomics, 29(6), 499-506.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In P. A. Hancock & N. Meshkati (Eds.), Human Mental Workload (pp. 139-183). Amsterdam: Elsevier.

Harwin, W. S., & Melder, N. (2002). Improved Haptic Rendering for Multi-Finger Manipulation Using Friction Cone based God-Objects. Paper presented at the EuroHaptics International Conference, Edinburgh, UK.

Hayward, V. (2004). Display of Haptic Shapes at Different Scales. Paper presented at the EuroHaptics International Conference, Munich, Germany.

Heller, M. A. (1983). Haptic dominance in form perception with blurred vision. Perception, 12, 607-613.

Henriques, D. Y., & Soechting, J. F. (2003). Bias and sensitivity in the haptic perception of geometry. Exp Brain Res, 150(1), 95-108.

Hermann, T. (2008, June 24-27, 2008). Taxonomy and Definitions for Sonification and Auditory Display. Paper presented at the Proceedings of the 14th International Conference on Auditory Display, Paris, France.

Hermann, T., & Hunt, A. (2005). An Introduction to Interactive Sonification. IEEE Multimedia, 12(2), 20-24.

Horst, B. J., & Kappers, A. (2007). Curvature discrimination in various finger conditions. Experimental Brain Research, 177, 304-311.

Hsia, P. T., & Drury, C. G. (1986). A simple method of evaluating handle design. Applied Ergonomics, 17(3), 209-213.

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

ISO/IEC 25062. (2006). Common Industry Format (CIF) for usability test reports. Retrieved. from.

Ittyerah, M., & Marks, L. E. (2007). Memory for Curvature of Objects: Haptic Touch vs. Vision. British Journal of Psychology, 98, 589-610.

Ittyerah, M., & Marks, L. E. (2008). Intramodal and cross-modal discrimination of curvature: Haptic touch versus vision. Current Psychology Letters, 24(1), 1-11.

Johnson, K. O. (2001). The Roles and Functions of Cutaneous Mechanoreceptors. Current Opinion in Neurobiology, 11, 455-461.

Johnston, A., & Passmore, P. J. (1994). Independent encoding of surface orientation and surface curvature. Vision Res, 34(22), 3005-3012.

Jousmaki, V., & Hari, R. (1998). Parchment-skin illusion: sound-biased touch. Current Biology, 8(6), R190-R191.

Jovanov, E., Starcevic, D., Wegner, K., Karron, D., & Radivojevic, V. (1998). Acoustic Rendering as Support for Sustained Attention during Biomedical Processing. Paper presented at the Proceedings of the Fifth International Conference on Auditory Display, Glasgow, UK.

Kalawsky, R. S., Bee, S. T., & Nee, S. P. (1999). Human Factors Evaluation Techniques to Aid Understanding of Virtual Interfaces. BT Technology Journal, 17(1), 128-141.

Kappers, A. M., & Koenderink, J. J. (1996). Haptic unilateral and bilateral discrimination of curved surfaces. Perception, 25(6), 739-749.

Kappers, A. M., Koenderink, J. J., & te Pas, S. F. (1994). Haptic discrimination of doubly curved surfaces. Perception, 23(12), 1483-1490.

Kappers, A. M. L., Koenderink, J. J., & Lichtenegger, I. (1994). Haptic identification of curved surfaces. Perception & Psychophysics, 56(1), 53-61.

Kappers, A. M. L., Koenderink, J. J., & Oudenaarden, G. (1996). Differences between haptic and visual judgements of curvature: DUP Delft University Press.

Klatzky, R. L., Lederman, S. J., & Metzger, V. A. (1985). Identifying objects by touch: An "expert system". Perception & Psychophysics, 37(4), 299-302.

Klatzky, R. L., Pai, D. K., & Krotkov, E. P. (2000). Perception of Material from Contact Sounds. Perception, 9(4), 399-410.

Koenderink, J. J., & van Doorn, A. J. (1992). Surface shape and curvature scales. Image & Vision Computing, 10(8), 557-565.

Kramer, G., Bargar, R., Barras, S., Berger, J., Evreinov, G., Fitch, W. T., et al. (1999). Sonification Report: Status of the Field and Research Agenda.

Lakatos, S., & Marks, L. E. (1999). Haptic Form Perception: Relative Salience of Local and Global Features. Perception and Psychophysics, 61(5), 895-908.

Lampton, D. R., Knerr, B. W., Goldberg, S. L., Bliss, J. P., Moshell, M. J., & Blau, B. S. (1994). Assessment Battery: Development and Evaluation. ????, ???

Lederman, S. J. (1976). The "callus-thenics" of touching. Canadian Journal of Psychology, 30, 82-89.

Lederman, S. J., & Klatzky, R. L. (1987). Hand Movements: A Window into Haptic Object Recognition. Cognitive Psychology, 19, 342-368.

Lederman, S. J., & Klatzky, R. L. (1997). Relative Availability of Surface and Object Properties During Early Haptic Processing. Journal of Experimental Psychology: Human Perception and Performance, 23(6), 1680-1707.

Lederman, S. J., Morgan, T., Hamilton, C., & Klatzky, R. L. (2002). Integrating Multimodal Information about Surface Texture via a Probe: Relative Contributions of Haptic and Touch-Produced Sound Sources. Paper presented at the Proceedings of the 10th Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems.

Lederman, S. J., & Taylor, M. M. (1972). Fingertip force, surface geometry, and the perception of roughness by active touch. Perception & Psychophysics, 12, 401-408.

Liu, A., Tharp, G., French, L., Lai, S., & Stark, L. (1993). Some of what one needs to know about using head-mounted displays to improve teleoperator performance. Robotics and Automation, IEEE Transactions on, 9(5), 638-648.

Livingston, M. A., Zanbaka, C., Swan, J. E., & Smallman, H. S. (2005). Objective Measures for the Effectiveness of Augmented Reality. Paper presented at the Proceedings of the IEEE Virtual Reality.

Loomis, J. M. (1981). Tactile pattern perception. Perception, 10, 5-27.

Loomis, J. M., Klatzky, R. L., & Lederman, S. J. (1991). Similarity of tactual and visual picture recognition with limited field of view. Perception, 20, 167-177.

Loomis, J. M., & Lederman, S. J. (1986). Tactual Perception. In K. R. Boff, L. Kaufman & J. P. Thomas (Eds.), Handbook of perception and human performance (Vol. Vol.2, Cognitive processes and performance, pp. 31/31-31/41). New York: Wiley.

Louw, S., Kappers, A., & Koenderink, J. (2002). Haptic discrimination of stimuli varying in amplitude and width. Experimental Brain Research, 146(1), 32-37.

Louw, S., Kappers, A. M. L., & Koenderink, J. J. (2000). Haptic detection thresholds of Gaussian profiles over the whole range of spatial scales. Experimental Brain Research, 132, 369-374.

Maciel, A., Sarni, S., Buchwalder, O., Boulic, R., & Thalmann, D. (2004). Multi-Finger Haptic Rendering of Deformable Objects. Paper presented at the Eurographics Symposium on Virtual Environments, Grenoble, France.

MathForum. (2008). Segments of Circles. Retrieved January, 2008, from http://mathforum.org/dr.math/faq/faq.circle.segment.html#7

McGee, M. R., Gray, P., & Brewster, S. (2001). Feeling Rough: Multimodal Perception of Virtual Roughness. Retrieved 29th August, 2007, from http://www.eurohaptics.vision.ee.ethz.ch/2001/mcgee.pdf

Moore, B. C. J. (2004). An Introduction to the Psychology of Hearing. London: Elsevier Academic Press.

Neuhoff, J., Knight, R., & Wayand, J. (2002, July 2-5). Pitch Change, Sonification, and Musical Expertise: Which Way is Up? Paper presented at the Proceedings of the 2002 International Conference on Auditory Display, Kyoto, Japan.

Nichols, S., Haldane, C., & Wilson, J. R. (2000). Measurement of presence and its consequences in virtual environments. International Journal of Human-Computer Studies, 52(3), 471-491.

Nichols, S., & Patel, H. (2002). Health and Safety Implications of Virtual Reality: A Review of Empirical Evidence. Applied Ergonomics, 33, 251-271.

Nielsen, J. (1993). Usability engineering ([New ed.] ed.). Boston ; London: AP Professional.

Noldus. (2009). Observer XT. Retrieved 1st July, 2009, from http://www.noldus.com/human-behavior-research/products/the-observer-xt

Norman, D. (1988). The Design of Everyday Things. New York: Basic Books.

Norman, J. F., Clayton, A. M., Lianekhammy, J., & Zielke, G. (2004). The Visual and Haptic Perception of Natural Object Shape. Perception and Psychophysics, 66(2), 342-351.

Pasquero, J., & Hayward, V. (2003, July). STReSS: A Practical Tactile Display System with One Millimeter Spatial Resolution and 700 Hz Refresh Rate. Paper presented at the EuroHaptics International Conference, Dublin, Ireland.

Patel, H., Stefani, O., Sharples, S., Hoffmann, H., Karaseitanidis, I., & Amditis, A. (2006). Human centred design of 3-D interaction devices to control virtual environments. International Journal of Human-Computer Studies, 64(3), 207-220.

Pont, S. C., Kappers, A. M., & Koenderink, J. J. (1997). Haptic curvature discrimination at several regions of the hand. Percept Psychophys, 59(8), 1225-1240.

Pont, S. C., Kappers, A. M., & Koenderink, J. J. (1999). Similar mechanisms underlie curvature comparison by static and dynamic touch. Percept Psychophys, 61(5), 874-894.

Preece, J., Rodgers, Y., & Sharp, H. (2002). Interaction Design: Beyond Human-Computer Interaction. New Jersey: John Wiley & Son, inc.

Provancher, W., R., Cutkosky, M., R., Kuchenbecker, K., J., & Niemeyer, G. (2005). Contact Location Display for Haptic Perception of Curvature and Object Motion. Int. J. Rob. Res., 24(9), 691-702.

Prytherch, D., & Jerrard, B. (2003). Haptics, the Secret Senses; the covert nature of the haptic senses in creative tacit skills.

Richard, P., Birebent, G., Burdea, G., Gomez, D., & Langrana, N. (1996). Effect of Frame Rate and Force Feedback on Virtual Object Manipulation. Presence: Teleoperators & Virtual Environments, 5(1), 95-108.

