A Framework for Adapative Capability Profiling

by

Matthew James Bell

A Doctoral Thesis

Submitted in partial fulfilment
of the requirements for the award of

Doctor of Philosophy
of
Loughborough University

6th May 2014

Abstract

This thesis documents research providing improvements in the field of accessibility modelling, which will be of particular interest as computing becomes increasingly ubiquitous. It is argued that a new approach is required that takes into account the dynamic relationship between users, their technology (both hardware and software) and any additional Assistive Technologies (ATs) that may be required. In addition, the approach must find a balance between fidelity and transportability.

A theoretical framework has been developed that is able to represent both users and technology in symmetrical (hierarchical) recursive profiles, using a vocabulary that moves from device-specific to device-agnostic capabilities. The research has resulted in the development of a single unified solution that is able to functionally assess the accessibility of interactions through the use of pattern matching between graph-based profiles. A self-efficacy study was also conducted, which identified the inability of older people to provide the data necessary to drive a system based on the framework. Subsequently, the ethical considerations surrounding the use of automated data collection agents were discussed and a mechanism for representing contextual information was also included. Finally, real user data was collected and processed using a practically implemented prototype to provide an evaluation of the approach.

The thesis represents a contribution through its ability to both: (1) accommodate the collection of data from a wide variety of sources, and (2) support accessibility assessments at varying levels of abstraction in order to identify if/where assistance may be necessary. The resulting approach has contributed to a work-package of the Sus-IT project, under the New Dynamics of Ageing (NDA) programme of research in the UK. It has also been presented to a W3C Research and Development Working Group symposium on User Modelling for Accessibility (UM4A). Finally, dissemination has been taken forward through its inclusion as an invited paper presented during a subsequent parallel session within the 8th International Conference on Universal Access in Human-Computer Interaction.

Keywords: Accessibility, Framework, Modelling, Profiling, Semantics.

Acknowledgements

I would like to dedicate this thesis to *everyone* who pushed, pulled, or held my hand as I travelled along this journey. While the research was born out of a desire to help others, it was only completed thanks to them.

Overview

Abstract ii

Acknowledgements iii

1 Introduction 1

2 Literature Review 6

3 Research Methodology 58

4 Developing an Approach 65

5 Design 94

6 Self-Efficacy Study 125

7 Ethical Considerations 172

8 Implementation 200

9 Evaluation 231

10 Conclusion 264

References 271

A Inference 293

B Self Efficacy Study 299

C Evaluation Study Materials 307

D Research Activity 316

Contents

Abstract ii

Acknowledgements iii

1 Introduction 1
1.1 Motivation . . . 2
1.2 Contributions . . . 2
1.2.1 Aims . . . 3
1.2.2 Contributions . . . 3
1.2.3 Objectives . . . 4
1.3 Thesis Outline . . . 4

2 Literature Review 6
2.1 Defining Accessibility, Usability & Personalisation . . . 6
2.1.1 Initial Concepts/Universal Accessibility . . . 7
2.1.2 WCAG POUR Principles . . . 8
2.1.3 Classification of Disability . . . 10
2.1.4 Modelling Interaction . . . 12
2.1.4.1 The Shannon-Weaver Model of Communication . . . 13
2.1.4.2 Universal Access Reference Model (UARM) . . . 15
2.1.5 Usability . . . 20
2.1.6 Personalisation . . . 22
2.1.7 Summary . . . 23
2.2 People . . . 23
2.2.1 People with Impairments . . . 23
2.2.2 Older People . . . 27
2.2.3 Everyone . . . 29
2.2.4 Summary . . . 30
2.3 Methods of Providing Accessibility . . . 31
2.3.1 Universal Design . . . 31
2.3.2 Standards . . . 32

2.3.3 Low and No Technology Solutions 33
2.3.4 Assistive Technology 35
2.3.5 Adaptation/Adaptive Design 36
2.3.6 Separation of Content and Presentation 38
2.3.7 Summary 38
2.4 Controlling Provision/Personalisation/Assistance 39
2.4.1 User Responsibility 40
2.4.2 Human Mediation 42
2.4.3 Technology-Mediation (Automation) 43
2.4.4 Summary 44
2.5 Modelling Accessibility for the Provision of Assistance 45
2.5.1 Recent Research – Demonstrating a Need for Interoperability 46
2.5.2 Definitions – Profile/Model/Framework 50
2.5.3 What Models are Required? 51
2.5.3.1 ISO 24751 53
2.5.3.2 ISO 24756 53
2.6 Aims, Objectives and Scope 54
2.6.1 Scope 54
2.6.2 Thesis Aims, Their Derived Contributions and Resulting Objectives 55
2.6.3 Objectives 56
2.6.4 Wider Goals 56

3 Research Methodology 58
3.1 The Overarching Research Approach 58
3.2 Initial Requirements Gathering 60
3.3 The Iterative Design Process 61
3.4 The Evaluation Process 62
3.5 Summary 63

4 Developing an Approach 65
4.1 Standardisation 65
4.1.1 Common Vocabulary (for Comparison) 67
4.1.1.1 Human-Focused Impairments 68
4.1.1.2 Technology-Focused Preferences 69
4.1.1.3 Technology-Focused Tasks 71
4.1.1.4 Human-Focused Capabilities 72
4.1.1.5 Summary 72
4.1.2 External Interfaces 73

4.1.2.1 Semantic Standardisation 73
4.1.2.2 Syntactic Standardisation 77
4.1.2.3 Summary 79
4.1.3 Internal Structure 80
4.2 Abstraction 83
4.2.1 Hierarchical Levels 85
4.2.1.1 Hierarchical Task Analysis 85
4.2.1.2 Fidelity 86
4.2.2 Meronomic (Whole-Part) Relationships 87
4.2.3 Functional Layers 88
4.2.3.1 Ontologies 89
4.2.3.2 Separation of Concerns 90
4.3 The Resulting Approach 91
4.3.1 The Need to Combine Existing Approaches 91
4.3.2 Recursive Hierarchical Structure 91
4.3.2.1 Modelling Communication 92
4.3.2.2 Functional Layers 93
4.3.3 Focus of the Approach 93
4.4 Summary 93

5 Design **94**
5.1 Particles, Atoms and Instances – A Metaphor for Hierarchical Levels and Functional Layers 95
5.1.1 Representing Particles and their Relationships Within the Framework 96
5.1.2 The '*type*' Relationship 97
5.2 The Structural Layer – Providing Basic Matching 97
5.2.1 Actors 98
5.2.1.1 The '*has*' Relationship 98
5.2.2 Modalities 100
5.2.2.1 The '*infers*' Relationship 100
5.2.2.2 Inferencing Functions 100
5.2.3 Capabilities 102
5.2.3.1 The '*inputsVia*' and '*outputsVia*' Relationships . . 102
5.2.3.2 Reusing the '*infers*' Relationship 104
5.2.3.3 The '*can*' Relationship 104
5.2.4 Processing Functions 105
5.2.4.1 The '*convertFrom*' and '*convertTo*' Relationships . 105
5.2.4.2 The '*can*' Relationship 107

5.2.5 Summary 107
5.3 The Data Layer – Describing the Quality of a Match 109
5.3.1 Data Items 110
5.3.1.1 Data Item Construction 110
5.3.1.2 The '*hasData*' Relationship 111
5.3.1.3 Scales of Measurement – The '*scale*' Relationship . 111
5.3.1.4 The '*dataType*' Relationship 113
5.3.1.5 Units of Measurement – The '*units*' Relationship . 114
5.3.2 Translation Functions 114
5.3.2.1 Construction and Storage 116
5.3.2.2 The '*hasTranslationFunction*' Relationship 116
5.4 The Contextual Layer – Appropriateness of a Match 117
5.4.1 Categories of Context 118
5.4.1.1 Subject Context 118
5.4.1.2 Partner Context 119
5.4.1.3 Environmental Context 119
5.4.2 Representing and Attributing Context – The '*hasContext*' Relationship 120
5.4.3 Comparison of Contextual Information 121
5.5 Summary 121
5.5.1 Contribution 122

6 Self-Efficacy Study 125
6.1 Rationale 126
6.1.1 Relation to Thesis Aims 126
6.1.2 Self-Efficacy 126
6.1.2.1 Predictive and Evaluative Self-Efficacy 127
6.1.3 Computer Self-Efficacy 128
6.1.4 The Use of Other Factors to Predict the Self-Efficacy Accuracy 130
6.1.5 Rationale Summary 132
6.2 Methodology 132
6.2.1 Variables 132
6.2.1.1 Participants 132
6.2.1.2 Devices 135
6.2.1.3 Environment 137
6.2.2 Instruments 137
6.2.2.1 Questionnaires 138
6.2.2.2 Objective Assessment Tasks 140
6.2.2.3 The Mini-Games 141

6.2.3 Test Procedure . . . 142
6.2.4 Methodology Summary . . . 143
6.3 Results . . . 144
6.3.1 Differences Resulting From Reduction in Dataset Size . . . 144
6.3.2 Is the Dataset Usable? . . . 144
6.3.2.1 Factors . . . 145
6.3.2.2 Subjective . . . 147
6.3.2.3 Objective . . . 148
6.3.3 Is the Data Representative? . . . 149
6.3.3.1 Technology Inclusion and Age . . . 150
6.3.3.2 Technology Inclusion and Computer Anxiety . . . 151
6.3.3.3 Gender and Technology Inclusion . . . 151
6.3.3.4 Gender and Computer Anxiety . . . 153
6.3.3.5 Gender and Age . . . 153
6.3.3.6 Age and Computer Anxiety . . . 153
6.3.3.7 Summary . . . 156
6.3.4 Can People Predict or Evaluate Ability? . . . 156
6.3.4.1 Touchscreen . . . 158
6.3.4.2 Mouse . . . 161
6.3.4.3 Mousepad . . . 161
6.3.4.4 Keyboard . . . 162
6.3.4.5 Summary . . . 164
6.3.5 How Do Other Factors Affect Prediction Ability? . . . 165
6.3.5.1 The Effect of Factors on Prediction Using Subjective Judgements . . . 165
6.3.5.2 The Effect of Factors on Prediction Using Objective Measures . . . 166
6.3.5.3 The Effect of Factors on the Suitability of Using Subjective Judgements and Objective Measures for Prediction . . . 167
6.4 Limitations . . . 169
6.5 Summary . . . 169

7 Ethical Considerations 172
7.1 Relation to Thesis Aims . . . 172
7.1.1 Creating Trust and Issues with Profile Transparency . . . 173
7.2 Stakeholders . . . 174
7.2.1 Users . . . 174
7.2.2 Accessibility System Providers . . . 176

7.2.3 AT Developers . . . 176
7.2.4 Accessibility Community . . . 177
7.2.5 External Entities . . . 179
7.2.6 Common Issues . . . 179
7.3 Informed Consent . . . 180
7.3.1 Ensuring Users are Informed . . . 180
7.3.1.1 Concerns Relating to Older People . . . 181
7.3.1.2 Providing Information . . . 181
7.3.2 Obtaining Consent . . . 182
7.3.2.1 Opt-In . . . 182
7.3.2.2 Opt-Out . . . 183
7.3.2.3 Continued Consent . . . 184
7.3.2.4 Withholding Consent . . . 184
7.4 Data Handling . . . 185
7.4.1 Collection . . . 185
7.4.2 Storage . . . 187
7.4.2.1 Location . . . 187
7.4.2.2 Time . . . 189
7.4.3 Use . . . 189
7.4.4 Profile Management . . . 191
7.4.4.1 Personal Identity . . . 192
7.4.4.2 Collaborative Working . . . 194
7.4.4.3 Recognising Users . . . 194
7.5 Data Management . . . 195
7.6 Summary . . . 198
7.6.1 Contribution . . . 198
7.6.2 Impact . . . 198

8 Implementation 200
8.1 Modularity . . . 200
8.2 Technologies and Techniques . . . 202
8.2.1 Representing and Storing Data . . . 202
8.2.2 Providing Reasoning Capabilities . . . 206
8.2.3 Dealing with Uncertainty . . . 207
8.3 Functional Modules Required for Implementation . . . 209
8.4 Data Storage . . . 210
8.4.1 RDF Files and Triple Stores . . . 210
8.4.2 Partitioning Graphs . . . 211
8.4.3 Data Storage Within the Prototype . . . 212