Rock, I., & Victor, J. (1964). Vision and touch: An experimentally created conflict between the two senses. Science, 143, 594-596.

Rose, D. (2006). Psychophysical Methods. In G. M. Breakwell, S. Hammond, C. Fife-Schaw & J. A. Smith (Eds.), Research Methods in Psychology. London: Sage.

SATIN Consortium. (2006). Project Presentation (D0). Milano: Politecnico di Milano. (P. d. Milano o. Document Number)

SATIN Consortium. (2007). Users' Specification and Scenarios (D3) (No. D3). Milan: Politecnico di Milano. (P. d. Milano o. Document Number)

SATIN Consortium. (2008a). Testing and Evaluation (D9.1 v1). Milano: Politecnico di Milano. (P. d. Milano o. Document Number)

SATIN Consortium. (2008b). Testing and Evaluation (D9.1 v2). Milano: Politecnico di Milano. (P. d. Milano o. Document Number)

SATIN Consortium. (2008c). Theoretical Foundations (D5.1). Milano: Politecnico di Milano. (P. d. Milano o. Document Number)

SATIN Consortium. (2009a). Heuristic Evaluation Report. Milano: Politecnico di Milano. (P. d. Milano o. Document Number)

SATIN Consortium. (2009b). SATIN Tool Testing and Evaluation (D9.2). Milano: Politecnico di Milano. (P. d. Milano o. Document Number)

Sekuler, R., Nash, D., & Armstrong, R. (1973). Sensitive, objective procedure for evaluating response to light touch. Neurology, 23, 1282-1291.

Sekuler, R., Sekuler, A. B., & Lau, R. (1997). Sound alters visual motion perception. Nature, 385, 308.

Sener, B. (2002). Evaluating a Haptic Modelling System with Industrial Designers. Paper presented at the EuroHaptics International Conference, Edinburgh, Scotland.

Sener, B., Pedgley, O., Wormald, P., & Campbell, I. (2003). Incorporating the FreeForm Haptic Modelling System into New Product Development. Dublin, Ireland.

Sensable. (2007, 2011). FreeForm Modelling and Modelling Plus Systems. Retrieved July, 2007, from http://www.sensable.com/products-freeform-systems.htm

Shams, L., Kamitani, Y., & Shimojo, S. (2000). What you see is what you hear. Nature, 408, 788.

Shelley, S. B., Alonso, M., Hollowood, J., Pettitt, M., Sharples, S., Hermes, D., et al. (2009, September 10-11). Interactive Sonification of Curve Shape and Curvature Data. Paper presented at the Haptic and Audio Interaction Design, Dresden, Germany.

Shih, R., Dubrowski, A., & Carnahan, H. (2009). Evidence for Haptic Memory. Paper presented at the Third Joint EuroHaptics Conference and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems.

Shneiderman, B., & Plaisant, C. (2005). Designing the User Interface: Pearson Education, Inc.

Soechting, J. F., Song, W., & Flanders, M. (2006). Haptic Feature Extraction. Cerebral Cortex, 16, 1168-1180.

Spence, C., Nicholls, M. E. R., & Driver, J. (2001). The cost of expecting events in the wrong sensory modality. Perception and Psychophysics, 63(330-336).

Spence, C., Pavani, F., & Driver, J. (2000). Crossmodal links between vision and touch in covert endogenous spatial attention. Journal of Experimental Psychology, Human Perception and Performance, 26, 1298-1319.

Stanney, K. M., Mollaghasemi, M., Reeves, L., Breaux, R., & Graeber, D. A. (2003). Usability engineering of virtual environments (VEs): identifying multiple criteria that drive effective VE system design. International Journal of Human-Computer Studies, 58(4), 447-481.

Stanney, K. M., Mourant, R. R., & Kennedy, R. S. (1998). Human Factors Issues in Virtual Environments: A Review of the Literature. Presence: Teleoperators and Virtual Environments, 7(4), 327-351.

Stanton, N. E., & Edworthy, J. (Eds.). (1999). Human Factors in Auditory Warnings. Aldershot, UK: Ashgate.

Stevens, J. C., & Patterson, M. Q. (1995). Dimensions of spatial acuity in the touch sense: changes over the life span. Somatosensory and Motor Research, 12, 29-47.

Stockman, T. (2005, July 6-9). Interactive Sonification of Spreadsheets. Paper presented at the Proceedings of the 5th International Conference on Auditory Display, Limerick Ireland.

Sutcliffe, A., & Gault, B. (2004). Heuristic evaluation of virtual reality applications. Interacting with Computers, 16(4), 831-849.

Swan, J. E., & Gabbard, J. L. (2005). Survey of User-Based Experimentation in Augmented Reality. Paper presented at the Proceedings of First International Conference in Virtual Reality.

Tan, H. Z., Srinivasan, M. A., Eberman, B., & Cheng, B. (1994). Human Factors for the Design of Force-Reflecting Haptic Interfaces. Dyamic Systems and Control, 55(1), 353-359.

Tao, N., Doug, A. B., & Jian, C. (2006). Increased display size and resolution improve task performance in Information-Rich Virtual Environments. Paper presented at the Proceedings of Graphics Interface 2006.

Teghtsoonian, R. (1971). On the Exponents in Steven's Law and The Constant in Ekman's Law. Psychological Review, 78(1), 71-80.

van Beers, R. J., Sittig, A. C., & van der Gon, J. J. D. (1998). The Precision of Proprioceptive Position Sense. Experimental Brain Research, 122, 367-377.

Van Doren, C. L., Pelli, D. G., & Verillo, R. T. (1987). A device for measuring tactile spatiotemporal sensitivity. Journal of the Acoustical Society of America, 81(6), 1906-1916.

Vogels, I. M., Kappers, A. M., & Koenderink, J. J. (1996). Haptic aftereffect of curved surfaces. Perception, 25(1), 109-119.

Vogels, I. M., Kappers, A. M., & Koenderink, J. J. (1999). Influence of shape on haptic curvature perception. Acta Psychol (Amst), 100(3), 267-289.

Wagner, C. R., Lederman, S. J., & Howe, R. D. (2002, March 24-25). A Tactile Shape Display Using RC Servomotors. Paper presented at the Tenth Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems, Orlando, USA.

Walker, B. N. (2002). Magnitude estimation of conceptual data dimensions for use in sonification. Journal of Experimental Psychology: Applied, 8(4), 211-221.

Walker, B. N., & Kramer, G. (1993, 4th-6th November). Mappings and Metaphors in Auditory Display. Paper presented at the Proceedings of the Third International Conference on Auditory Display, Palo Alto, USA.

Walker, B. N., & Mauney, L. M. (2004, July 6-9). Individual Differences, Cognitive Abilities, and the Interpretation of Auditory Graphs. Paper presented at the Tenth International Conference on Auditory Display, Sydney, Australia.

Walker, B. N., & Nees, M. A. (2005, July 6-9). An agenda for research and development of multimodal graphs. Paper presented at the Proceedings of International Conference on Auditory Display (ICAD2005), Limerick, Ireland.

Walker, B. N., & Nees, M. A. (2005, July 6-9). Brief Training for Performance of a Point Estimation Sonification Task. Paper presented at the Eleventh Meeting of the International Conference on Auditory Display, Limerick, Ireland.

Walker, B. N., & Nees, M. A. (2010). Theory of Sonification. In H. A. Hunt & J. Neuhoff (Eds.), Handbook of Sonification. New York: Academic Press.

Weinstein, S. (1968). Intensive and extensive aspects of tactile sensitivity as a function of body part, sex and laterality. In D. R. Kenshalo (Ed.), The skin senses. Springfield, Illinois.: Thomas.

Whelan, R. J. C. R. (1996). Pilot study for the validation of a virtual environment cognitive ergonomics evaluation tool.

Wickens, C. D., Dixon, S. R., & Seppelt, B. (2005). Auditory Pre-emption Versus Multiple Resources: Who Wins in Interruption Management? Human Factors and Ergonomics Society Annual Meeting Proceedings, 49, 463-167.

Wickens, C. D., Lee, J. D., Liu, Y., & Gordon Becker, S. E. (2004). An Introduction to Human Factors Engineering (Second ed.). New Jersey: Pearson Educational International.

Wijntjes, M., Sato, A., Hayward, V., & Kappers, A. (2009). Local Surface Orientation Dominates Haptic Curvature Discrimination. IEEE Transactions on Haptics.

Wijntjes, M., Sato, A., Kappers, A., & Hayward, V. (2008). Haptic Perception of Real and Virtual Curvature. In Haptics: Perception, Devices and Scenarios (pp. 361-366).

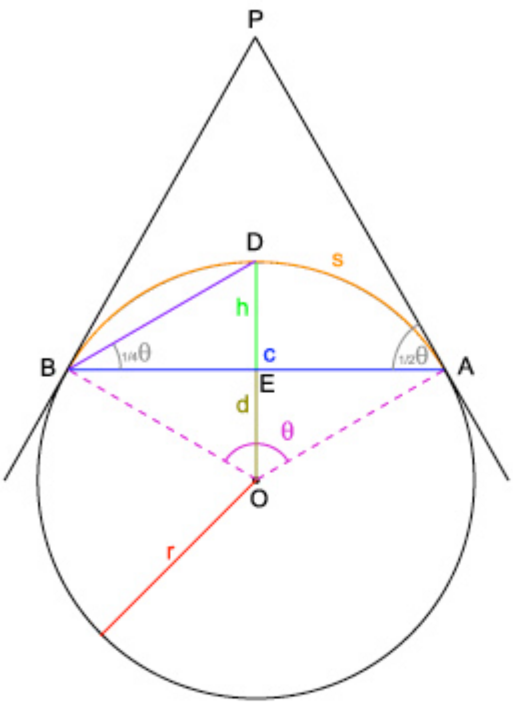
Wilson, J. R., & Eastgate, R. M. (2002). Structured Development of Virtual Environments. In K. M. Stanney (Ed.), Handbook of Virtual Environments: CRC.

Witmer, B. G., & Singer, M. J. (1998). Measuring Presence in Virtual Environments: A Presence Questionnaire. Presence: Teleoperators & Virtual Environments, 7(3), 225-240.

Wolfe, J. M., Kluender, K. R., Levi, D. M., Bartoshuk, L. M., Herz, R. S., Klatzky, R. L., et al. (2006). Sensation and Perception (First ed.). Sunderland, Massachusetts: Sinauer Associates, Inc.

Appendix A: Geometric Relationships between Various Stimulus Dimensions

The following geometric formulae were used to transpose thresholds given in one dimension e.g. base-to-peak height (h) to another e.g. curvature (through calculating r).



s = arc length

c = chord length

h = height

d = sagitta

r = Radius

By knowing two of either the cord (c), radius (r), or height (h), the following dimensions may be calculated:

If c and r are known, then:

θ = 2 arcsin(c/[2r])
s = r theta
d = r cos(theta/2)
h = r - d

If c and h are known, then:

r = (c2+4h2)/(8h)
θ = 2 arcsin(c/[2r])
s = r theta
d = r - h

These formulae are generally known, although these particular ones were obtained from an online source (MathForum 2008).

Appendix B: Study 1 Experiment Documents

This appendix gives details of the documents that were used to conduct Study 1. The documents can be found on the accompanying CD: /study-1_ documents/.

Study Information and Consent Form (one for each session conducted):
info-consent_HAPTIC.doc
info-consent_VISUAL-HAPTIC.doc
info-consent_VISUAL-HAPTIC.doc

Study Procedure (one for each session conducted):
study-procedure_HAPTIC.doc
study-procedure_VISUAL-HAPTIC.doc
study-procedure_VISUAL.doc

'More' or 'Less' Curved? Comprehension Test
more-less-curved.doc

Appendix C: Stimuli Specification (Studies 1 and 3)

Geometric Properties to which Specification Relate

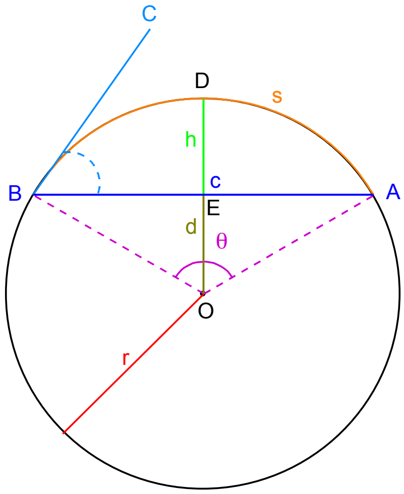


Table C-1: Stimulus Specification - Study 1

Pres. Ref.	Set Ref.	Curvature (/m)	Radius (r)	Chord (c)	Theta (radians)	arc (s)	Segment Height (h)	ABC	Gradient
G1	A	0.28	358.57	20	0.056	20.003	0.139	1.598	0.014
F1	A	0.30	335.27	20	0.060	20.003	0.149	1.709	0.015
E1	A	0.32	313.47	20	0.064	20.003	0.160	1.828	0.016
D1	A	0.34	293.1	20	0.068	20.004	0.171	1.955	0.017
C1	A	0.36	274.05	20	0.073	20.004	0.183	2.091	0.018
B1	A	0.39	256.23	20	0.078	20.005	0.195	2.236	0.020
A1	A	0.42	239.58	20	0.084	20.006	0.209	2.392	0.021
SA	A	0.45	224.01	20	0.089	20.007	0.223	2.558	0.022
A2	A	0.48	209.45	20	0.096	20.008	0.239	2.736	0.024
B2	A	0.51	195.83	20	0.102	20.009	0.255	2.927	0.026
C2	A	0.55	183.1	20	0.109	20.010	0.273	3.130	0.027
D2	A	0.58	171.2	20	0.117	20.011	0.292	3.348	0.029
E2	A	0.62	160.07	20	0.125	20.013	0.313	3.581	0.031
F2	A	0.67	149.67	20	0.134	20.015	0.334	3.830	0.033
G2	A	0.71	139.94	20	0.143	20.017	0.358	4.097	0.036
G1	B	0.94	106.95	20	0.187	20.029	0.469	5.364	0.047
F1	B	1.00	100	20	0.200	20.033	0.501	5.738	0.050
E1	B	1.06	93.9	20	0.213	20.038	0.534	6.113	0.053
D1	B	1.13	88.17	20	0.227	20.043	0.569	6.511	0.057
C1	B	1.21	82.78	20	0.242	20.049	0.606	6.937	0.061
B1	B	1.29	77.73	20	0.258	20.056	0.646	7.391	0.065
A1	B	1.37	72.99	20	0.275	20.063	0.688	7.874	0.069
SB	B	1.46	68.53	20	0.293	20.072	0.734	8.390	0.073
A2	B	1.55	64.35	20	0.312	20.081	0.782	8.939	0.078
B2	B	1.66	60.42	20	0.333	20.092	0.833	9.526	0.083
C2	B	1.76	56.74	20	0.354	20.105	0.888	10.150	0.089
D2	B	1.88	53.27	20	0.378	20.119	0.947	10.819	0.095

Pres. Ref.	Set Ref.	Curvature (/m)	Radius (r)	Chord (c)	Theta (radians)	arc (s)	Segment Height (h)	ABC	Gradient
E2	B	2.00	50.02	20	0.403	20.136	1.010	11.531	0.101
F2	B	2.13	46.97	20	0.429	20.154	1.077	12.291	0.108
G2	B	2.27	44.1	20	0.457	20.175	1.149	13.105	0.115
G1	C	10.28	9.73	5	0.520	5.057	0.327	14.886	0.131
F1	C	10.94	9.14	5	0.554	5.065	0.349	15.872	0.139
E1	C	11.66	8.58	5	0.591	5.074	0.372	16.938	0.149
D1	C	12.42	8.05	5	0.632	5.084	0.398	18.091	0.159
C1	C	13.23	7.56	5	0.674	5.096	0.425	19.308	0.170
B1	C	14.08	7.1	5	0.720	5.110	0.455	20.614	0.182
A1	C	14.99	6.67	5	0.768	5.125	0.486	22.010	0.194
SC	C	15.97	6.26	5	0.822	5.143	0.521	23.535	0.208
A2	C	17.01	5.88	5	0.878	5.164	0.558	25.158	0.223
B2	C	18.12	5.52	5	0.940	5.189	0.599	26.926	0.239
C2	C	19.31	5.18	5	1.007	5.218	0.643	28.853	0.257
D2	C	20.53	4.87	5	1.078	5.251	0.691	30.883	0.276
E2	C	21.88	4.57	5	1.158	5.291	0.744	33.160	0.298
F2	C	23.31	4.29	5	1.244	5.338	0.804	35.640	0.321
G2	C	24.81	4.03	5	1.338	5.394	0.869	38.337	0.348

Table C-2: Stimulus Specification - Study 3

Pres. Ref.	Set Ref.	Curvature (/m)	Radius (r)	Chord (c)	Theta (radians)	arc (s)	Segment Height (h)	ABC	Gradient
8M	D	1.08	92.55	10	0.108	10.005	0.135	3.10	0.027
4M	D	1.27	78.74	10	0.127	10.007	0.159	3.64	0.032
2M	D	1.36	73.28	10	0.137	10.008	0.171	3.91	0.034
1M	D	1.41	70.83	10	0.141	10.008	0.177	4.05	0.035
Std	D	1.46	68.53	10	0.146	10.009	0.183	4.18	0.037
1P	D	1.51	66.38	10	0.151	10.009	0.189	4.32	0.038
2P	D	1.55	64.36	10	0.156	10.010	0.195	4.46	0.039
4P	D	1.65	60.67	10	0.165	10.011	0.206	4.73	0.041
8P	D	1.84	54.43	10	0.184	10.014	0.230	5.27	0.046
8M	E	1.09	92.10	30	0.327	30.134	1.230	9.37	0.082
4M	E	1.27	78.54	30	0.384	30.185	1.446	11.01	0.096
2M	E	1.37	73.18	30	0.413	30.214	1.554	11.83	0.104
1M	E	1.41	70.78	30	0.427	30.229	1.608	12.23	0.107
Std	E	1.46	68.53	30	0.441	30.245	1.662	12.64	0.111
1P	E	1.51	66.43	30	0.456	30.261	1.716	13.05	0.114
2P	E	1.55	64.45	30	0.470	30.278	1.770	13.46	0.118
4P	E	1.64	60.85	30	0.498	30.312	1.878	14.27	0.125
8P	E	1.83	54.78	30	0.555	30.388	2.094	15.89	0.140
8M	F	3.26	30.70	10	0.327	10.045	0.410	9.37	0.082
4M	F	3.82	26.18	10	0.384	10.062	0.482	11.01	0.096
2M	F	4.10	24.39	10	0.413	10.071	0.518	11.83	0.104
1M	F	4.24	23.59	10	0.427	10.076	0.536	12.23	0.107
Std	F	4.38	22.84	10	0.441	10.082	0.554	12.64	0.111
1P	F	4.52	22.14	10	0.456	10.087	0.572	13.05	0.114
2P	F	4.65	21.48	10	0.470	10.093	0.590	13.46	0.118
4P	F	4.93	20.28	10	0.498	10.104	0.626	14.27	0.125
8P	F	5.48	18.26	10	0.555	10.129	0.698	15.89	0.140

Appendix D: Summary Results

The research undertaken has generated a large number of hypotheses, the pattern and significance of which may be difficult to follow in the reported results. It was therefore felt that it would be helpful to present these in a summarised graphical format. Therefore the results for each study are presented within this appendix. Non-significant results are shown by a grey box, significant results are shown as follows: >0.05 = Green, >0.01 = light green.

Study 1

Perception (H1-H3)

H1: There was a significant difference between Weber fractions for high and low curvature.

H2: There is no significant difference between Weber fractions for modality.

H3: There is no interaction effect between curvature and modality.

Performance and Interaction

Table D-1: Hypothesis H4-H21 and H86-H91

Metrics	Main Effects			Interactions		
	Curvature	Modality	MagDiff	Curv*Modality	Curv*MagDiff	Mod*MagDiff
RT	✓	✓	✓			✓
ACC	✓	✓	✓		✓	
CON	✓	✓	✓			✓
COMP	✓	✓	✓			✓

Predictability and Performance Characteristics

Table D-2: Hypothesis H22-H237 and H38-H85

Metrics	Curvature		Modality			Magnitude Difference		
	Low	High	Haptic	Visual	VisHap	Small	Medium	Large
Response Time Correct > Incorrect	✓	✓		✓	✓			
Response Time High > Low Confidence	✓	✓	✓	✓	✓	✓	✓	✓
Response Time HCC > LCC	✓	✓	✓	✓	✓	✓	✓	✓
HCC > HCI				✓				
HCC > LCI	✓	✓	✓	✓	✓	✓	✓	
Correlation Response Time & Accuracy								
Correlation Response Time & Confidence				✓				
Correlation Accuracy & Confidence			✓		✓			

Study 2 – Experiment 1 (Chapter 5)

Performance

Table D-3: Hypothesis H1-H15

Metrics	Main Effects			Interactions	
	Practice	SoundType	Orientation	Practice*ST	Practice*Ori
RT	✓				
ACC	✓	✓	✓		
CON	✓				

Practice

H16: There was no significant association between practice and error rates.