8.5 Reasoning . . . 213
8.5.1 Structural Layer Comparisons . . . 214
8.5.1.1 Input/Retrieval . . . 215
8.5.1.2 Comparison . . . 216
8.5.1.3 Output . . . 218
8.5.2 Data Layer Comparisons . . . 218
8.5.2.1 Input/Retrieval . . . 219
8.5.2.2 Comparison . . . 220
8.5.2.3 Output . . . 222
8.5.3 Contextual Layer Comparisons . . . 222
8.5.3.1 Input/Retrieval . . . 223
8.5.3.2 Comparison . . . 224
8.5.3.3 Outputs . . . 226
8.6 Output . . . 226
8.7 Overview of the Prototype Using a Worked Example . . . 227
8.8 Summary . . . 229
8.8.1 Contents . . . 229
8.8.2 Impact . . . 230

9 Evaluation 231
9.1 Approach . . . 231
9.2 The Data Collection Study . . . 233
9.2.1 Data Collection Methodology . . . 234
9.2.2 Data Gathering Mini-Games . . . 235
9.2.2.1 The Audio Mini-Game . . . 235
9.2.2.2 The Visual Mini-Game . . . 236
9.2.2.3 The Motor Mini-Game . . . 236
9.2.3 Higher-Level Evaluation Tasks . . . 238
9.2.4 Assuring the Ontology's Integrity and Representativeness . . 238
9.2.4.1 Representativeness . . . 239
9.2.4.2 Integrity . . . 240
9.2.5 Summary . . . 241
9.3 Evaluation Against the Functional Contributions . . . 241
9.3.1 Representation and Storage . . . 241
9.3.1.1 Differentiating Between Capabilities . . . 242
9.3.1.2 Measurement of Capabilities at Different Hierarchical Levels . . . 245
9.3.2 Supporting Comparison . . . 247
9.3.3 Mediating Between Data from Different Sources . . . 248

9.3.4 Identification of Mismatches 249
9.3.5 Enabling Speculative Augmentation 250
9.3.6 Summary 251
9.4 Evaluation Against Thesis Aims 251
9.4.1 Functional Assessment of Accessibility 252
9.4.1.1 Theoretical Validity 252
9.4.1.2 Evidence From the Practical Evaluation 253
9.4.2 Variability Between and Within Individuals 253
9.4.2.1 Theoretical Validity 253
9.4.2.2 Evidence From the Practical Evaluation 254
9.4.3 Variety Within and Interaction Between Accessibility Barriers 254
9.4.3.1 Theoretical Validity 255
9.4.3.2 Evidence From the Practical Evaluation 255
9.4.4 Variety Within and Interaction Between Accessibility Solutions 256
9.4.4.1 Theoretical Validity 256
9.4.4.2 Evidence From the Practical Evaluation 257
9.4.5 Variety of Agents Producing and Using Data 257
9.4.5.1 Theoretical Validity 258
9.4.5.2 Evidence From the Practical Evaluation 258
9.4.6 Variability of Data Quality 259
9.4.6.1 Theoretical Validity 259
9.4.6.2 Evidence From the Practical Evaluation 260
9.5 Summary 260
9.5.1 Limitations 260
9.5.1.1 "Brute Force" Approach to Data Storage and Comparison Process 260
9.5.2 Dependencies 261
9.5.3 Wider Achievements 262
9.5.4 Impact 263

10 Conclusion 264
10.1 Contributions 264
10.1.1 Individual Chapters 264
10.1.2 Objectives 265
10.1.3 Thesis Contributions 266
10.1.4 Aims 267
10.2 Practical Reuse of the Research 267
10.2.1 The Approach 267

10.2.2 The Implementation . . . 267
10.3 Extensions and Further Research . . . 268
10.3.1 Data Aggregation . . . 268
10.3.2 Additional Functional Layers . . . 268
10.3.3 Context . . . 268
10.3.4 Standardisation . . . 269
10.3.5 Use Within Alternative Domains . . . 269
10.4 Summary . . . 270

References **271**

A Inference **293**
A.1 Deductive (Top-Down) Inference . . . 294
A.2 Bottom-Up Inference . . . 296
A.3 Relevance to the CAP . . . 298

B Self Efficacy Study **299**
B.1 Instruments . . . 299
B.1.1 Self Efficacy Questionnaires . . . 299
B.1.2 Computer Anxiety Questionnaire . . . 300
B.2 Objective Task . . . 301
B.2.1 Recorded Data . . . 301
B.2.1.1 Keyboard Task . . . 302
B.2.1.2 Touchscreen Task . . . 302
B.2.1.3 Mouse and Touchpad Task . . . 302
B.3 Test Procedure . . . 302
B.4 Additional Results Tables and Subjective Measure Histograms . . . 303

C Evaluation Study Materials **307**
C.1 Data Collection Study . . . 307
C.1.1 Selection Criteria . . . 307
C.1.2 Procedure . . . 307
C.1.3 Introductory Interview . . . 308
C.1.4 Mini-Games . . . 308
C.1.5 Participant Instructions . . . 312
C.2 Description of Validity (Messick, 1995) . . . 313
C.2.1 Content Relevance and Representativeness (Content) . . . 314
C.2.2 Substantive Theories, Process Models, and Process Engagement (Substantive) . . . 314

C.2.3 Scoring Models As Reflective of Task and Domain Structure (Structural) 314
C.2.4 Generalisability and the Boundaries of Score Meaning 314
C.2.5 Convergent and Discriminant Correlations With External Variables (Convergent) 315
C.2.6 Consequences As Validity Evidence (Consequential) 315

D Research Activity 316

D.1 Publications and Presentations 316
D.2 Research Activity 317
D.3 Undergraduate Project Co-Supervision 317
D.4 Events Attended 317
D.5 Other Activities 318

List of Figures

2.1 The Shannon-Weaver Model of Communication (Adapted from Shannon & Weaver (1949)). 14
2.2 Functional models depicting a user and a system within the UARM (Adapted from Carter & Fourney (2004b)). 17
2.3 A multichannel representation of interaction (Carter & Fourney, 2004b). 18
2.4 A representation of the Universal Access Reference Model (UARM) identifying its components (from Carter & Fourney (2004b)). 21
2.5 The elements that constitute the Global Public Inclusive Infrastructure (GPII) (Vanderheiden *et al.*, 2013). 48
2.6 The use of a central exchange format within the VUMS Cluster (Biswas *et al.*, 2013). 49

3.1 The use of aims and functional contributions in evaluating the approach. 63

4.1 Interaction Model – A basic model including two actors. 66
4.2 The elements of the interaction model that can be standardised. . . 66
4.3 Two standard vocabularies that can be used to model interaction. . 67
4.4 Accessibility options offered through the BBC's "My Web My Way". 69
4.5 Interaction Model – Focusing on the use of standard external interfaces. 74
4.6 A diagram representing the use of standard interfaces to compare abilities between two actors. 77
4.7 Fixing a communication gap using an Assistive Technology. 78
4.8 An interaction model including three actors. 80
4.9 Interaction Model – Focusing on the use of standard internal structures. 80
4.10 The fixed four level structure of the Common Accessibility Profile (Fourney, 2007). 81
4.11 Combining hierarchical levels and functional layers of abstraction. . 84

4.12 Interaction Model – Focusing on the use of a hierarchical internal structure. . . . 85
4.13 Interaction Model – Focusing on the use of functional layers. 89
4.14 Interaction Model – Focusing on the use of a recursive internal structure 92

5.1 Two methods of graphically representing actor particles and their relationships. . . . 99
5.2 Representing modality particles and their relationships. 101
5.3 Graphical representations of capability particles and their relationships. . . . 103
5.4 The graphical representation of capabilities between and within actors. . . . 104
5.5 A box-based representation of sub-capabilities and their relationships. 105
5.6 The processing function particle and its relationships. 106
5.7 The use of processing functions inside a hearing aid. 108
5.8 The attribution of capabilities to actors at multiple hierarchical level including inheritance. . . . 108
5.9 A channel of communication between two capabilities. 110
5.10 The comparison and measurement of capabilities using different measurement scales. . . . 112
5.11 The use of a translation function within an extended channel of communication. . . . 115
5.12 An interaction model highlighting sources of context. 122
5.13 A generic graph of the particles and relationships. 123

6.1 A diagram describing the effect of self-efficacy on outcome (adapted from Bandura (1978). . . . 127
6.2 Example interfaces of the objective assessment games. 142
6.3 Histogram showing the distribution of participants' ages. 146
6.4 Histograms showing the distribution of results for two factors. . . . 147
6.5 Scatter plot comparing participants' Technology Inclusion and Age. 151
6.6 Box plots comparing participants' Gender and Technology Inclusion. 152
6.7 Box plots comparing gender and computer anxiety. 154
6.8 Box plots describing the Gender and Age of participants. 155
6.9 Scatter-plots displaying Touchscreen Objective measure residuals. . 171

7.1 An example of the hierarchical structuring of profiles using actors and the '*has*' relationship. . . . 193

8.1 Using modular construction to provide separation between data and processes, based on layers of abstraction. 201
8.2 A graphical representation of the example structural layer graph. . 206
8.3 Using confidence intervals to provide a measure of the accuracy of a prediction. 209
8.4 Implementation – The basic modules. 210
8.5 Implementation – The data storage module. 212
8.6 Implementation – Including the structural layer reasoning module. . 214
8.7 Implementation – Including the data layer reasoning module. 219
8.8 Implementation – Including the contextual layer reasoning module. 223
8.9 Visual representations of the comparisons required to match between the contextual layer information of two data items. 225
8.10 Implementation – Including the outputs module. 227
8.11 Worked Example – The graphical user interface of the prototype. . 228
8.12 Worked Example – Using the prototype to compare Actor $\mathbb{A}$ against the sub-actors of Actor $\mathbb{B}$ to predict their accessibility. 228
8.13 Worked Example – Using the prototype to compare Actor $\mathbb{A}$'s ability to interact with each of Actor $\mathbb{B}$'s capabilities. 229

9.1 The use of aims and functional contributions in evaluating the approach. 232
9.2 A high-level representation of the scenario (using the notation developed in chapter 5). 234
9.3 A graphical representation of the profiles created for the the evaluation study in terms of actors, capabilities and modalities. 237
9.4 A 'box and graph'-based interaction model depicting a user and the evaluation task. 242
9.5 A model of the interaction between a user and the evaluation task including an additional hierarchical layer. 243
9.6 An ability-based audiogram. 246
9.7 Relationship Between Contrast and Acuity Capabilities 246

B.1 Predictive Self-Efficacy Judgement Sheet 299
B.2 Computer Anxiety Rating Scale 300
B.3 Example Games 301
B.4 Touchscreen Self-Efficacy Judgement Histograms 304
B.5 Mouse Self-Efficacy Judgement Histograms 304
B.6 Mousepad Self-Efficacy Judgement Histograms 305
B.7 Keyboard Self-Efficacy Judgement Histograms 305

C.1 Visual Mini-Games 310
C.2 Motor Skills Mini-Game 311
C.3 Example Arrangements of the Evaluation Task 311

List of Tables

2.1 Comparison of accessibility features in common operating systems (from Atkinson (2012)) . . . 36

4.1 Standards Demonstrating Vocabularies Based on Different Points of Comparison . . . 68
4.2 Examples of Media (Fourney, 2007) . . . 76

5.1 Examples of Particles, Atoms and Instances Within the Framework 97

6.1 Independent t-tests to Demonstrate Effect of Removing Records . . 145
6.2 Shapiro-Wilk Test of Normality for Factors . . . 146
6.3 Age Profiles of Data Collection Sessions . . . 146
6.4 Comparison Between Predictive and Evaluative Judgements 148
6.5 Correlations Demonstrating the Usefulness of Subjective and Objective Measures for Predictive Performance . . . 157
6.6 Mathematical Models Used in Multiple Linear Regression. 157
6.7 Mathematical Models Predicting the Touchscreen Objective Measure 159
6.8 Linear Regression Modelling of Objective Mousepad Performance Against Subjective and Objective Measures . . . 159
6.9 Mathematical Models Predicting Mouse Objective Measure 162
6.10 Mathematical Models Predicting Mousepad Objective Measure . . . 162
6.11 Mathematical Models Predicting Keyboard Objective Measure . . . 164
6.12 Mathematical Models Evaluating the Effect of Factors on Subjective Measures . . . 166
6.13 Summary of Multiple Linear Regressions Investigating the Effect of Factors on Prediction Using Subjective Judgements . . . 166
6.14 Mathematical Models Evaluating the Effect of Factors on Objective Measures . . . 167
6.15 Summary of Multiple Linear Regressions Investigating the Effect of Factors on Prediction Using Objective Measures . . . 167

6.16 Mathematical Models Evaluating the Suitability of Subjective and Objective Measures for Prediction with Factors Included 168
6.17 Summary of Multiple Linear Regressions Investigating the Suitability of Subjective and Objective Measures for Prediction with Factors Included 168

9.1 Summary of Evaluation Tasks 239
9.2 Definition of Ability-Based Scores 244
9.3 Paired T-Test Results Between Visual Capabilities 245
9.4 Paired T-Test Results Between Different Visual Mini-Games 249

B.1 Records Excluded Due to Data Loss 303
B.2 Changes in Descriptive Statistics as a Result of Removing Records . 306
B.3 Shapiro-Wilk Test of Normality for Subjective Measures 306
B.4 Shapiro-Wilk Test of Normality for Objective Measures 306

Chapter 1

Introduction

This thesis documents research providing improvements in the field of accessibility modelling, which will be of particular interest as computing becomes increasingly ubiquitous. It is argued that a new approach is required that takes into account the dynamic relationship between users, their technology (both hardware and software) and any additional Assistive Technologies (ATs) that may be required. In addition, the approach must find a balance between fidelity and transportability.