H17: There was a significant association between sound type and error rates.

User Experience

H18: There was no significant effect of practice on attitude to user experience issues.

Table D-4: Hypothesis H19-H23

	User Experience Issue				
	Enjoyment	Difficulty	Performance	Improvement	Irritability
Association of frequency with attitude	✓		✓		

Workload (H24-H30)

H24: There was no significant effect of practice on workload score

Table D-5: Hypothesis H23-H30

	Workload Sub-factors					
	Mental Demand	Physical Demand	Temporal Demand	Performance	Effort	Frustration
Effect of practice on workload score	✓					

Study 2 - Experiment 2 (Chapter 5)

Performance

Curve Shape

Table D-6: Hypothesis H1-H18 (Curve Shape)

Metrics	Main Effects			Interactions		
	Practice	SoundType	MagDiff	Practice*ST	Practice*MagDiff	ST*MagDiff
RT	✓		✓			✓
ACC	✓		✓			
CON		✓	✓			✓

Curvature

Table D-7: Hypothesis H1-H18 (Curvature)

Metrics	Main Effects			Interactions		
	Practice	SoundType	MagDiff	Practice*ST	Practice*MagDiff	ST*MagDiff
RT	✓	✓				✓
ACC	✓	✓		✓		✓
CON		✓				✓

User Experience Issues

Curve Shape

H19: There was no significant effect of practice on attitude to user experience issues

Table D-6: Hypothesis H20-H24 (Curve Shape)

	User Experience Issue				
	Enjoyment	Difficulty	Performance	Improvement	Irritability
Association of frequency with attitude	✓	✓	✓		

Curvature

H19: There was no significant effect of practice on attitude to user experience issues

Table D-7: Hypothesis H20-H24 (Curvature)

	User Experience Issue				
	Enjoyment	Difficulty	Performance	Improvement	Irritability
Association of frequency with attitude	✓	✓	✓		

Workload (H25-31)

Curve Shape

H25: There was a significant effect of practice on workload score.

Table D-8: Hypothesis H26-H31 (Curve Shape)

	Workload Sub-factors					
	Mental Demand	Physical Demand	Temporal Demand	Performance	Effort	Frustration
Effect of practice on workload score	✓				✓	

Curvature

H25: There was a significant effect of practice on workload score.

Table D-9: Hypothesis H26-H31 (Curvature)

	Workload Sub-factors					
	Mental Demand	Physical Demand	Temporal Demand	Performance	Effort	Frustration
Effect of practice on workload score	✓				✓	

Study 3 - Experiment 3 (Chapter 6)

Performance (H1-H3)

Curve Shape

Table D-10: Hypothesis H1-H3 (Curve Shape)

Main Effect	Metrics		
	RT	ACC	MagDiff
Sound Type	✓		✓

Curvature

Table D-10: Hypothesis H1-H3 (Curvature)

Main Effect	Metrics		
	RT	ACC	MagDiff
Sound Type			

User Experience Issues

Curve Shape

Table D-10: Hypothesis H4-H8 (Curve Shape)

	User Experience Issue				
	Enjoyment	Difficulty	Performance	Improvement	Irritability
Association of frequency with attitude			✓	✓	

Curvature

Table D-11: Hypothesis H4-H8 (Curvature)

	User Experience Issue				
	Enjoyment	Difficulty	Performance	Improvement	Irritability
Association of frequency with attitude		✓		✓	

Workload

Curve Shape

H9: There was no significant effect of presentation set on workload score

Workload

Curvature

H9: There was no significant effect of presentation set on workload score

Study 3 (Chapter 7)

Perception

H1: There was a significant effect of sound on difference threshold

H2: There was no significant effect of stimulus on difference threshold

H3: There was no interaction between stimuli and sound on difference threshold

Performance (H4-H21)

Table D-12: Hypothesis H4-H21

Metrics	Main Effects			Interactions			
	Sound	Stimulus	MagDiff	Sound*Stimulus	Sound*MagDiff	Stimulus*MagDiff	All
RT	✓		✓	✓		✓	✓
ACC	✓	✓	✓	✓	✓	✓	✓
CON	✓		✓		✓	✓	✓

Performance Characteristics

Table D-13: Hypothesis H22-H45

Metrics	Sound		
	NoS	CS	Curv
Response Time Correct > Incorrect	✓	✓	
Response Time High > Low Confidence	✓	✓	
Response Time HCC > LCC	✓	✓	
HCC > HCI	✓		
HCC > LCI	✓	✓	
Correlation Response Time & Accuracy			
Correlation Response Time & Confidence			
Correlation Accuracy & Confidence			✓

Interaction

H46: There was a significant association between sound and interaction style
H47: There was a significant effect of sound on exploration duration
H48: There was a significant effect of location on exploration duration
H49: There was a significant interaction effect between sound and location on exploration duration

User Experience

Table D-14: Hypothesis H50-H55

	User Experience					
	Ease of Use	Difficulty	Helpfulness	Concentration	Sense Used	Sense Conflict
Association of sound with	✓	✓	✓	✓	✓	

Workload (H56-H62)

H56: There was a significant effect of sound on workload score

Table D-15: Hypothesis H56-H62

	Workload Sub-factors					
	Mental Demand	Physical Demand	Temporal Demand	Performance	Effort	Frustration
Effect of sound on workload score	✓			✓	✓	

Study 3 (Further Analysis – Chapter 8)

Perception (H1-H2)

H1: There was a significant effect of Stimulus on JND
H2: There was no significant effect of Stimulus on Weber Fractions

Performance

Table D-16: Hypothesis H3-H11

Metrics	Main Effects		Interactions
	Stimulus	MagDiff	Stimulus*MagDiff
RT	✓	✓	✓
ACC	✓	✓	✓
CON	✓	✓	✓

Predictability (H12-H20)

Table D-17: Hypothesis H12-H20

Stimuli	Hypothesis	Metric	R^2 Contribution & Significance		R^2
			Stimulus Dimensions	MagDiff	
DEF	H12	RT	0.67	0.13	0.80
	H13	ACC	0.33	0.46	0.79
	H14	CON	0.33	0.27	0.60
BC	H15	RT	0.44	0.46	0.90
	H16	ACC	0.07	0.48	0.55
	H17	CON	0.10	0.66	0.76
BCDEF	H18	RT	0.00	0.00	0.01
	H19	ACC	0.34	0.38	0.72
	H20	CON	0.34	0.37	0.71

Appendix E: Psychophysical data for Haptic, Visual, and Visual-Haptic Modalities in Low and High Curvature Conditions

Calculation of Threshold

The method applied to produce the psychometric function is that described by Gescheider (1985). The percentage 'more than' scores were converted to a z-score. This converted data were then plotted against curvature to produce psychometric functions for each participant (see Figure for example). A linear regression line was fitted to this using SPSS. The point at which 75% (z=0.67) of responses were correct gave the upper difference threshold (DL_u), and the point at which 25% (z=-0.67) of responses were correct gave the lower difference threshold (DL_l). The JND (DL) was calculated by halving the difference between the upper and lower thresholds.

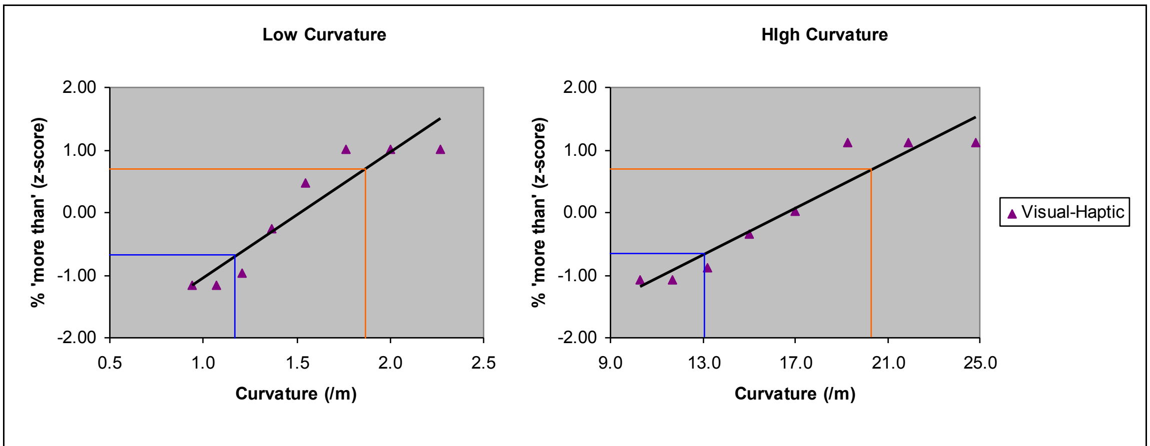


Figure D-1: Visual-Haptic psychometric function (Participant 8) for low and high curvature

Note: The upper difference threshold is shown in orange, and the lower difference threshold is shown in blue.

Limitations of the Method

The psychometric function produces a sigmoid (s-shaped) curve. Theoretically conversion of the data to z-scores will produce a straight line (Gescheider, 1985). In practice this had varying degrees of success as can be seen in Figure D-1. Even so it should be noted that across all conditions and participants the mean fit was 87%, which was considered to be acceptable. In addition, the fact that the haptic thresholds generated here were in alignment with those reported in the haptic literature, was seen as providing validation for the sufficiency of the method adopted (see Chapter 4 for comparison with literature).