A theoretical framework has been developed that is able to represent both users and technology in symmetrical (hierarchical) recursive profiles, using a vocabulary that moves from device-specific to device-agnostic capabilities. The research has resulted in the development of a single unified solution that is able to functionally assess the accessibility of interactions through the use of pattern matching between graph-based profiles. A self-efficacy study was also conducted, which identified the inability of older people to provide the data necessary to drive a system based on the framework. Subsequently, the ethical considerations surrounding the use of automated data collection agents were discussed and a mechanism for representing contextual information was also included. Finally, real user data was collected and processed using a practically implemented prototype to provide an evaluation of the approach.

The thesis represents a contribution through its ability to both: (1) accommodate the collection of data from a wide variety of sources, and (2) support accessibility assessments at varying levels of abstraction in order to identify if/where assistance may be necessary. The resulting approach has contributed to a work-package of the Sus-IT project, under the New Dynamics of Ageing (NDA) programme of research in the UK. It has also been presented to a W3C Research and Development Working Group symposium on User Modelling for Accessibility (UM4A). Finally, dissemination has been taken forward through its inclusion as an invited paper presented during a subsequent parallel session within the 8th International Conference on Universal Access in Human-Computer Interaction.

1.1 Motivation

The accessibility solution that is appropriate for an individual in a given situation may be provided through variations in the behaviour of the user, as well as the choice of device, AT and adaptations used (Sloan *et al.*, 2010). To this end there has been significant interest in user modelling, with profiles being used to allow the identification of both potential accessibility barriers and the solutions that can be used to overcome them. Existing approaches require that a trade-off be made between fidelity and transportability, and as such there is not currently (and likely never will be) a universally accepted format for creating profiles for use in holistic interaction modelling.

The research has been performed in the context of the Sus-IT project – a *multi-disciplinary* investigation of factors that lead to technology disengagement in older people and methods which can be used to sustain their technology use. As such it incorporates an appreciation of the temporally sensitive capabilities of users and devices within different environments. Additionally, diversity observed in the older populations leads to variability in the accuracy with which they are able to provide data.

The success of any accessibility assessment or reasoning process is dependent on the availability of appropriate data. The higher the fidelity of a simulation, the more accurately it is able to represent the person or situation under investigation. This accuracy is however gained at the expense of transportability, resulting in an inability to reuse the simulation to model other situations. Inversely, reducing the fidelity of a simulation increases its transportability, while reducing its ability to accurately represent a particular situation. Both fidelity and transportability are then also negatively impacted as data accuracy is reduced.

An approach is therefore required that balances the conflicting needs of specificity and transportability through the representation of both users and technology at multiple levels of granularity while mediating between data that has been collected from different sources in order.

1.2 Contributions

The goal of this work is to develop a framework that guides the collection, storage and provision of data as described above. The research has contributed to a work-package within a funded project and a number of derivative papers have been included in its list of attributed publications and official outputs (Damodaran, 2014) (further details in Appendix D). Further dissemination is also planned through a participation in a parallel session within Universal Access in Human-

Computer Interaction (http://www.hcii2014.org/uahci).

1.2.1 Aims

The work complements existing research and builds on established theory (such as hierarchical task analysis and communication theory) to provide a compatible approach. As such it aims to develop a novel approach to modelling that combines the following features:

Aim 1: Functional Assessment of Accessibility.

Aim 2: Variability Between and Within Individuals.

Aim 3: Variety Within and Interaction Between Accessibility Barriers.

Aim 4: Variety Within and Interaction Between Accessibility Solutions.

Aim 5: Variety of Agents Producing and Using Data.

Aim 6: Variability of Data Quality.

1.2.2 Contributions

Through achieving the aims the thesis will provide the following contributions:

- Identify existing approaches used to represent people or technology and develop a format that is suitable to store data for use on different platforms and devices.
- Investigate the ability of older people to provide data and develop a mechanism for identifying and describing data from different sources with the aim of understand the confidence with which predictions can be made.
- The format should enable descriptions to be made at different levels of granularity, in order to facilitate comparisons in different contexts.
- Develop an approach for comparing between profiles that is able to quantify the accessibility of an interaction and provide a description of the assistance that is required if a barrier is identified.
- Enabling speculative augmentation to depict different technology configurations or the use of different forms of assistance.

1.2.3 Objectives

In order to focus the research towards achieving the aims, a number of objectives will be used:

- A search of relevant literature will be used to better define the problem and identify existing approaches that provide partial solutions.
- The potential solutions will be combined to create a novel approach from which a design can be developed.
- An investigation of the ability of older people to provide the data required to drive the approach will be conducted.
- The ethical implications of automated data collection will be explored, resulting in the identification of a mechanism to allow mediation between data from different sources.
- The design and data mediation mechanism will be integrated and a prototypical system will be implemented.
- Data collection and evaluation studies will be used to provide data from real users for use within the implementation to evaluate the effectiveness of the approach in solving the problem.

1.3 Thesis Outline

Chapter 2 (Literature Review) examines relevant literature in order to determine the scope of the problem on which the work is based and identify the aims and objectives that the thesis will work towards.

Chapter 3 (Research Methodology) describes the methodology used to achieve the aims and guide the research contained within the thesis.

Chapter 4 (Developing an Approach) examines existing approaches that provide solutions for individual goals in order to develop an approach that is taken forward through the rest of the thesis. Recursive relationships and communication theory are used to model interaction between any actors involved in a communication.

Chapter 5 (Design) describes a theoretical design that is based on the approach and details the use of a series of standard elements, connected together by a series of relationships that allow models of people, technology and various forms of assistance to be created.

Chapter 6 (Self-Efficacy Study) investigates the ability of users to accurately provide the information needed to drive the approach.

Chapter 7 (Ethical Considerations) discusses the ethical issues surrounding the collection and use of user data and their implications on the application of the theoretical design.

Chapter 8 (Implementation) provides examples of technologies and techniques that can be used to allow implementation and describes the development of a prototypical system.

Chapter 9 (Evaluation) discusses the ability of the theoretical and practical approaches to satisfy the goals identified at the start of the thesis.

Chapter 10 (Conclusion) presents the conclusions that can be drawn from this work and discusses the potential for its future use.

Additional material is also provided in the following appendices:

Appendix A (Inference) provides a discussion relating to the use of inferencing.

Appendix B (Self-Efficacy Study Supporting Materials) provides details relating to the study materials, test procedure and raw data upon which the results of chapter 6 are based.

Appendix C (Evaluation Supporting Materials) provides supplementary material to support chapter 9.

Appendix D (Research Activity) details publications and activities that have been generated by and used to support this research.

Chapter 2

Literature Review

This chapter will be used to examine existing literature in order to both build up a justification for the research that will be presented within this thesis. As this thesis aims to provide a contribution within the area of accessibility modelling, section 2.1 begins by examining what accessibility is and how it can be measured. Section 2.2 then moves on to identify three groups of people who may be identified as benefiting from the provision of accessibility and their individual requirement. With the potential users and their reasons for use identified, section 2.3 investigates the *methods* that are used for providing accessibility and section 2.4 follows by describing the different agents that can be used to perform accessibility assessments in order to identify and/or actually *provide* the support that an individual requires.

While the previous listed sections all result in the identification of a series of aims that are used to guide the research performed and detailed within this thesis, section 2.5 instead focuses on existing modelling techniques. Finally section 2.6 will provide a summary of the scope in which this thesis will operate and uses the aims defined throughout the chapter to identify a series of objectives (in the form of functional contributions) that the thesis will aim to deliver.

2.1 Defining Accessibility, Usability & Personalisation

The Oxford English Dictionary, defines accessibility as "[t]he quality or condition of being accessible (*in various senses*)" (OED, Accessed 11/4/14) and is relevant to a wide range of academic disciplines as a research interest. This thesis aims to make a contribution to the field of accessibility. Specifically it will focus on providing improvements to the modelling of interactions between people and technology, in order to identify their accessibility. Even within the field of Human Computer

Interaction (HCI), there is however no single definition of "accessibility" (Petrie & Kheir, 2007). This chapter will therefore build up a working definition which will be used to scope the research described and identify an aim against which it can eventually be evaluated.

2.1.1 Initial Concepts/Universal Accessibility

The ISO 9241 set of related standards deal with the ergonomics of HCI and cover issues relating to both hardware and software. As a set of standards they aim to improve the accessibility of HCI and are primarily aimed at providing guidance to designers and developers. Within the family, ISO 9241-171 (2008) is a standard which specifically deals with the accessibility of software and provides a definition of accessibility as the "usability of a product, service, environment or facility by people with the widest range of capabilities." While ISO 9241-171 (2008) is limited in scope to software, the definition has also been adopted by ISO 9241-20 (2008) and demonstrates the applicability of accessibility to hardware, software and services.

Although the definition relies on an understanding of the term usability (discussed in section 2.1.5) it conveys the intention that accessibility is concerned with accommodating the diversity that is apparent in the population. ISO 9241 recognises the requirements of users with a range of physical, sensory and cognitive abilities, with specific interest in those related to engaging in interactions. Physical abilities are those related to motion and affecting a person's surroundings. Sensory abilities relate to the traditional five senses and taking in information; in reality, sight, sound and touch are given the most attention. Cognitive abilities describe a range of 'unseen' abilities that are related to the functioning of the brain. This demonstrates a tendency for accessibility to be tied to the notion of disability (discussed in section 2.1.3), however the ISO also caters for people with temporary disabilities and the elderly.

The tendency for accessibility to be tied to the notion of disability can be further demonstrated via examination of the publicly curated Wikipedia article describing accessibility[1]. As a barometer of public opinion, the article specifically states that "[t]he concept [of accessibility] often focuses on people with disabilities" before being organised around a number of disability related topics. The specific mention of older people in the ISO should be unnecessary given that they can be described in terms of the abilities already mentioned. This highlights the tendency for accessibility to deal with disabilities in terms of single profound impairments (e.g. sight/hearing) as opposed to a set of dynamic requirements (as

[1] http://en.wikipedia.org/wiki/Accessibility – Accessed 21/03/2014

seen in people with multiple minor disabilities) or non-disability related needs. The complex nature of older people results in the production of specific guidelines for them as a group.