Regression and Psychophysical Data

Table D-1: Haptic Low Curvature (1.46/m)

P-ID	St	R^2	Intercept (B_0)	Slope (B_1)	PSE	CE	Upper (DL_u)	Lower (DL_l)	JND (DL)	Weber Fraction
P1	1.46	0.95	-3.20	2.11	1.52	0.06	1.84	1.20	0.32	21.79
P2	1.46	0.94	-3.18	2.09	1.52	0.06	1.84	1.20	0.32	21.94
P3	1.46	0.87	-3.07	2.02	1.52	0.06	1.85	1.19	0.33	22.77
P4	1.46	0.86	-3.05	2.00	1.52	0.06	1.86	1.19	0.33	22.91
P5	1.46	0.84	-3.00	1.97	1.52	0.06	1.86	1.18	0.34	23.31
P7	1.46	0.89	-3.08	2.03	1.52	0.06	1.85	1.19	0.33	22.64
P8	1.46	0.90	-3.11	2.04	1.52	0.06	1.85	1.19	0.33	22.48
P9	1.46	0.97	-3.23	2.12	1.52	0.06	1.84	1.21	0.32	21.64
mean		0.90			1.52	0.06	1.85	1.19	0.33	22.43
SD		0.05			0.00	0.00	0.01	0.01	0.01	0.59

Table D-2: Haptic High Curvature (15.97/m)

P-ID	St	R^2	Intercept (B_0)	Slope (B_1)	PSE	CE	Upper (DL_u)	Lower (DL_l)	JND (DL)	Weber Fraction
P1	15.97	0.79	-2.90	0.17	16.67	0.70	20.52	12.82	3.85	24.11
P2	15.97	0.89	-3.08	0.19	16.67	0.70	20.29	13.05	3.62	22.68
P3	15.97	0.83	-2.98	0.18	16.63	0.66	20.37	12.88	3.74	23.44
P4	15.97	0.89	-3.09	0.19	16.61	0.64	20.22	13.01	3.60	22.56
P5	15.97	0.81	-2.95	0.18	16.68	0.71	20.46	12.89	3.79	23.70
P7	15.97	0.84	-3.00	0.18	16.66	0.69	20.38	12.93	3.72	23.31
P8	15.97	0.92	-3.14	0.19	16.68	0.71	20.24	13.11	3.56	22.32
P9	15.97	0.88	-3.07	0.19	16.61	0.64	20.23	12.99	3.62	22.68
mean		0.86			16.65	0.68	20.34	12.96	3.69	23.10
SD		0.04			0.03	0.03	0.11	0.10	0.10	0.63

Table D-3: Visual Low Curvature (1.46/m)

P-ID	St	R^2	Intercept (B_0)	Slope (B_1)	PSE	CE	Upper (DL_u)	Lower (DL_l)	JND (DL)	Weber Fraction
P1	1.46	0.88	-3.07	2.02	1.52	0.06	1.85	1.19	0.33	22.72
P2	1.46	0.89	-3.09	2.03	1.52	0.06	1.85	1.19	0.33	22.56
P3	1.46	0.85	-3.02	1.98	1.52	0.06	1.86	1.18	0.34	23.17
P4	1.46	0.88	-3.09	2.03	1.52	0.06	1.85	1.19	0.33	22.62
P5	1.46	0.84	-3.00	1.98	1.52	0.06	1.86	1.18	0.34	23.24
P7	1.46	0.85	-3.02	1.99	1.52	0.06	1.86	1.18	0.34	23.11
P8	1.46	0.91	-3.13	2.06	1.52	0.06	1.85	1.20	0.33	22.32
P9	1.46	0.89	-3.09	2.03	1.52	0.06	1.85	1.19	0.33	22.62
mean		0.87			1.52	0.06	1.85	1.19	0.33	22.79
SD		0.02			0.00	0.00	0.01	0.00	0.00	0.33

Table D-4: Visual High Curvature (15.97/m)

P-ID	St	R^2	Intercept (B_0)	Slope (B_1)	PSE	CE	Upper (DL_u)	Lower (DL_l)	JND (DL)	Weber Fraction
P1	15.97	0.79	-2.91	0.18	16.63	0.66	20.46	12.80	3.83	23.97
P2	15.97	0.85	-3.01	0.18	16.64	0.67	20.34	12.94	3.70	23.18
P3	15.97	0.86	-3.04	0.18	16.60	0.63	20.26	12.93	3.66	22.93
P4	15.97	0.82	-2.97	0.18	16.69	0.72	20.45	12.92	3.76	23.57
P5	15.97	0.91	-3.12	0.19	16.66	0.69	20.25	13.08	3.58	22.44
P7	15.97	0.85	-3.01	0.18	16.63	0.66	20.33	12.93	3.70	23.18
P8	15.97	0.86	-3.04	0.18	16.62	0.65	20.28	12.96	3.66	22.93
P9	15.97	0.83	-2.98	0.18	16.63	0.66	20.37	12.88	3.74	23.44
mean		0.85			16.64	0.67	20.34	12.93	3.71	23.20
SD		0.04			0.03	0.03	0.08	0.08	0.07	0.47

Table D-5: Visual-Haptic Low Curvature (1.46/m)

P-ID	St	R^2	Intercept (B_0)	Slope (B_1)	PSE	CE	Upper (DL_u)	Lower (DL_l)	JND (DL)	Weber Fraction
P1	1.46	0.87	-3.06	2.01	1.52	0.06	1.85	1.19	0.33	22.79
P2	1.46	0.91	-3.12	2.05	1.52	0.06	1.85	1.19	0.33	22.41
P3	1.46	0.87	-3.06	2.01	1.52	0.06	1.86	1.19	0.33	22.84
P4	1.46	0.87	-3.06	2.01	1.52	0.06	1.85	1.19	0.33	22.81
P5	1.46	0.95	-3.20	2.11	1.52	0.06	1.84	1.20	0.32	21.80
P7	1.46	0.95	-3.19	2.10	1.52	0.06	1.84	1.20	0.32	21.86
P8	1.46	0.87	-3.06	2.01	1.52	0.06	1.85	1.19	0.33	22.81
P9	1.46	0.90	-3.12	2.05	1.52	0.06	1.85	1.19	0.33	22.39
mean		0.90			1.52	0.06	1.85	1.19	0.33	22.46
SD		0.03			0.00	0.00	0.01	0.01	0.01	0.43

Table D-6: Visual-Haptic High Curvature (15.97/m)

P-ID	St	R^2	Intercept (B_0)	Slope (B_1)	PSE	CE	Upper (DL_u)	Lower (DL_l)	JND (DL)	Weber Fraction
P1	15.97	0.78	-2.90	0.17	16.66	0.69	20.51	12.81	3.85	24.11
P2	15.97	0.79	-2.90	0.17	16.65	0.68	20.50	12.80	3.85	24.11
P3	15.97	0.84	-2.99	0.18	16.60	0.63	20.32	12.88	3.72	23.31
P4	15.97	0.88	-3.07	0.18	16.66	0.69	20.30	13.02	3.64	22.80
P5	15.97	0.80	-2.93	0.18	16.65	0.68	20.45	12.84	3.81	23.84
P7	15.97	0.79	-2.91	0.17	16.70	0.73	20.55	12.84	3.85	24.11
P8	15.97	0.90	-3.11	0.19	16.64	0.67	20.22	13.06	3.58	22.44
P9	15.97	0.85	-3.01	0.18	16.64	0.67	20.34	12.94	3.70	23.18
mean		0.83			16.65	0.68	20.40	12.90	3.75	23.49
SD		0.05			0.03	0.03	0.12	0.10	0.10	0.65

Appendix F: Non-Normal Data and ANOVA

For ANOVA there is a requirement that the data should meet a number of assumptions. One of these is that it should be normally distributed. The normality of a distribution can be analysed in a number of ways; visual checking with a histogram, analysis of Skew or Kurtosis, or applying the Kolmogorov-Smirnov test (Field 2009). In the case of data here, normality was determined by testing for significant levels of Skew or Kurtosis. This was judged by converting the level of Skew or Kurtosis into a z-number, and a significant departure from normality was said to exist if z was greater than 1.96 (Field 2009). For some of the variables that were analysed within these studies there were varying degrees of non-normality; that is they had z-scores that exceeded 1.96 (for detailed results see CD /other-documents/DATA_discriptives-normality.xls).

Does this matter? The ANOVA is considered to be a 'robust' test. By this it is meant that even with violation of its underlying assumptions it should provide reliable results. It therefore becomes not so much a question of whether normality is violated, but whether the level of violation will affect the veracity of the results. So what we need to know is whether the effect of Skew and Kurtosis is such that the level of probability is no longer reliable i.e. the true probability of the distribution is too divergent from the normal 5% tails. This question is discussed by Davies (1956), who concludes that 'even extreme non-Normality has little serious effect on the probability levels' (see Table E-1 for effect of Skewness and Kurtosis on probability).

Table E-1: True percentage probability at various degrees of Skewness and Kurtosis (Davis, 1956)

		Skewness		
		0	1	2
Kurtosis	-1.5	5.36		
	0.0	5.00	5.10	5.20
	2.0	4.52	4.62	4.72

The normal distribution has a Skewness and Kurtosis of 0, 0, the probability is 5%. Probability values can be seen to diverge from normal depending on the level of Skewness or Kurtosis.

Because of this it was decided that in reporting significance exact figures would be given so that marginal p-values could be easily identified. Given the figures provided by Davis (1956) caution should be exercised with interpretation resulting from marginal p-values.

Appendix G: Study 2 Experiment Documents

This appendix gives details of the documents that were used to conduct Study 2. The documents can be found on the accompanying CD: /study-2_ documents/.

Session Protocol

session-1_protocol.doc

session-2_protocol.doc

Consent Form

consent-form.doc

Demographics Questionnaire

demographics-questionnaire.doc

Session Introduction

session-1_introduction.doc

session-2_introduction.doc

Introduction to Curvature

intro-curavature.doc

NASA-TLX Workload & User Experience Questionnaire

NASA-TLX_UE-questionnaire.doc

Appendix H: Eowave Sensor Data Sheet

This is reproduced from information available at www.eowave.com.

Eowave Position Sensors are resistive ribbons. The device outputs the voltage corresponding to the position with 2,5kΩ/100mm. This technology allows an infinite accuracy.