The right of people to access a variety of products and services is protected by legislation in many countries, for example American and Canada (Kovacs Burns & Gordon, 2010), Australia (Basser & Jones, 2002), and the UK (Gooding, 2000). The emphasis is again placed on catering for the specialised needs of people with disabilities. The UK legislation provides details of expectations with regards to the accessibility of education and regulations regarding public transport. Otherwise it identifies a requirement for businesses not to discriminate based on the abilities (or lack thereof) of their customers and employees. Despite the range of abilities or impairments that need to be catered for this demonstrates the second common factor, namely the placing of responsibility on the product or service providers (developer, designer) rather than the user themselves. For the purposes of this thesis, this attempt to provide technology that is accessible to the whole range of abilities is defined as 'Universal Accessibility'. The same approach is taken with the standards; despite acknowledging the needs of users, ISO 9241 is concerned with guidance to ensure that the computer-systems people use are fit for purpose.

2.1.2 WCAG POUR Principles

Accessibility is a growing concern for technology manufacturers and the World Wide Web Consortium (W3C), Web Accessibility Initiative (WAI), Web Content Accessibility Guidelines (WCAG) (W3C, 2008) are widely used as the standard against which online accessibility is measured. WCAG 2.0 is based on four principles:

Perceivability - It should be possible for content to pass from the device to the user (content cannot be invisible to all of their senses).

Operability - The user should be able to operate the interface (it cannot require interaction that the user is unable to perform).

Understandability - This extends the previous two requirements by requiring that the user understand the content they are exposed to and the interactions they must perform (they cannot be beyond the user's understanding).

Robustness - Expands the definition further by requiring that content be rendered reliably as technologies advance (technological evolution should not render the content inaccessible).

The POUR principles—as they are known—are directed towards interaction between people and content or soft-interfaces rather than physical devices. They do however present a fundamental definition of accessibility that can be applied to any piece of technology. Although their application (in terms of guidelines) provides limitations to their usefulness outside of a web context[2] the principles define accessibility in terms of a number of relationships between user, content and technology. The principles identify that accessibility is a multi-directional, multi-layered concept as will be described in the following discussion.

The first principle requires that the user be able to access the content, in other words, the content is transferred to the user. As there is no requirement for understanding at this stage, this principle effectively defines accessibility as a monotonic relationship between a device and a user. The relationship is then measured in terms of the physical and sensory abilities mentioned in section 2.1.1.

The second principle defines the same type of access, but in the opposite direction; rather than the device affecting the user, the user must be able to affect the device. This principle upgrades the previous one in defining accessibility as a uni-directional relationship; although all principles are written from the perspective of the user, accessibility involves information both entering and leaving them.

The third principle adds a layer of comprehension by defining a difference between access and understanding. This principle extends the nature of accessibility, from a physical and sensory concern, to include cognitive abilities. The definition of accessibility is upgraded again through the inclusion of elements that mirror the first two principles; requiring that the user understand both the content they receive and the implications of the actions they use to affect the device.

Finally the fourth principle highlights that the user does not hold a monopoly on requiring the accessibility of their interactions to be considered. Although not explicitly stated, the principle could be understood as a requirement for the device to be able to access and display the content or interface. At a technical level this could be demonstrated in terms of an ability of the system to parse and render a file type, potentially mirroring the third principle by imitating the requirement for understanding. The description of accessibility in purely technical terms is a departure from the person-centred view taken in the previous section and a move towards the functional definition of disability that will be described in the next section.

[2]Described further in section 2.3.2

2.1.3 Classification of Disability

The relationship between accessibility and the notion of disability has already been identified. In a paper which investigated the relationship between accessibility and usability of websites, Petrie & Kheir (2007) defines accessibility problems as problems that are encountered by people with disabilities. This raises the need to define the notion of disability and once again, different definitions exist which can be classified by their philosophical view of the people they define as 'disabled'. Three different view-points will now be discussed and their relationship to the process of providing accessibility will be explored.

The Medical Model

The medical model is named due to its focuses on the deficiencies that a disabled person has and attempts to use medical diagnoses to allow simple categorisation of individuals. The historic popularity of the medical model is due, at least in part, to its ability to attribute disability to an illness or injury which can subsequently be 'fixed' (Parsons, 1975).

The UK Equality Act (HMSO, 2010), states that a person has a disability if they have a physical or mental impairment that has a 'substantial' and 'long-term' negative effect on their ability to perform normal daily activities. In the online guidance published to promote the act[3], the definition is unpacked and clarified. 'Substantial' is classified using the duration of time taken to perform tasks of daily living and 'long-term' is classified as a duration of 12 months or more. The act allows people with progressive conditions to be classified as disabled and people with HIV are a special case, automatically meeting the disability criteria "from the day they are diagnosed". Some conditions are also explicitly excluded from the definition including: addiction, seasonal allergic rhinitis and tendencies to commit various criminal activities.

The use of the qualifiers 'substantial' and 'long-term' is designed to separate 'people with disabilities' from 'the general population' in order to provide them with protection from discrimination. For the same reason exclusions are also included to limit the scope of the act and prevent its misuse. Ultimately, the definition is provided for use in legal proceedings and as such should be precise, however, the use of qualifiers does leave the definition open to subjective interpretation.

Amundson (2000) criticises the medical model for its negative stance on disability; which is described in relation to a prescribed notion of normality. The model defines disability in terms of a series of conditions that have a negative ef-

[3] https://www.gov.uk/definition-of-disability-under-equality-act-2010 accessed 20/08/2013

fect on a person's existence by imposing limitations or restrictions. This results in assistance being viewed as a means of "curing" the individual and by identifying disabled people in terms of their disabilities, the person is viewed as intrinsically substandard and there are fears that this categorisation will lead to prejudice.

The Social Model

In response to these criticisms the social model of disability takes an opposing view to the medical model. Burchardt (2004) acknowledges that there have been a number of different versions, but they all agree on a number of points. Rather than focusing on the limitations of the individual, the social model places the focus on the failure of society to provide an environment that enables individuals to achieve their full potential. The shifting of 'blame' from user to device addresses the criticisms of the medical model but results in widening the definition of a disability dramatically.

Under the social model an impairment is a condition of the body or mind which, if not taken into consideration may result in a disability. A disability is then a situation resulting in the loss of opportunity caused by an impairment. This distinction results in distancing the negativity attached to disability from an individual that has an impairment. Although an impairment may exist, it will not necessarily result in disability.

Returning to the Equality Act (HMSO, 2010), despite its medical classification of disability, the spirit of the Act as a whole is in line with the social model of disability. The Act requires that reasonable provision is made for people with disabilities to avoid discrimination. By providing additional support, disabled people are able to access the same level of service as non-disabled people, effectively negating their impairments.

Under the social model, disability is not equated to personal tragedy and provisions to minimise the disability caused by an impairment are expected as a basic right (Mitra, 2006). Taking the social model to its extreme, any situation in which a person is unable to perform an action is not only classified as disability, but the responsibility for finding a solution is taken away from the individual. For example, when browsing websites from a foreign country, an inability to understand the local language results in a disability when trying to decipher the content. Rather than suggesting that some attempt should be made to prepare for the trip abroad, the social model advocates that all websites should have multi-lingual provisions for foreign visitors. This situation is clearly infeasible and as such there is a need for an approach that takes the middle-ground.

The Functional Model

The medical and social models present opposing views of accessibility by defining disability in terms of the deficiencies of either the user or their device/environment. After its creation in 1980 the International Classification of Impairments, Disabilities, and Handicaps (ICIDH) was a well cited example of widespread use of the medical model (Bickenbach *et al.*, 1999). In 2001 it was revised and the International Classification of Functioning, Disability and Health (icf, 2001) was approved. The ICF provides a functional approach to assessing disability by focusing on the outcomes (or functionality) that a person can achieve. Disability is defined in terms of the health status of an individual and the activities that they are able to accomplish, given a series of environmental conditions, and is assessed based on four categories:

1. Body functions.
2. Body structures.
3. Activities and participation.
4. Environmental factors.

Unlike the medical and social models, the functional model is not constrained to an examination that focuses solely on the needs of the user or the affordances of the device. Rather it assesses accessibility in terms of the match that is observed between their respective capabilities and as a result is also referred to as the capability approach (Mitra, 2006).

Mitra (2006) identifies that the advantages of the approach stem from its ability to model disability at multiple levels. At the "capability level" impairments are identified and the issues that they may cause can be established, as per the medical model. Impairments can then be used to identify potential disabilities that may arise as a result of interactions, as per the social model. Finally, at the "functional level" disabilities can be understood in terms of their impact on actual interaction. For this reason, a functional approach will be used in the assessment of accessibility; identifying the potential for interactions to take place rather than the deficiencies of either of the participants in the interaction.

2.1.4 Modelling Interaction

So far the term accessibility has been scoped in terms of its relationship to disability, and the resulting approaches to assessment that arise as a consequence of this. The functional approach focused on the interaction between a person and a task, with disability being a result of mismatch between them. This thesis is concerned with interactions between people and technology and this subsection will now explore the modelling of interaction as a process in order to develop a better understanding of its use in the assessment of accessibility.

2.1.4.1 The Shannon-Weaver Model of Communication

The Shannon-Weaver Model of Communication (Shannon & Weaver, 1949) is targeted at the 'physical' act of communicating. It describes how information travels from an information source to a destination via a "channel of communication" that is formed between a transmitter and a receiver. While it consists of three stages, for the purposes of this discussion the two outer stages have been expanded to emphasise the cognitive aspects of the process, as seen in figure 2.1:

Encoding: Messages start as concepts within the transmitter, this is the only place they can be truly understood, surrounded by the knowledge and experiences that gave rise to them. As the concept itself cannot leave the transmitter, it is encoded into a form that can be externalised.

Transmitting: Once encoded the message is able to leave the transmitter via the use of skills that move it to an external medium. The choice of medium may be dictated by the encoding and it is possible that the message may require multiple channels.

Travelling over a medium: Assuming that an appropriately encoded message has been well transmitted it is possible that it may never reach the receiver, or that when it does it has been significantly altered. In this stage of the channel the message is travelling through the real world and is subject to interference or noise.

Receiving: If the message does reach the receiver it is observed using perceptual skills that mirror those used in the transmitting stage. The skills internalise the message, moving it from the external medium to inside the receiver.

Decoding: Decoding is the inverse of encoding and involves trying to interpret the message to understand the concept that was being communicated.

Chandler (1994) criticises the model for its over-simplification of the communication process. The use of the "conduit metaphor" reduces communication to a stereotypical, mathematical model where in reality it is a complex process. By only recognising the flow of information in a single direction, the model cannot cope with a bi-directional interaction and while feedback loops were introduced

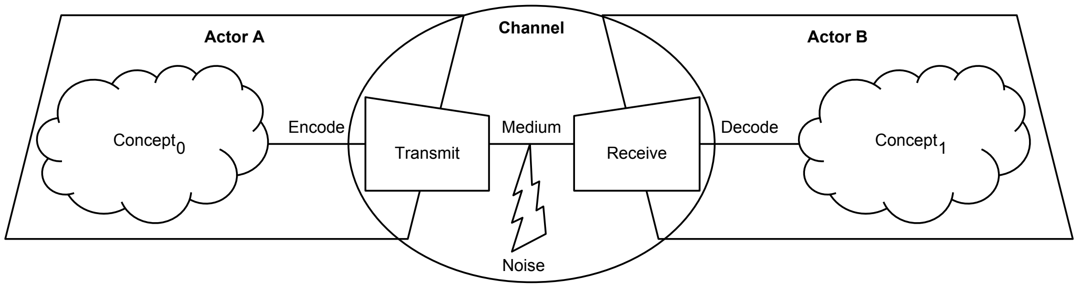


Figure 2.1: The Shannon-Weaver Model of Communication (Adapted from Shannon & Weaver (1949)).

by later theorists, the model remains linear. Weaver (1949) identifies three levels of communication problems:

Technical: How accurately the message can be transmitted.

Semantic: How precisely the meaning can be conveyed.

Effectiveness: How the received meaning affects behaviour.

The Shannon-Weaver model assumes that by addressing technical problems, accessibility will automatically follow at other levels. This may not be the case however as there is also no explicit mention of context other than an appreciation of the existence of "noise" that can degrade the message. Chandler (1994) on the other hand approaches communication from a social perspective and recognises the additional information that is implicitly provided by the different stages of the process. Not only is the context in which the message originates essential to understanding the underlying concept, but the method of transmission, the medium chosen and even the relationship between transmitter and receiver can all provide additional information to aid understanding.

Despite its simplicity, the Shannon-Weaver model does provide a method of describing communication and has been used as the basis for technical solutions to be developed (Carter & Fourney, 2004b; Iglesias-Perez, 2010). Its simplicity and generality have allowed it to be applied to a number of applications within the domain of Information Theory. By modelling communication in the form of a process it is possible to identify the different components and analyse them separately. For example, by treating encoding and decoding as separate processes to transmitting and receiving, an understanding of the context of a message can be represented and some of the above concerns addressed; further discussion on this point will be seen in section 2.1.4.2.

Additionally, by representing communication as a linear process deficiencies early in the process have a demonstrable effect on later stages by reducing the quality of the message that is delivered. This will either result in a reduced quality of communication or require increased effort in later stages of the process to compensate. There is therefore a need to start with quality content to ensure that a quality experience is provided. Burzagli *et al.* (2009) demonstrates this principle through the "Design for All" philosophy which will be described further in section 2.3.1. As an example, web content meeting WCAG 2.0 standards will render reliably across different browsers; the correct encoding of information using a standard format, allows simple decoding later on.

As each of the stages can be quantified, 'weak links' in the process can be identified and targeted solutions used to improve overall performance. The symmetry seen in the model allows solutions to be provided both to the 'weak link' itself and other stages of the process that can be used to compensate. For example, although every stage needs to be successful, when working with content intended for users with cognitive disabilities (reducing abilities at the decoding stage) solutions are often targeted at the encoding stage (e.g. replacing words with pictures or removing unnecessary detail) (Bohman & Anderson, 2005). Alternatively, problems with a user's fine motor skills (as seen in users with Parkinson's Disease) can be reduced through solutions that tailor the sensitivity of their input device (Trewin *et al.*, 2006).

2.1.4.2 Universal Access Reference Model (UARM)

The Universal Access Reference Model (UARM) (Carter & Fourney, 2004b) was developed as a theoretical approach to validating ISO 16071 (the precursor to ISO 9241-171 (2008)). As a self-proclaimed "reference model", the UARM takes an impartial view of accessibility, illustrating its major functions and relations without being tied to any particular design philosophy. Communication is viewed as the flow of information from transmitters to receivers, as it is based on Shannon & Weaver (1949), however multiple transmitters and receivers are used to identify multiple channels of communication. This facilitates the modelling of bi-directional communication, i.e. interaction.

Within the UARM, the metaphor chosen for accessibility is that of a valve. In the same way that a valve can be used to constrict the flow of water through a pipe, different factors can affect the accessibility of an interaction; reducing the capacity of the channel to transfer information and the quality of the resulting interaction. The UARM is constructed using six components (seen in figure 2.4) that allow interaction between a person and a computer system to be modelled.

Users: Users are people who wish to use the system and "Universal Accessibility" has previously been described as "the usability of a single product to as wide a range of people as possible". Although the reference model is intended to describe accessibility for a wide range of users, it does this by acknowledging the heterogeneity (individuality) in individuals. The acknowledgement that users are individuals with needs that change over time (Benyon & Murray, 1993) is central to the UARM's ability to deal with the variety found in the population. Rather than producing advice on the creation of accessible systems or designing for generic user groups, the reference model requires knowledge about the personal needs and abilities of each user. This is also in line with a general trend for use of personalisation that is discussed in section 2.1.6.

Users are described as the focal point of the UARM and this mirrors the way that accessibility assessments often revolve around identifying and meeting people's needs. This results in accessibility being seen as a burden when approached in terms of the medical model of disability. However, as the UARM takes a functional approach it simply acknowledges that the user has a set of abilities which they can use to interact with the world around them. These abilities are exposed via the user's "interface"; the conceptual line distinguishing between the user and the rest of the world.

In defining the interface, the UARM identifies the point at which the user interacts with the world around them. Everything on the user's side of that point can be attributed to the user and used in the creation of a model describing their ability to take part in interactions. Anything on the other side of the point is external to the user and the user's ability to interact with it can be assessed, providing a measure of their accessibility. The UARM identifies four components of users:

Interface – The boundary between the user and the outside world. The interface is responsible for creating, managing and selecting channels over which interaction will take place.

Interaction Abilities and Skills – Discrete (low-level) abilities that allow users to transmit or receive information via the interface. Users have a large number and variety of these skills and their presence dictates what forms of interaction the user can successfully participate in.

Task Abilities and Skills – More complex (higher-level) abilities that allow the user to make sense of interactions. Tasks are meaningful activities that may require one or more specific interaction abilities.

Personal Preferences – Where there are multiple interaction abilities that can be used to complete a task, preferences may dictate which is chosen.

Rather than three discrete components, the user can be viewed as a continuum, as demonstrated on the left of figure 2.2. As described already, the interface is the point at which the user interacts with the outside world and interactions are based on abilities and skills which can be combined to achieve more complex activities, which are governed by preferences. Together they describe a user's ability to interact with any situation in terms of the internal flow of information from brain to interface.

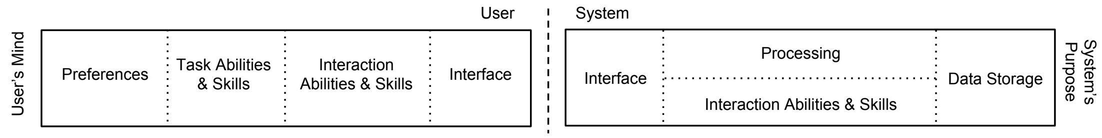


Figure 2.2: Functional models depicting a user and a system within the UARM (Adapted from Carter & Fourney (2004b)).

System: Within the context of ISO 16701, the accessibility of interactions between users and software-based human-computer interfaces are considered. The UARM however accepts that user-software interaction is dependent on the hardware through which the software is accessed and as such considers "computer systems" which can be composed of both hardware and software. The ISO also placed responsibility for ensuring accessibility on the software – in line with the social model of disability. Once again, due to the functional approach taken by the UARM, it does not hold the system responsible for the ability of the user to use it, simply identifying what the system's interaction capabilities are.

Like the user, information about a computer system can be stored in a model and the UARM identifies four parts of a system:

Interface: The boundary between the system and the outside world and responsible for managing its interaction.

Interaction Components: Include the interaction styles and media that the system can provide.

Processing: Represents the system's internal logical processes which determine how the interaction components function.

Data Storage: Provides the data on which the system makes decisions.

As with the user's model, the right half of figure 2.2 demonstrates how a system's elements form a continuum that runs from the system's purpose through to its interactions with the outside world (although interaction components and processing are considered to be equivalent).

Interaction: Within the UARM, interactions are considered the "tangible components of a task" (Carter & Fourney, 2004b). Tasks may require a series of interactions, which take place in either of two directions (user-to-systems or system-to-user). As a result, multiple channels may be constructed between the user and the system as seen in figure 2.3. Each of the channels represents communication via a different modality and has a valve attached to it that can be adjusted separately. The valves represent the ability for interaction to be restricted through the reduction in (or lack of) compatible capabilities.

The multimodal nature of interaction is well researched, both as a method for adding more natural human communication channels into HCI and as a means of delivering accessibility (Obrenovic *et al.*, 2007; Edwards, 2002). By being able to deliver content in multiple modalities it is possible for content to be rerouted where disabilities are observed. Similar to the UARM, Obrenovic *et al.* (2007) proposes a "Unified Framework" which models interaction between users and user interfaces in terms of "effects" (messages) that are sent and received via different modalities. A number of accessibility issues are identified (including: user, device, environment and social constraints) and each of the issues has the potential to either block or filter the effects. The framework is intended to provide developers with a better understanding of the multimodal design and helps to identify the potential for messages to be rerouted via alternative modalities.

However, as highlighted in Edwards (2002) different modalities have different qualities, which often result in their use for communicating different types of information. An audio-based channel for example will be able to convey a range of subtle contextual information through the use of tone and inflection in addition to the actual content of a text-based message. Mapping the information to a visual channel can result in the loss of the contextual information, however there is the potential for this to be conveyed through the use of formatting. More complex, non-linear information (as found in many graphical user interface components) is harder to translate for transmission via an audio channel.

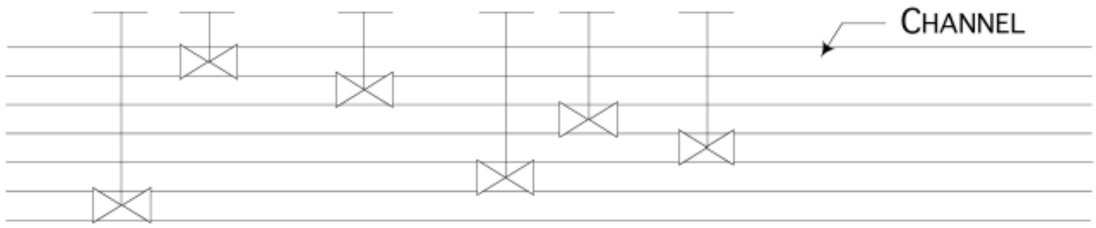


Figure 2.3: A multichannel representation of interaction (Carter & Fourney, 2004b).

Handicaps: In the UARM, handicaps are the name given to anything that interferes with the accessibility of interaction and can be experienced by disabled and non-disabled people alike. Handicaps are therefore equivalent to disabilities, as described previously[4] however they apply to the interaction rather than the person. The UARM identifies two spectra by which handicaps/disabilities may be measured: "degree" and "duration" (Carter & Fourney, 2004b). Returning to the metaphor of a valve, the degree of the handicap dictates the amount to which the channel is constricted and the duration relates to the length of time that the constriction is in place.

Using the functional approach, impairments do not automatically lead to disabilities but are rather the result of a mismatch between the abilities of the user and device required to perform a task in a given environment. This means that the degree to which a channel is constrained is dependent on a number of different factors and a user with a certain impairment may face different handicaps when using different devices, or the same device in different situations. The constriction (or blocking) of a channel affects interaction by reducing the bandwidth available for the transfer of information via that modality (where bandwidth is a measure of the speed/amount of information that can be transferred).

Context: The UARM identifies that both users and systems use context when they encode and interpret information and context and can therefore either enhance or mitigate the effects of handicaps affecting interaction. As an example Carter & Fourney (2004b) describes a situation where a message is created by a system in a particular language and encoded in a particular character-set. If the message is encoded in a character-set that the user does not regularly use (e.g. traditional rather than simplified Chinese) the user is likely to have greater difficulty understanding the message owing to the lack of a shared context. Alternatively, the use of a shared context can also mitigate the effects of a handicap. If for example a user misses part of a message that is delivered via an audio-based channel, they may be able to fill in the gaps through their knowledge of the language used to transmit the message.

Although the word "context" is a useful/descriptive term, it is not however used consistently between different areas of computer science research and can be used to refer to a very broad range of factors (Chen *et al.*, 2000). Contextual awareness is often associated with mobile computing and in a survey of research detailed in Chen *et al.* (2000) four categories of context were identified:

User Context: Examples include: information stored within a user's profile, location and social situation.

[4] Restrictions that are caused as a result of an impairment.

Computing Context: Examples include: network connectivity and nearby resources such as printers.

Physical Context: Examples include: lighting levels, background noise, temperature.

Time Context: Examples include: time of a day, week, month, season (separated as it did not easily fit into the other categories).

While more up-to-date literature has focused on the collection and storage of contextual information (Bettini *et al.*, 2010) the categories are still relevant. Delivery context for example includes the device capabilities and user preference and can be used to guide the adaptation of content presented to that device and is an amalgam of the first two categories. The Composite Capabilities/Preferences Profiles is a recognised W3C specification for describing the "delivery context" of software-based user agents (Klyne *et al.*, n.d.). According to Chen *et al.* (2000), the aim of context-aware computing is either to allow adaptations to be made based on the user's current context (active awareness), or store details of the current context for later retrieval (passive awareness). Within the UARM, the identification of shared context is used to assess the accessibility of interactions.

Environments: "Environments provide additional contexts that focus the user or system on particular portions of their own contexts" (Carter & Fourney, 2004b). Within the UARM the term environment is used to describe both the physical or socio-cultural surroundings of an interaction and provide a source of both context and noise. In this way they either help or hinder communication.