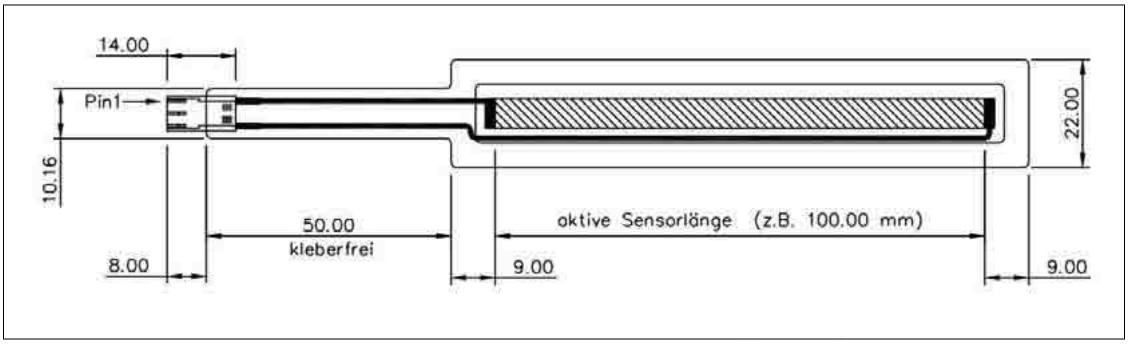


Figure G-1: Eowave Sensor (10cm length)

Technical specifications:

Resistive field: 2,5kΩ/100mm

Tolerance: +/-30%

Average pressure: 1 to 3 N

Max. sensor thickness: 0,5mm

Response time: 0,1m/s

Accuracy: infinite

Temperature: from -25° to 70°C

Ultra-low noise

Size: 22 mm wide

Weight: 5 g

Compatible Eobody1 & 2:

Wiring: 1 x 6.35 mm TRS jack, 2 m-long cable

Compatible Eobody2 Wireless System:

Wiring: 3 pins

Output format: Continuous 0 to 127 (7 bit), 4095 (12 bit)

Power: Built-in phantom 5V DC

Appendix I: Study 3 Experiment Documents

This appendix gives details of the documents that were used to conduct Study 3. The documents can be found on the accompanying CD: /study-3_ documents/.

Consent Form

consent-form.doc

Demographic Questionnaire

demomgraphic-questionnaire.doc

Session Introduction

session-intro.doc

Evaluation Introduction

no-sound-intro.doc

curve-shape-intro.doc

curvature-intro.doc

'More' or 'Less' Curved? Explanation and Comprehension Test

More-less-curved.doc

Workload Assessment

workload-assessment.doc

User Experience Questionnaire

UE-questionnaire.doc

Appendix J: Regression and Psychophysical Data (Study 3)

No Sound

Table I-1: Stimulus D

P-ID	Gradient	Model Intercept (B0)	Slope (B1)	R^2	PSE	CE	Threshold Upper (DLu)	Lower (DLl)	JND (DL)	% Change
P1	0.037	-4.70	128.75	0.57	0.037	0.000	0.042	0.031	0.005	14.065
P2	0.037									
P3	0.037	-5.55	151.89	0.79	0.037	0.000	0.041	0.032	0.004	11.922
P4	0.037	-5.25	143.81	0.71	0.037	0.000	0.041	0.032	0.005	12.592
P5	0.037	-5.64	154.35	0.82	0.037	0.000	0.041	0.032	0.004	11.732
P6	0.037	-5.98	163.82	0.92	0.037	0.000	0.041	0.032	0.004	11.054
P7	0.037	-4.21	115.2	0.45	0.037	0.000	0.042	0.031	0.006	15.719
P8	0.037	-3.75	102.71	0.36	0.037	0.000	0.043	0.030	0.007	17.630
			Mean	0.66	0.037	0.000	0.042	0.032	0.005	13.530
			SD	0.20	0.000	0.000	0.001	0.001	0.001	2.406

Table I-2: Stimulus E

P-ID	Gradient	Model Intercept (B0)	Slope (B1)	R^2	PSE	CE	Threshold Upper (DLu)	Lower (DLl)	JND (DL)	% Change
P1	0.111	-5.89	53.16	0.89	0.111	0.000	0.123	0.098	0.013	11.35
P2	0.111	-5.52	49.83	0.78	0.111	0.000	0.124	0.097	0.013	12.11
P3	0.111	-5.58	50.37	0.80	0.111	0.000	0.124	0.097	0.013	11.98
P4	0.111	-5.96	53.84	0.91	0.111	0.000	0.123	0.098	0.012	11.21
P5	0.111	-5.97	53.88	0.91	0.111	0.000	0.123	0.098	0.012	11.20
P6	0.111	-5.79	52.30	0.86	0.111	0.000	0.124	0.098	0.013	11.54
P7	0.111	-5.73	51.72	0.84	0.111	0.000	0.124	0.098	0.013	11.67
P8	0.111	-6.01	54.22	0.93	0.111	0.000	0.123	0.098	0.012	11.13
			Mean	0.87	0.111	0.000	0.124	0.098	0.013	11.53
			SD	0.05	0.000	0.000	0.000	0.000	0.000	0.37

Table I-3: Stimulus F

P-ID	Gradient	Model Intercept (B0)	Slope (B1)	R^2	PSE	CE	Threshold Upper (DLu)	Lower (DLl)	JND (DL)	% Change
P1	0.111	-5.40	48.73	0.75	0.111	0.000	0.125	0.097	0.014	12.39
P2	0.111	-5.58	50.41	0.80	0.111	0.000	0.124	0.097	0.013	11.97
P3	0.111	-6.06	54.71	0.94	0.111	0.000	0.123	0.099	0.012	11.03
P4	0.111	-5.70	51.49	0.83	0.111	0.000	0.124	0.098	0.013	11.72
P5	0.111	-5.84	52.69	0.87	0.111	0.000	0.124	0.098	0.013	11.46
P6	0.111	-5.80	52.39	0.86	0.111	0.000	0.124	0.098	0.013	11.52
P7	0.111	-5.40	48.78	0.75	0.111	0.000	0.125	0.097	0.014	12.37
P8	0.111	-5.65	51.01	0.82	0.111	0.000	0.124	0.098	0.013	11.83
			Mean	0.83	0.111	0.000	0.124	0.098	0.013	11.79
			SD	0.07	0.000	0.000	0.001	0.001	0.001	0.46

Curve Shape

Table I-4: Stimulus D

P-ID	Gradient	Model Intercept (B0)	Slope (B1)	R^2	PSE	CE	Threshold Upper (DLu)	Lower (DLl)	JND (DL)	% Change
P1	0.037	-5.65	154.62	0.82	0.037	0.000	0.041	0.032	0.004	11.71
P2	0.037	-5.20	142.42	0.69	0.037	0.000	0.041	0.032	0.005	12.71
P3	0.037	-5.58	152.81	0.80	0.037	0.000	0.041	0.032	0.004	11.85
P4	0.037	-4.18	131.92	0.60	0.032	-0.005	0.037	0.027	0.005	13.73
P5	0.037	-5.60	153.17	0.80	0.037	0.000	0.041	0.032	0.004	11.82
P6	0.037	-5.45	149.22	0.76	0.037	0.000	0.041	0.032	0.004	12.14
P7	0.037	-5.19	142.02	0.69	0.037	0.000	0.041	0.032	0.005	12.75
P8	0.037	-5.55	151.84	0.79	0.037	0.000	0.041	0.032	0.004	11.93
			Mean	0.74	0.036	-0.001	0.040	0.031	0.005	12.27
			SD	0.08	0.002	0.002	0.002	0.002	0.000	0.73

Table I-5: Stimulus E

P-ID	Gradient	Model Intercept (B0)	Slope (B1)	R^2	PSE	CE	Threshold Upper (DLu)	Lower (DLl)	JND (DL)	% Change
P1	0.111	-5.91	53.36	0.90	0.111	0.000	0.123	0.098	0.013	11.31
P2	0.111	-5.06	45.70	0.66	0.111	0.000	0.125	0.096	0.015	13.21
P3	0.111	-4.88	44.03	0.61	0.111	0.000	0.126	0.096	0.015	13.71
P4	0.111	-4.52	40.84	0.53	0.111	0.000	0.127	0.094	0.016	14.78
P5	0.111	-5.56	50.19	0.79	0.111	0.000	0.124	0.097	0.013	12.03
P6	0.111	-5.88	53.05	0.89	0.111	0.000	0.123	0.098	0.013	11.38
P7	0.111	-5.29	47.71	0.72	0.111	0.000	0.125	0.097	0.014	12.65
P8	0.111	-4.69	42.35	0.56	0.111	0.000	0.127	0.095	0.016	14.25
			Mean	0.71	0.111	0.000	0.125	0.096	0.014	12.91
			SD	0.14	0.000	0.000	0.001	0.001	0.001	1.30

Table I-6: Stimulus F

P-ID	Gradient	Model Intercept (B0)	Slope (B1)	R^2	PSE	CE	Threshold Upper (DLu)	Lower (DLl)	JND (DL)	% Change
P1	0.111	-5.17	46.67	0.69	0.111	0.000	0.125	0.096	0.014	12.93
P2	0.111	-5.70	51.48	0.83	0.111	0.000	0.124	0.098	0.013	11.72
P3	0.111	-5.93	53.51	0.90	0.111	0.000	0.123	0.098	0.013	11.28
P4	0.111	-4.48	40.45	0.52	0.111	0.000	0.127	0.094	0.017	14.92
P5	0.111	-5.50	49.60	0.77	0.111	0.000	0.124	0.097	0.014	12.17
P6	0.111	-6.11	55.17	0.96	0.111	0.000	0.123	0.099	0.012	10.94
P7	0.111	-6.05	54.65	0.94	0.111	0.000	0.123	0.099	0.012	11.04
P8	0.111	-5.43	49.04	0.76	0.111	0.000	0.124	0.097	0.014	12.31
			Mean	0.80	0.111	0.000	0.124	0.097	0.014	12.17
			SD	0.15	0.000	0.000	0.001	0.001	0.001	1.31

Curvature (original)

Table I-7: Stimulus D

P-ID	Gradient	Model Intercept (B0)	Model Slope (B1)	R^2	PSE	CE	Threshold Upper (DLu)	Threshold Lower (DLl)	JND (DL)	% Change
P1	0.037	-5.08	139.04	0.66	0.037	0.000	0.041	0.032	0.005	13.02
P2	0.037	-5.08	139.04	0.66	0.037	0.000	0.041	0.032	0.005	13.02
P3	0.037	-5.12	140.24	0.67	0.037	0.000	0.041	0.032	0.005	12.91
P4	0.037	-5.08	139.04	0.66	0.037	0.000	0.041	0.032	0.005	13.02
P5	0.037	-4.88	133.58	0.61	0.037	0.000	0.042	0.032	0.005	13.56
P6	0.037	-5.08	139.04	0.66	0.037	0.000	0.041	0.032	0.005	13.02
P7	0.037	-5.12	140.24	0.67	0.037	0.000	0.041	0.032	0.005	12.91
P8	0.037	-5.08	139.04	0.66	0.037	0.000	0.041	0.032	0.005	13.02
			Mean	0.66	0.037	0.000	0.041	0.032	0.005	13.07
			SD	0.02	0.000	0.000	0.000	0.000	0.000	0.22

Table I-8: Stimulus E

P-ID	Gradient	Model Intercept (B0)	Model Slope (B1)	R^2	PSE	CE	Threshold Upper (DLu)	Threshold Lower (DLl)	JND (DL)	% Change
P1	0.111	-5.58	50.34	0.80	0.111	0.000	0.124	0.097	0.013	11.99
P2	0.111	-5.08	45.85	0.66	0.111	0.000	0.125	0.096	0.015	13.17
P3	0.111	-5.12	46.24	0.67	0.111	0.000	0.125	0.096	0.014	13.05
P4	0.111	-5.08	45.85	0.66	0.111	0.000	0.125	0.096	0.015	13.17
P5	0.111	-5.23	47.20	0.70	0.111	0.000	0.125	0.097	0.014	12.79
P6	0.111	-5.08	45.85	0.66	0.111	0.000	0.125	0.096	0.015	13.17
P7	0.111	-5.08	45.85	0.66	0.111	0.000	0.125	0.096	0.015	13.17
P8	0.111	-5.08	45.85	0.66	0.111	0.000	0.125	0.096	0.015	13.17
			Mean	0.69	0.111	0.000	0.125	0.096	0.014	12.96
			SD	0.05	0.000	0.000	0.000	0.000	0.000	0.41

Table I-9: Stimulus F

P-ID	Gradient	Model Intercept (B0)	Model Slope (B1)	R^2	PSE	CE	Threshold Upper (DLu)	Threshold Lower (DLl)	JND (DL)	% Change
P1	0.111	-5.57	50.25	0.79	0.111	0.000	0.124	0.097	0.013	12.01
P2	0.111	-5.15	46.50	0.68	0.111	0.000	0.125	0.096	0.014	12.98
P3	0.111	-5.07	45.72	0.66	0.111	0.000	0.125	0.096	0.015	13.20
P4	0.111	-5.08	45.85	0.66	0.111	0.000	0.125	0.096	0.015	13.17
P5	0.111	-5.08	45.85	0.66	0.111	0.000	0.125	0.096	0.015	13.17
P6	0.111	-5.07	45.72	0.66	0.111	0.000	0.125	0.096	0.015	13.20
P7	0.111	-5.08	45.85	0.66	0.111	0.000	0.125	0.096	0.015	13.17
P8	0.111	-5.08	45.85	0.66	0.111	0.000	0.125	0.096	0.015	13.17
			Mean	0.68	0.111	0.000	0.125	0.096	0.014	13.01
			SD	0.05	0.000	0.000	0.000	0.000	0.000	0.41

Curvature (adjusted)

Table I-10: Stimulus D

P-ID	Gradient	Model Intercept (B0)	Model Slope (B1)	R^2	PSE	CE	Threshold Upper (DLu)	Threshold Lower (DLl)	JND (DL)	% Change
P1	0.037	-28.78	787.92	1.00	0.037	0.000	0.037	0.036	0.001	2.30
P2	0.037	-28.78	787.92	1.00	0.037	0.000	0.037	0.036	0.001	2.30
P3	0.037	-29.34	803.59	1.00	0.037	0.000	0.037	0.036	0.001	2.25
P4	0.037	-28.78	787.92	1.00	0.037	0.000	0.037	0.036	0.001	2.30
P5	0.037	-27.97	766.26	1.00	0.037	0.000	0.037	0.036	0.001	2.36
P6	0.037	-28.78	787.92	1.00	0.037	0.000	0.037	0.036	0.001	2.30
P7	0.037	-29.37	803.59	1.00	0.037	0.000	0.037	0.036	0.001	2.25
P8	0.037	-28.78	787.92	1.00	0.037	0.000	0.037	0.036	0.001	2.30
			Mean	1.00	0.037	0.000	0.037	0.036	0.001	2.30
			SD	0.00	0.000	0.000	0.000	0.000	0.000	0.03

Table I-11: Stimulus E

P-ID	Gradient	Model Intercept (B0)	Model Slope (B1)	R^2	PSE	CE	Threshold Upper (DLu)	Threshold Lower (DLl)	JND (DL)	% Change
P1	0.111	-14.19	131.26	1.00	0.108	-0.003	0.113	0.103	0.005	4.60
P2	0.111	-28.78	259.8	1.00	0.111	0.000	0.113	0.108	0.003	2.32
P3	0.111	-29.37	264.97	1.00	0.111	0.000	0.113	0.108	0.003	2.28
P4	0.111	-28.78	259.8	1.00	0.111	0.000	0.113	0.108	0.003	2.32
P5	0.111	-29.30	264.28	1.00	0.111	0.000	0.113	0.108	0.003	2.28
P6	0.111	-28.78	259.8	1.00	0.111	0.000	0.113	0.108	0.003	2.32
P7	0.111	-28.78	259.8	1.00	0.111	0.000	0.113	0.108	0.003	2.32
P8	0.111	-28.78	259.8	1.00	0.111	0.000	0.113	0.108	0.003	2.32
			Mean	1.00	0.110	-0.001	0.113	0.108	0.003	2.60
			SD	0.00	0.001	0.001	0.000	0.002	0.001	0.81

Table I-12: Stimulus F

P-ID	Gradient	Model Intercept (B0)	Model Slope (B1)	R^2	PSE	CE	Threshold Upper (DLu)	Threshold Lower (DLl)	JND (DL)	% Change
P1	0.111	-16.692	152.53	1.00	0.109	-0.002	0.114	0.105	0.004	3.96
P2	0.111	-26.968	242.89	1.00	0.111	0.000	0.114	0.108	0.003	2.49
P3	0.111	-29.374	264.97	1.00	0.111	0.000	0.113	0.108	0.003	2.28
P4	0.111	-28.782	259.8	1.00	0.111	0.000	0.113	0.108	0.003	2.32
P5	0.111	-28.782	259.8	1.00	0.111	0.000	0.113	0.108	0.003	2.32
P6	0.111	-29.374	264.97	1.00	0.111	0.000	0.113	0.108	0.003	2.28
P7	0.111	-28.782	259.8	1.00	0.111	0.000	0.113	0.108	0.003	2.32
P8	0.111	-28.782	259.8	1.00	0.111	0.000	0.113	0.108	0.003	2.32
			Mean	1.00	0.111	0.000	0.113	0.108	0.003	2.54
			SD	0.00	0.001	0.001	0.000	0.001	0.001	0.58

:

Appendix K: Performance Characteristics – Correlation Statistics

Table J-1: Correlation between Accuracy and Response Time

	Pearson's r				Spearman's Rho			
	NoS	CS	Curv	ALL	NoS	CS	Curv	ALL
Correlation	-0.08	0.30	-	0.11	-	-	-0.61	-
Sig. (2-tailed)	0.84	0.48	-	0.79	-	-	0.11	-
N	8	8	-	8	-	-	8	-

Table J-2: Correlation between Confidence and Response Time

	Pearson's r				Spearman's Rho			
	NoS	CS	Curv	ALL	NoS	CS	Curv	ALL
Correlation	0.58	0.38	-	0.46	-	-	-0.60	-
Sig. (2-tailed)	0.13	0.35	-	0.25	-	-	0.12	-
N	8	8	-	8	-	-	8	-

Table J-3: Correlation between Confidence and Accuracy

	Pearson's r				Spearman's Rho			
	NoS	CS	Curv	ALL	NoS	CS	Curv	ALL
Correlation	0.57	0.14	-	0.87	-	-	0.90	-
Sig. (2-tailed)	0.14	0.74	-	0.01	-	-	0.00	-
N	8	8	-	8	-	-	8	-

Appendix L: Participants Subjective Views about Conflict of Senses

The table below gives responses to the following questions:

10. When there was ‘no sound’ did you feel that you received the same or conflicting information from the different senses?
11. When there was the ‘curve shape’ sound did you feel that you received the same or conflicting information from the different senses?
12. When there was the ‘curvature’ sound did you feel that you received the same or conflicting information from the different senses?

Table K-1: User Experience Questionnaire (Responses Questions 10-12)

Q-ID	P-ID	Stim-ID	Comment
10	1	D	Eyes would say one thing, touch would say another… tended to believe touch over eyes.
10	2	D	Touch sometimes conflicted with vision and made visual judgement less secure.
10	3	D	I couldn't really feel the curvature so I depended on my vision.
10	5	D	Between vision and touch as sometimes it looked as though one side was more curved when it felt less. This may have been influenced by the lines on the sensors.
10	8	D	Sometimes touch would seem different to vision
10	2	E	Vision info came first, touch appeared to contradict sometimes - went with vision
10	2	E	I mainly used sight as my touch didn't tell me much.
10	8	E	Sometimes it would look more curved than it felt.
10	2	F	Left hand end, with touch, mostly appeared lower = more curvature.
10	3	F	I didn't get as much information from touch as I did from sight.
11	1	D	Sound would not match touch or eye.
11	2	D	Vision sometimes conflicted with sound, went with sound unless ambiguous - then used vision if needed.
11	3	D	Sometimes I would think one side looked more curved than what I could hear.
11	7	D	Conflicting with what feeling. Was in two minds so took longer to work out.
11	1	E	Hearing not same as touch/vision.
11	2	E	Vision came first, sound appeared to contradict sometimes - decision went with vision (rarely).
11	2	E	The pitches at either end did not seem to correspond with what I saw.
11	5	E	At times I could not hear a difference between sounds but could see a difference.
11	6	E	Gave me no indication of curve relied on vision.
11	7	E	The curve shape sound sometimes differed from my initial though based on vision.
11	2	F	Sometimes not sure if got pitch right at end of curve, and sound dominated for me, so important info might have mislead.
11	3	F	Sometimes the sounds (high/low) seemed different to what I saw.
11	5	F	At times the sound varied each time I touched the block.
11	7	F	It was hard to differentiate between sides which had similar curvature. The 'curve shape' sound did not always agree with my first thoughts using sight and touch.
11	8	F	Sometimes it would look more curved than it should.
12	1	D	Sound and touch/eyes not always in agreement tended to go with sound over touch.
12	2	D	Used sound predominantly, but vision occasionally disagreed - went with sound.
12	3	D	I only used hearing.
12	2	E	Did not use touch, relied on sound pitch -effort was in translating pitch difference to answer.
12	2	E	I only used hearing.
12	2	F	Sometimes noticed conflict from visual but went with pitch.
12	3	F	I only really used my hearing, and didn't look at the block too closely.
12	6	F	Looking at the block, I thought it was say more curved but the sound would conflict so I went with the sound.