Summary Figure 2.4 provides an overview of the UARM and shows how all of its parts interact with each other. As in the Shannon-Weaver model, the focus is on communication between a user and a system, however this is described in terms of their interactions, which may be bi-directional. The bandwidth of interactions may be reduced through a valve that is dependent on both the environment in which it is taking place and the shared context (A) of the the user and system.

2.1.5 Usability

Whereas the definition of accessibility is dependent on the context in which it is used, the definition of usability is well defined and widely used. The ISO 9241 series of standards define usability 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). Unlike universal

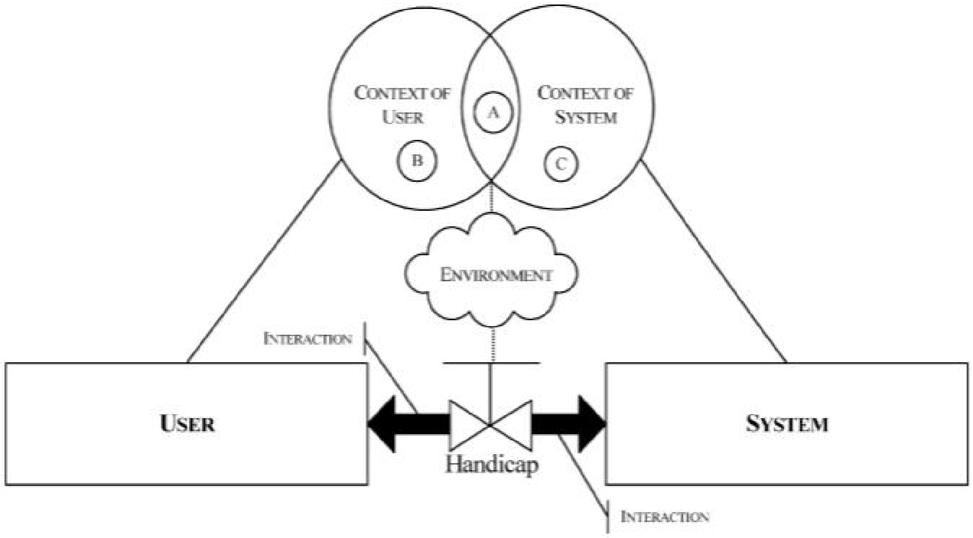


Figure 2.4: A representation of the Universal Access Reference Model (UARM) identifying its components (from Carter & Fourney (2004b)).

accessibility, which has a broad scope (everyone) and low success criterion (some form of usability), usability appears to have a narrower scope (a specific user) and a more descriptive success criterion (described above). There has however been significant discussion on the relationship between the two.

Petrie & Kheir (2007) presents a concise discussion of the potential relationships between accessibility and usability—in an online context—through the organisation of their associated problems into sets. The paper defines accessibility problems as those related to people with disabilities and usability issues are those that affect non-disabled people. By examining the relationship between the sets, the relationship between accessibility and usability was tested against four possible scenarios:

Distinct non-intersecting sets: The situation observed in the development of most websites due to responsibility for each problem set being given to the different individuals, at different points of the development process. Although this view is common amongst the web accessibility community (Thatcher *et al.*, 2006) it is incorrect given the ability for accessibility solutions to assist non-disabled people.[5]

Accessibility as a sub-set of usability: A situation where accessibility issues are dealt with as part of an all encompassing usability assessment. However given the definitions previously provided in terms of disability, this description is incorrect as there are issues that only affect disabled people.

[5] Discussed further in 2.2

Usability as a sub-set of accessibility: This situation is supported by considering accessibility to be a prerequisite to usability, as proposed by Shneiderman (2000). However, considering all usability problems to be within the scope of accessibility clearly contradicts the second portion of the ISO definition developed previously.

Intersecting sets: Petrie & Kheir (2007) settles on the view that accessibility and usability problems form overlapping sets to produce three categories of problem: pure accessibility (affecting disabled people only), pure usability (affecting non-disabled people only) and universal usability (affecting both disabled and non-disabled people).

2.1.6 Personalisation

Personalisation is the process of making adaptations in response to the wants or needs of an individual. As it is often used as a method of providing accessibility or usability, it is not considered to be a peer alongside them. Methods of providing/controlling personalisation will be discussed further in section 2.4 but is included here to illustrate the potential effect of placing personalisation on the same spectrum as accessibility and usability.

Personalisation as a technique is employed for a number of reasons. Firstly the ability to adapt to the needs of an individual make personalisation an ideal candidate for delivering both accessible (Kurniawan *et al.*, 2006) and usable (Sutcliffe *et al.*, 2003) technologies. As one of many potential solutions to the above problems, personalisation intersects both accessibility and usability within the literature[6]. Just as the boundary between accessibility and usability problems is blurred, the point at which personalisation is used purely for vanity rather than to solve a problem is difficult to define.

It follows, therefore, that if accessibility can be modelled in terms of channels of communication with different modalities and personalisation provides a method of creating accessible interfaces, the model could be logically extended to encompass any form of personalisation. For example, Kelly & Belkin (2002) applies personalisation to information retrieval (IR); using a user model to identify search behaviour and personalise IR interaction over a number of devices. While the study had no intention of providing an accessibility related solution, it provides an example of a system that could be considered as providing cognitive assistance, by reducing the requirement for users to remember the location of their files.

[6]As will be demonstrated by the discussion in section 2.3.

2.1.7 Summary

This section has introduced *accessibility*, as both an aim to be strived for and therefore a concept that can be observed and measured. Using the definitions provided thus far, this thesis will contribute to the ideal situation of universal accessibility of computer systems. It is aiming to do this through the identification of accessibility barriers, as it is only after barriers have been identified that appropriate methods of overcoming them can then be proposed.

Accessibility is a measure of the ability of two actors (e.g. user and technology) to interact with each other in a given context. As interaction can be described in terms of a series of channels of communication, accessibility can therefore be measured in terms of the bandwidths of those channels. Accessibility barriers are then either constrictions or absences of channels and a method of identifying them has been proposed.

Aim 1 (Functional Assessment of Accessibility)**.** *Support the functional assessment of accessibility by recognising the benefits provided through its ability to identify accessibility barriers in terms of the assistance required to overcome them.*

With a basic definition and method of assessing accessibility in place, subsequent sections will now discuss the people for whom accessibility is often a problem and the methods by which it can be delivered.

2.2 People

As seen in the previous section, accessibility is a measure of the barriers encountered by people attempting to use technology. This section will now focus on two subsets of the population who are stereotypically described as facing accessibility barriers and discuss the problems that they face. Their experiences will then be compared to the wider "general population" to identify the ubiquitous nature of accessibility.

2.2.1 People with Impairments

Section 2.1 identified that accessibility is often linked to the notion of disability and section 2.1.3 investigated disability further, providing three different perspectives that can be used to guide its definition. While a functional definition has been proposed, this section will begin by considering the impairments that people have. In particular it will focus on single profound impairments of the kind that may be linked to a person's sense of identity. This will allow both the identification of a number of fundamental accessibility barriers, as well as providing examples of accessibility provision at the more acute end of the spectrum.

Shakespeare (1996) discusses the topic of identity both as: (1) an active verb used to describe the process of discovering who disabled people are and (2) a means of declaring membership to a collective or wider group. As already discussed in section 2.1.3 although disability provides a simple method of classification that is based on their disability, there are often negative connotations attached which must be acknowledged and avoided. The latter however, may be used as a means of empowering disabled people through the connections and sense of community it can provide.

Impairments are often considered as falling into one (or more) of four categories when identifying computer related assistance; those affecting visual, hearing, motor and cognitive abilities (Crow, 2008). Categorising impairment in this way provides four relatively discrete sets of disabilities that have their own underlying associated impairments and potential solutions.

Visual Impairments

Visual impairments affect a person's ability to receive visual information and can be broken down into three levels: blindness, low-vision and colour-blindness [7]. Each type represents a different level of severity, both in terms of the impairment itself and the resulting assistance required. Blindness was traditionally noted as the most often cited disability regarding both web accessibility and within the literature, owing to its reliance on graphical presentation of information (Crow, 2008). It is also a very common form of disability with large numbers of people regularly using either glasses or contact lenses.

Blindness implies the total loss of useful vision, however many blind people have some form of residual light sensitivity. As an impairment, blindness precludes the use of any form of visual stimuli, effectively 'closing the valve', completely blocking or not acknowledging the existence of, any channels that transfer graphical information. Support may therefore be needed that diverts graphical information into another medium; often sound, however tactile devices may also be used.

Low-vision is a term that covers a range of sight-related impairments, resulting in a person having difficulty perceiving graphical information. As low-vision does not completely preclude the use of visual stimuli, accessibility can be described in terms of the amount that the bandwidth of the visual channel has been reduced. While assistance may be provided that switches channels, as the person still has a capacity to receive some visual information it may also be provided within the

[7] http://webaim.org/articles/visual

channel (e.g. magnification).

Finally, colour-blindness is an impairment that reduces a person's ability to distinguish between different colours (red/green, blue/yellow or even all colours). While colour-blindness does not reduce visual acuity, it could still be represented as a reduction in the bandwidth of a channel as any information that has been encoded in colour (e.g. a red, amber, green 'traffic light' type system) will not be perceived. Physiologically colour-blindness is the result of a reduction in the number of cones that are available to sense colour[8], which is effectively reduces the bandwidth available. As colour-blindness does not affect visual acuity, assistance that provides magnification will not be appropriate. Rather, support is either required to reduce the dependency on colour as a means of transferring information, or alternatively, algorithms are available to remap colours to those within a person's capabilities (Jefferson & Harvey, 2006).

Hearing

Hearing impairments affect a person's ability to receive information that is transmitted using sound. Similar to visual impairments, audio impairments could be categorised into total deafness, partial deafness and the loss of specific pitches [9]. As with blindness, deafness is the total loss of a person's hearing and results in an inability to receive information that is transmitted audibly. Hearing loss is the audio equivalent of low-vision, where general hearing is reduced to certain degree, ranging from minor to major (and eventually total deafness). Finally, the ability to hear high or low-tone noises may be reduced leading to difficulty hearing sounds at the relevant pitch.

Audio impairments are dealt with in one of two ways. Firstly, where a person has a degree of hearing that they can use, sounds can be changed to fit their appropriate range. This can involve changing the volume, pitch of sounds and the speed at which any spoken language is delivered. Secondly, where there is no available hearing, a change of medium is required with information presented either as visual text or through tactile feedback.

Motor

Unlike the two previous 'sensory' impairments motor impairments often affect a person's ability to transmit information to their device due to a lack of motor

[8]Rather than an impairment, "Tetrachromacy" (Jakab & Wenzel, 2004) is the name given to the condition of having four cones rather than the usual three. While usual in some birds, it is unusual in humans and results in a widening of the visible spectrum to include ultra-violet. Those that have it could therefore be said to have a wider bandwidth with regards to sensing colour than most people.

[9]http://webaim.org/articles/auditory

control. Motor impairments may either constrain a person's ability to move their limbs or make their movements jerky and erratic. As with the previous categories, the level of impairment is dependent on the individual and there are two main methods of providing assistance. Firstly, the sensitivity of many input devices can be altered to make them more tolerant to erratic behaviour, or easier to use by someone with constrained movements. Secondly, an alternative input device can be used, changing the method by which input it provided (e.g. switches, head/mouth sticks, vision tracking or even voice recognition) [10].

Cognitive

Rather than directly affecting a person's ability to transmit or receive information, cognitive impairments affect how it is processed and/or understood. There are a wide range of cognitive impairments and they are often more difficult, both the identify and then provide assistance for. Keates *et al.* (2007) support for this view by acknowledging that there are very few accommodations for cognitive difficulties available in the workplace, in direct contrast to physical difficulties.

Bohman & Anderson (2005) provides a "conceptual framework" that can be used to aid the development of accessibility tools that are intended to benefit users with cognitive disabilities. Rather than providing a simple tick-box list of criteria, the framework includes six lists, which contain different factors that are intended to provide developers with a deeper understanding of the nature of cognitive disabilities and methods by which they can be addressed. The use of a framework as toolbox from which appropriate elements can be chosen, provides a dynamic method of dealing with the diversity that is apparent, given the range of different factors that are identified:

- Categories of functional cognitive disabilities.
- Principles of cognitive disability accessibility.
- Units of Web content analysis.
- Aspects of analysis.
- Realms of responsibility.

Commonalities

While each of the categories identifies a different form of impairment, a number of similarities can be observed between them. All of the impairments can be described as either constricting or completely blocking their associated channel of

[10] http://webaim.org/articles/motor/assistive

communication due to the restrictions they place on a person's abilities to interact. Disability can however be overcome through the use of appropriate assistive technologies that work in one of two ways, both of which serve the same fundamental purpose. Assistive technologies bridge the gap between the abilities of people and technology by either: (1) transforming abilities so that their ranges match up (volume vs hearing level) or (2) translating between abilities in different mediums (e.g. text to speech).

Another important commonality is the existence of similar types of impairments within each classification and that all of the impairments could be plotted on a scale from low to high severity. The higher the severity of an impairment, the greater the level of assistance required. Where an impairment results in a channel being completely blocked (or non-existent), assistance is required that can translate between mediums. Where an impairment constricts a channel, the transformation provided by the required assistance may attempt to maximise the bandwidth of the channel.

2.2.2 Older People

Discussion has recently focused on the provision of accessibility to both disabled and older people. Ageing is a process that is made up of a number of physical and cognitive changes that ultimately result in death, however there is little agreement on the exact age at which a 'person' becomes an 'older person'. As each element of the ageing process may progress at its own speed, using age as a measure of ageing is rarely informative. At the lower end of the scale, the physical and cognitive effects may be beginning to take effect at the age of 45 (Hawthorn, 2000) however many, if not most people at that age would object to being categorised as old. This is understandable as it is acknowledged in the same paper that variance of individuals (e.g. in terms of health) increases in line with age and many changes are reported as being widespread in the 80+ age bracket. This section will not therefore propose a definition of older people in terms of age, preferring instead to use the point at which age-related impairments are manifested in the form of disabilities.

The study of older computer users is gaining popularity according to Wagner *et al.* (2010). A steady increase in the number of article's published in the period 1997—2008 was observed, mirroring similar trends in business and marketing related literature. The increase in the production of literature is coinciding with the increasing age of the population. By 2050 (UN) (2006) predicts that there will be more people in the world over the age of 60 than under 18 and older people are already the fastest growing demographic in many countries. Of further interest is the impending retirement of the 'baby-boomer generation' with research activity predicted to continue in line with the needs of this generation (Wagner *et al.*, 2010).

As a subset of the general population, older people are often treated in the same manner as people with disabilities Newell & Gregor (2002). To some extent this would appear to be a valid approach given the positive correlation between ageing and disability prevalence (Olphert *et al.*, 2005). However older people tend not to self-identify as being disabled (Sloan & Sayago, 2012) and as a result may be less likely to use assistance than their younger disabled counterparts. This is understandable as the ageing process is often characterised by the gradual onset of multiple minor disabilities which may vary on a daily basis (Hanson, 2011).

Hanson (2011) has investigated older computer users from the perspective of comparison with future generations and identifies two characteristics. Firstly in terms of age-related disabilities, a decline is noticeable in many different abilities. Hearing, vision, mobility and dexterity may all be affected and solutions are proposed in terms of various existing assistive technologies. A range of cognitive impairments affecting memory are also identified that produce the significant reductions in ability that are stereotypical of older generations. Many of the coping strategies developed to combat perceptual and physical impairments rely on fluid intelligence. Although older people measure well in tests of crystallised intelligence, fluid intelligence is one of the many abilities that shows a dramatic decrease with age. The result is that many older computer users stick to tasks that are already within their repertoire and have an inability to deal with new situations is instrumental in causing the second characteristic.

The second characteristic identified in Hanson (2011) was related to a lack of experience stemming from reduced exposure to technology. Whereas future generations will be subject to age-related impairments, there is an argument that experience-related issues will be less prevalent due to the increased and constant exposure that younger generations will benefit from. The lack of exposure is well documented, but given the constantly changing nature of technology, the nature of specific experiences is less important than the willingness to engage in new experiences. Older people are less likely to keep up-to-date with current trends for a number of reasons.

Financial barriers are cited, however this contrasts with the view found in Olphert *et al.* (2005) that thanks in part to prudence and the welfare state the current generation of older people is the wealthiest to date. The existence of initiatives to provide technology either free of charge or at very low prices (Sloan & Sayago, 2012) also contradict the idea of financial barriers. If technology is either gifted by well-meaning family or re-purposed from equipment rendered unnecessary through upgrading it is likely to be under-powered and present a substandard experience to its new owner. The substandard nature of the technology provided under these initiatives gives a better insight into the lack of interest that older people show.

Psychological barriers in the form of fear (Olphert *et al.*, 2005) and anxiety (Charness & Boot, 2009)[11] are presented as the main reason for older people avoiding new technology experiences. Even if future generations of older people are unaffected by the issues of fear and computer anxiety due to the security provided by a wealth of previous experience (Hunter *et al.*, 2007), it is argued that a gradual lack of interest in keeping up-to-date will combine with a number of step-changes in technology to create gradual waves of stagnation, leading to eventual abandonment.

2.2.3 Everyone

People from the two groups described above are often stereotypically referred to as facing accessibility barriers. A functional assessment of accessibility can however be used to explain how many of the barriers described above are also shared by the rest of the population. The following discussion will demonstrate how the functional definition used above extends the notion of disability and how the wider population may therefore benefit from the provision of similar accessibility focused assistance.

Just as people from the two groups are 'disabled' by the mismatch between their abilities and those of the technologies they are attempting to interact with, people from the wider population may face barriers due to a similar mismatch. In the first place, the boundary that was defined in section 2.1.3 to separate people with impairments from the rest of the population was in reality an arbitrary one, as demonstrated by the variation in severity that was observed. Secondly, just as older people have multiple minor-to-moderate versions of the impairments described in section 2.1.3, people can be 'disabled' by their environment in similar ways. For example a significant amount of background noise can result in a requirement for the same kinds of solutions (turning up the volume) as a person with a hearing impairment in a quiet environment.

The view that everyone has a disability is supported by Atkinson *et al.* (2008) which asks whether the term 'accessibility' is even appropriate. Rather than extending the definition of the term disability, the paper observes a dilution of the term. It argues that most users with minor disabilities can be provided with an accessible experience simply by invoking and adjusting options that are already available to them. For most users, adjustments to their current operating system

[11] The difference between fear and anxiety is defined in relation to the imminence of the danger.

or program would be sufficient. As these options are available to all users, the distinction between their use as an assistive technology or non-AT personalisation is based on the abilities of the user using them.

Shneiderman (2000) argues that design for extremes, such as those encountered by people with disabilities, can benefit people without disabilities.

ISO 9241 recognises that "usability problems impact all users equally, regardless of ability" Fourney & Carter (2006b).

2.2.4 Summary

This section has discussed two subsets of the population that are regularly identified as facing accessibility barriers and their relationship to the population as a whole. Although they are still the subject of significant research efforts, people who have a single profound impairment are not the only subjects of accessibility research. People with multiple dynamic impairments—as caused by the ageing process—have also attracted attention and accessibility problems may likewise affect people without any recognised impairments at all.

While three overlapping groups of people have been identified and discussed, the people within those groups are not homogeneous and may in reality have a diverse range of abilities and requirements. Rather than trying to match people to pre-existing stereotypes (which are at best approximations), it is therefore preferable to treat them as individuals.

Aim 2 (Variability Between and Within Individuals)**.** *Acknowledge that the variability that exists between different people results in a need for accessibility to be considered on an individual basis. Further, there may also be variability within an individual over time in response to environmental factors, resulting in a need to recognise individuals.*

As identified in section 2.1, disabilities are caused by a mismatch between the abilities of a user, their technology and the environment in which the interaction is taking place. The variety that can be observed between and within individuals' capabilities therefore also leads to variety in the accessibility barriers that they face.

Aim 3 (Variety Within and Interaction Between Accessibility Barriers)**.** *Recognise that people with different abilities (or the same abilities in different environments) will face different accessibility barriers. Additionally recognise that multiple impairments may interact with each other and result in people having multiple conflicting requirements.*

2.3 Methods of Providing Accessibility

With the concept of accessibility having been defined in the previous section, this section will now investigate the methods by which it is delivered. A range of methods exist for providing accessibility, varying both in terms of the point at which they are delivered and the size of the audience at which they are aimed.

2.3.1 Universal Design

"Universal Design" advocates the inherent provision of accessibility or usability in the design phase and the related term "Design for All" expands the definition to include provision for as wide an audience as possible without the need for adaptation Steinfeld (2008). By being aware of user needs early on in the development life-cycle, products are more likely to be both accessible and usable and the requirements of a wider audience can be accommodated.

There are many examples of products initially designed for people with disabilities which have subsequently provided benefits to other user groups. Shneiderman (2000) uses the often used example of dropped kerbs to highlight both the potential for unexpected benefits and the importance of incorporating accessibility at the design phase. While dropped kerbs were initially designed to facilitate the needs of wheelchair users when crossing the road, their use subsequently resulted in benefits for pushchairs, cyclists and travellers with roller bags. In retrospect the commonality between these groups (their use of wheels to facilitate movement) is a functional link. The disability caused by the step-change of a kerb will also affect anybody with an impairment that causes difficulty with tasks related to either physically mounting the curb or perceiving that it is there (resulting in a tripping hazard).

Although designing for extremes can provide improved usability for non-disabled people, it is clearly impractical to incorporate the demands of people that have requirements at extreme opposite ends of the same spectrum into a single design. For this reason adaptation is preferred, where possible, over universal design as will be discussed in section 2.3.5. Universal design is however still a relevant design principle.

The aim of universal design is to provide accessibility "out-of-the-box" without the need for adaptation; which may be costly in terms of money, time or require experience. This should result in reduced effort for the user in terms of meeting any accessibility needs that they have. The benefits that have been described are tied to the notion of the increased effort related to the provision of adaptation. This is best demonstrated in relation to the static nature of many physical products and the practical difficulties of providing dynamic features. By performing adequate requirements analysis, efficiencies developed in the design phase will result in improvements and savings throughout the development process (Keates & Clarkson, 2003).

Additionally, no matter how much adaptability is provided, the functionality will be ineffective if the user does not use it. Adaptation may not be used either due to objection or lack of awareness on the part of the user (Hanson, 2011). In these circumstances universal design can ensure that a minimum level of accessibility is provided, reducing the chances of abandonment due to failures that could be attributed to the inaction of the user.

2.3.2 Standards

"[A] standard is an agreed way of doing something."[12] and standards can be used to provide either guidance that assists or requirements that force the technology to be designed and built in an accessible way. Standards exist to inform the accessibility of both hardware and software and have also been developed to inform the development of profiles that describe a user's accessibility needs (Fourney & Carter, 2006b).

ISO 9241

ISO 9241 has already been introduced in section 2.1 and is a standard that focuses on the ergonomics of computers. It takes a design-for-all approach and provides guidance for ensuring the accessibility of different hardware and software components of a computer system through a series of related standards.

WCAG

As a platform built upon standards (W3C recommendations), the internet would appear to be an ideal candidate for assessing the use of standards for delivering accessibility. Unfortunately the use of the readily available and widely advertised Web Content Accessibility Guidelines (WCAG) (now used within BS 8878 (2010)) is less than encouraging.

Gilbertson & Machin (2012) describes a study conducted on the homepages of 100 UK based development companies. Although accessibility and validation as a skill set was gaining in visibility, the mention of accessibility on the homepage had no impact on the actual accessibility of the site. Out of the 100 homepages checked, only 20 successfully validated for HTML or XHTML and 17 passed validation for CSS. In addition, despite its acceptance as the primary standard for

[12] http://www.bsigroup.com/en-GB/standards/Information-about-standards/what-is-a-standard/

producing accessibility on the internet, the study found that only five pages mentioned WCAG 2.0 on their website. In terms of actual accessibility, a total of eight sites passed WCAG 2.0 Level A compliance (the lowest possible level). The failure of professional developers to widely acknowledge and adhere to standards indicates that there is still work to be done for a standards-based approach to succeed.