Appendix M: Regression and Psychophysical Data for Individual Thresholds (Study 3 – Further Analysis)

Table L-1: Stimulus D regression and psychophysical data (gradient 0.037)

P-ID	St	R^2	Intercept (B_0)	Slope (B_1)	PSE	CE	Upper (DL_u)	Lower (DL_l)	JND (DL)	Weber Fraction
P1	0.04	0.57	-4.70	128.75	0.04	0.00	0.04	0.03	0.01	14.06
P2	0.04	0.01	-0.63	-17.38	-0.04	-0.07	-0.08	0.00	-0.04	-98.73
P3	0.04	0.79	-5.55	151.89	0.04	0.00	0.04	0.03	0.00	11.92
P4	0.04	0.71	-5.25	143.81	0.04	0.00	0.04	0.03	0.00	12.59
P5	0.04	0.82	-5.64	154.35	0.04	0.00	0.04	0.03	0.00	11.73
P6	0.04	0.92	-5.98	163.82	0.04	0.00	0.04	0.03	0.00	11.05
P7	0.04	0.45	-4.21	115.20	0.04	0.00	0.04	0.03	0.01	15.72
P8	0.04	0.36	-3.75	102.71	0.04	0.00	0.04	0.03	0.01	17.63
mean		0.66			0.04	0.00	0.04	0.03	0.01	13.53
SD		0.20			0.00	0.00	0.00	0.00	0.00	2.41

Note: P2 has been excluded from calculation of means due to extremely low r-squared

Table L-2: Stimulus E regression and psychophysical data (gradient 0.111)

P-ID	St	R^2	Intercept (B_0)	Slope (B_1)	PSE	CE	Upper (DL_u)	Lower (DL_l)	JND (DL)	Weber Fraction
P1	0.11	0.89	-5.89	53.16	0.11	0.00	0.12	0.10	0.01	11.35
P2	0.11	0.78	-5.52	49.83	0.11	0.00	0.12	0.10	0.01	12.11
P3	0.11	0.80	-5.58	50.37	0.11	0.00	0.12	0.10	0.01	11.98
P4	0.11	0.91	-5.96	53.84	0.11	0.00	0.12	0.10	0.01	11.21
P5	0.11	0.91	-5.97	53.88	0.11	0.00	0.12	0.10	0.01	11.20
P6	0.11	0.86	-5.79	52.30	0.11	0.00	0.12	0.10	0.01	11.54
P7	0.11	0.84	-5.73	51.72	0.11	0.00	0.12	0.10	0.01	11.67
P8	0.11	0.93	-6.01	54.22	0.11	0.00	0.12	0.10	0.01	11.13
mean		0.87			0.11	0.00	0.12	0.10	0.01	11.53
SD		0.05			0.00	0.00	0.00	0.00	0.00	0.37

Table L-3: Stimulus F regression and psychophysical data (gradient 0.111)

P-ID	St	R^2	Intercept (B_0)	Slope (B_1)	PSE	CE	Upper (DL_u)	Lower (DL_l)	JND (DL)	Weber Fraction
P1	0.11	0.75	-5.40	48.73	0.11	0.00	0.12	0.10	0.01	12.39
P2	0.11	0.80	-5.58	50.41	0.11	0.00	0.12	0.10	0.01	11.97
P3	0.11	0.94	-6.06	54.71	0.11	0.00	0.12	0.10	0.01	11.03
P4	0.11	0.83	-5.70	51.49	0.11	0.00	0.12	0.10	0.01	11.72
P5	0.11	0.87	-5.84	52.69	0.11	0.00	0.12	0.10	0.01	11.46
P6	0.11	0.86	-5.80	52.39	0.11	0.00	0.12	0.10	0.01	11.52
P7	0.11	0.75	-5.40	48.78	0.11	0.00	0.12	0.10	0.01	12.37
P8	0.11	0.82	-5.65	51.01	0.11	0.00	0.12	0.10	0.01	11.83
mean		0.83			0.11	0.00	0.12	0.10	0.01	11.79
SD		0.07			0.00	0.00	0.00	0.00	0.00	0.46

Appendix N: ANOVA Weber Fractions (gradient) Stimulus B and C (Study 1)

Modality (1=Haptic, 2=Visual, 3=Visual-Haptic)

Stimulus (1=B, 2=C)

Mauchly's Test of Sphericity(b)

Measure: MEASURE_1

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon(a)		
	Greenhouse-Geisser	Huynh-Feldt	Lower-bound	Greenhouse-Geisser	Huynh-Feldt	Lower-bound	Greenhouse-Geisser
Modality	.455	4.725	2	.094	.647	.732	.500
Stimulus	1.000	.000	0	.	1.000	1.000	1.000
Modality * Stimulus	.637	2.710	2	.258	.733	.880	.500

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.
a May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.
b Design: Intercept Within Subjects Design: Modality+Stimulus+Modality*Stimulus

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Modality	Sphericity Assumed	.611	2	.305	1.379	.284	.165
	Greenhouse-Geisser	.611	1.294	.472	1.379	.283	.165
Error(Modality)	Sphericity Assumed	3.100	14	.221			
	Greenhouse-Geisser	3.100	9.061	.342			
Stimulus	Sphericity Assumed	136.123	1	136.123	248.499	.000	.973
	Greenhouse-Geisser	136.123	1.000	136.123	248.499	.000	.973
Error(Stimulus)	Sphericity Assumed	3.834	7	.548			
	Greenhouse-Geisser	3.834	7.000	.548			
Modality * Stimulus	Sphericity Assumed	.807	2	.403	1.070	.370	.133
	Greenhouse-Geisser	.807	1.467	.550	1.070	.356	.133
Error(Modality*Stimulus)	Sphericity Assumed	5.280	14	.377			
	Greenhouse-Geisser	5.280	10.268	.514			

a Computed using alpha = .05

Tests of Within-Subjects Contrasts

Measure: MEASURE_1

Source	Modality	Stimulus	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Modality	Level 1 vs. Level 3		.499	1	.499	1.553	.253	.182
	Level 2 vs. Level 3		.004	1	.004	.067	.803	.010
Error(Modality)	Level 1 vs. Level 3		2.249	7	.321			
	Level 2 vs. Level 3		.416	7	.059			
Stimulus		Level 1 vs. Level 2	90.749	1	90.749	248.499	.000	.973
Error(Stimulus)		Level 1 vs. Level 2	2.556	7	.365			
Modality * Stimulus	Level 1 vs. Level 3	Level 1 vs. Level 2	1.379	1	1.379	1.264	.298	.153
	Level 2 vs. Level 3	Level 1 vs. Level 2	3.114	1	3.114	1.288	.294	.155
Error(Modality*Stimulus)	Level 1 vs. Level 3	Level 1 vs. Level 2	7.635	7	1.091			
	Level 2 vs. Level 3	Level 1 vs. Level 2	16.919	7	2.417			

a Computed using alpha = .05

Appendix O: The Effect of Trial Group Position on Percentage of Correct Judgements

Background to Data

In Study 1 participants were asked to judge if a comparison stimuli was 'more' or 'less' curved than the standard stimuli. They had to do this 12 times for each of eight comparison stimuli. The number of successful judgments was calculated out of the 12 trails, and this gave a percentage correct figure for each of the eight comparison stimuli (A1, A2, C1, C2, E1, E2, G1, and G2). The data used for this analysis were extracted from the data for stimulus 'C' in the haptic condition.

Data Analysis

The purpose of this analysis was to understand if early trials were less successful than later trials. In other words, did the participant learn from undertaking the earlier trials and so become more successful in later trials. The data collected for the 12 trials were split into three groups; early (trials 1-4), mid (trials 5-8), and late (trials 9-12). This was done for each of the eight comparison stimuli, so that rather than one mean, each comparison stimuli had three means, each corresponding to early, mid, or late trials. A two-way analysis of variance was undertaken on the data to determine the effect of trial group position (early, mid, and late) and comparison stimulus (A1, A2, C1, C2, E1, E2, G1, and G2) on percentage of correct judgements.

Results

The ANOVA revealed that there was no effect of trial group position on percentage correct judgements, $F(2,12)=1.90$, $p=0.19$, partial $\eta^2=0.24$. Post-hoc dependent t-tests, using a Bonferroni adjustment for multiple tests, also showed that there was no significant difference between any of the groups (see Table N-1). The results for each of the trial groups and comparison stimuli is summarised in Table N-2.

Table N1: Repeated-measures t-tests

Trials Compared		Mean Difference	Std. Error	Sig.
1	2	-0.45	3.44	1.00
	3	-4.91	2.72	0.36
2	1	0.45	3.44	1.00
	3	-4.46	2.03	0.21
3	1	4.91	2.72	0.36
	2	4.46	2.03	0.21

Table N-2: Percentage of correct judgments by trial group and comparison stimulus

Comparison Stimulus	Trials							
	1 to 4		5 to 8		9 to 12		All Trials	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
G1	96.43	9.45	100.00	0.00	100.00	0.00	98.81	3.15
E1	82.14	18.90	92.86	18.90	100.00	0.00	91.67	11.79
C1	89.29	13.36	82.14	23.78	92.86	12.20	88.10	6.56
A1	57.14	12.20	53.57	22.49	60.71	24.40	57.14	7.50
A2	64.29	24.40	67.86	23.78	85.71	13.36	72.62	15.75
C2	92.86	12.20	92.86	12.20	85.71	19.67	90.48	10.12
E2	100.00	0.00	96.43	9.45	100.00	0.00	98.81	3.15
G2	96.43	9.45	96.43	9.45	92.86	18.90	95.24	9.45