Despite their apparent benefits, standards cannot actually guarantee accessibility. Vigo *et al.* (2007) demonstrates the potential for a standards-compliant website to be inaccessible and non-compliant websites to be accessible to specific user groups owing to their use of static guidelines. The study highlighted how the use of the universal accessibility approach proved detrimental to the success of WCAG at delivering accessibility on a personal level. By attempting to provide for all types of users at once, the experience of individual users is impaired. In contrast, it suggested that guidelines should be held in a machine-readable repository and tagged in relation to their applicability to different user groups and situations. This results in the ability to assess a website based on the needs of an individual.

A move from a standards to a repository based approach is also advocated by Loitsch *et al.* (2012) in relation to the storage of preferences describing a user's accessibility needs. The approach combats the static nature of standards by permitting the addition, modification and deletion of individual components. The approach provides a means for experts from different domains to collaborate by contributing their personal knowledge to the central repository. Additionally as individual components can be debated separately, the use of the repository as a whole will not be disrupted by the need to make small changes. It is dependent on adequate curation of the repository and the expertise of contributors, but this is however also true of the standards based approach.

2.3.3 Low and No Technology Solutions

Given that fully universal design is an unreachable ideal, accessibility must either be initiated or configured during (or immediately prior to) use. The propensity for older people to stick with known success strategies over unproven but potentially more successful ones (Eisma *et al.*, 2004; Hanson, 2011) leads to the widespread use of low and no-tech solutions to problems encountered with technology. A combination of improved speed of deployment and reduced effort are the driving factors behind the use of these (sometimes) far from ideal solutions.

Where the duration of a problem is short and its occurrence is infrequent, there may be little perceived need for a long-term solution. A user faced with infrequent areas of small text within an otherwise acceptable document[13] may find that the increased effort taken to activate and subsequently deactivate additional support outweighs the perceived benefits of access to the content. In relation to a paper-based document, the idea of finding a magnifying glass is unlikely to occur to many people when faced with this situation, instead the usual reaction is to move the paper closer, in effect creating a manual zoom. When viewing the document on a piece of immobile technology (e.g. a desktop computer), movement of the user's head towards and away from the screen will produce the same result.

When using a desktop, the potential exists to use a number of technological solutions and the point at which these are employed is the result of a cost-benefit analysis (Keates *et al.*, 2007). The lower the experience of the user, the higher the cost associated with using a technological solution. Similarly, the smaller the degree and duration of the handicap, the lower the cost that can be justified for its solution.

There is however a two-fold danger associated with low- and no-tech solutions. Firstly they can mask the need for more sophisticated assistance. Many accessibility barriers faced due to impairment exhibit gradual onset, with the initially low impact resulting in simple workarounds being successfully employed. The use of workarounds may however lead to impairments becoming invisible and as such the user's potential need for assistance is difficult to identify. As a result alternative assistance is unlikely to be proposed.

Secondly, as the degree or frequency of the impairment increases, the increased effort required for workaround becomes gradually more expensive. Eventually either the workaround will fail or its cost (in terms of time or effort) will become higher than the user is willing to accept. At this point a real solution is required, however the increased degree of the impairment may result in an increased cost of implementation and the cost may outweigh the perceived benefits of access, increasing the likelihood of abandonment.

This situation is a self-fulfilling prophecy, and is similar to the process described in Hanson (2011) whereby a 'step change' in technology requirements leads to abandonment. Although no-tech solutions may have a lower initial costs associated with them, a long-term strategy should revolve around providing the user with a solution that is able to sustain accessibility, by predicting and matching the user's needs. This does not make the use of low and no-tech solutions invalid, however it does highlight both the need to acknowledge their existence and include them in a comparison with of more traditional solutions.

[13] For example footnotes.

2.3.4 Assistive Technology

Cullen *et al.* (2012) is a report on the provision of Assistive Technology (AT) to support assistive living across the life-cycle in Ireland, with reference to other countries. The report identifies ISO 9999 (2011) as providing the most widely used definition of AT. As with accessibility (in section 2.1), AT has been implicitly linked to people with impairments and the ISO provides the following definition:

> "Any product (including devices, equipment, instruments and software), especially produced or generally available, used by or for persons with disability: for participation; to protect, support, train, measure or substitute for body functions/structures and activities; or to prevent impairments, activity limitations or participation restrictions." ISO 9999 (2011)

In practice however the report acknowledges that the functional definition of disability is increasingly being used and this is demonstrated in the broad scope of the standard, which takes a functional approach in its high-level classification. The ISO has several sections devoted to assistive technology used to improve physical access to and use of computers:

- 22.33 Computers and terminals.
- 22.36 Input devices for computers.
- 22.39 Output devices for computers.

In recent years there has been interest in developing more operationally useful classification systems and given the prominence of the WHO ICF, efforts have been made to link the two standards (Cullen *et al.*, 2012). As an example Bougie & Heerkens (2009) provides a mapping between the ISO and the icf (2001), classifying the Assistive Technologies in terms of:

- Assistive products which primarily support a function.
- Assistive products which primarily enable the performance of activities.
- Assistive products which are a prerequisite for participation.
- Assistive products which are primarily used for training.
- Assistive products which are primarily used for measuring (or controlling) functioning or environmental/personal factors.

Table 2.1: Comparison of accessibility features in common operating systems (from Atkinson (2012))

Feature	Mac OS X	GNOME	Windows
Full-screen magnification	I	I	C
Colour deficit support	P	P	C
Resolution and text size	P	I	I
Screenreader	I	I	F/C
Read specific text	I	F	C
On-screen keyboard	C	I	C
Voice recognition	P	C	I

I = Integrated; C = Commercial add-on; F = Free add-on; P = Partial support.

There is frequent mention in ISO 9999 (2011) of specific computer based hardware and software that is used as an assistive technology, a trend supported by the number of different technology based areas identified in Cullen *et al.* (2012). Cullen *et al.* (2012) however provides less information on the specific use of ATs to improve the accessibility of different types of technology. While this suggests that technology use is viewed as a ubiquitous activity there are also many ATs that can be used to support technology use that may also be used to benefit non-technology based tasks.

There are however ATs that have been developed specifically to improve the accessibility of technology, one set of which is software based ATs. A number of these may be designed in to the operating system as shown in table 2.1.

2.3.5 Adaptation/Adaptive Design

The methods of providing accessibility discussed so far could be placed on a spectrum describing their ability to change once implemented. Universal design is static, with the emphasis being placed on provisions during the design. Assistive technologies are able to improve the accessibility of systems that are already in use, however they are external to the system. Adaptive design is the process of designing adaptations into a system, allowing changes to be made during (or before) run-time. It is favoured as a method of providing accessibility over universal design (Benyon, 1993; Shneiderman, 2000; Burzagli *et al.*, 2009).

The AVANTI project (Fink *et al.*, 1996) identified the advantages of using adaptation to tailor websites to the needs of individual users. Adaptive design is preferred over universal design for this reason (Atkinson *et al.*, 2010; Shneiderman, 2000). The project demonstrates the wide range of adaptations that are possible within an interactive system and the uses that they may be put to. The AVANTI project aimed to provide tourist style information about a metropolitan area that could be used by a variety of people. Adaptations were suggested for every aspect of the project and although it was stated that only a sub-section could practically be implemented, the range demonstrates the diversity of uses that adaptations can provide for. Although a categorisation was not proposed, adaptations were described as being available for both the user interface and hypermedia pages. This demonstrates the potential for adaptations to be made at various points in the channel of communication between user and system, mirroring the way that accessibility is not limited to a single aspect.

Given the diversity in range of possible adaptations, there are several ways of classifying them. A functional classification was proposed in Kobsa *et al.* (2001) which distinguished three broad types of adaptations: presentation and modality, structure and content. All three relate primarily to displaying information to the user, which is understandable as they were developed in relation to hypermedia. The classification has been re-used and augmented (Razmerita *et al.*, 2012; Loitsch *et al.*, 2012) and although its domain is restricted, the functional approach mirrors that seen in the UARM.

Benyon & Murray (1993) classified adaptations in terms of different levels of abstraction. Scope can vary from low-level (e.g. size or colour of text), through to high-level (e.g. the reconfiguration of an interface to suit a mobile device). Higher-level adaptations may encompass a number of lower level adaptations. This adds depth to the functional approach by differentiating based on the purpose or scope of the adaptation.

An important distinction that has been made in the literature is between "Adaptability" (user-driven) and "Adaptivity" (system driven) (Findlater & McGrenere, 2004; Oppermann, 1994). This is inconsistent with the use of the term "Adaptable" to imply a general ability to adapt and "Adaptive" to imply a system-driven subset thereof. All four terms will be used with the meanings given here. Although all the terms imply an ability for adaptation to occur, there are significant difference between user and system driven adaptation as seen in the discussion in section 2.4.

Browne *et al.* (1990) proposes a classification based on the level of sophistication of an adaptive system. Each progressive level requires a reduced commitment in terms of designing and evaluation of the resulting interface. This is due to higher-level systems taking responsibility for this themselves.

(Simple) Adaptive Systems – Simple stimulus-response. These systems are limited in scope and adaptations are effectively hard-wired, responding directly to pre-determined stimuli.

Self-Regulating – Provides feedback on the success of adaptations in terms of their effect on the quality of interaction. This may be as a result of inference being performed.

Self-Mediating – Allows adaptations to be trialled before they are given to the user. This allows only useful adaptations to be provided, reducing the burden on the user and increasing the likelihood that adaptations will be accepted.

Self-Modifying – The previous levels rely on static reference sets. Self-Modifying systems can update themselves to allow reasoning to be performed.

2.3.6 Separation of Content and Presentation

The increasing desire for adaptation has led to the use of abstract user interface languages to support accessibility (Trewin *et al.*, 2002). After examining four different languages, the conclusion is however reached that extensions are necessary in order to meet the specific requirements of universal usability. Modelling functionality at a high level allows the generation of interfaces to suit the user, device and context of use placing the burden of providing accessibility on the interface provider.

2.3.7 Summary

This section has demonstrated both the diversity of potential accessibility solutions that are available and the potential for different solutions to be used in combination. The accessibility solution that is appropriate for an individual in a given situation may be provided through variations in the behaviour of the user, as well as the choice of device, AT and adaptations used (Sloan *et al.*, 2010). This means that there may be more than one solution that can be used to fix any particular barrier and as such it should be possible to represent this in the modelling of communication for the provision of assistance.

Aim 4 (Variety Within and Interaction Between Accessibility Solutions)**.** *Recognise that there may be zero, one or more than one potential solution to any single accessibility barrier and different solutions may be provided in different ways. In addition, as solutions may have multiple objectives and components, it is possible for them to be combined.*

2.4 Controlling Provision/Personalisation/Assistance

Although a range of methods are available for providing accessibility, adaptation (which incorporates targeted assistive technologies) has shown the greatest potential for providing the personalisation identified in the section 2.1.7. When describing adaptive interfaces there is a differentiation between user- and system-controlled adaptations. Razmerita *et al.* (2012) uses this differentiation to redefine personalisation (automated adaptations driven by a user profile), contrasting it with customisation (requiring effort on the part of the user).

Four stages were identified in Dieterich *et al.* (1993) regarding the provision of adaptation within user interfaces by either the user or an automated agent. As well as being limited in scope to the provision of adaptive user interfaces, Dieterich *et al.* (1993) considered only the user or the system as potential 'agents', able to take responsibility at each stage. This section is interested in the provision of any of the forms of support and in this wider context the use of a spectrum between user and automated agent is inappropriate. Although automated support is generally agreed as the ideal end-goal (for reasons to be discussed), there has been significant interest demonstrated by the social literature regarding the advantages of assistance provided by people.

Dieterich *et al.* (1993) classifies systems based on their functionality, with regards to the stages of the adaptation process. While this provides information about the placement of responsibility at each stage of the process, it does not describe how each stage is actually carried out. More recently Edlin-White *et al.* (2010)[p20] identifies a general consensus that adaptive systems can be considered to be composed of three processes:

1. **Afference:** Collecting information.
2. **Inference:** Building a model based on the information.
3. **Efference:** Using the model to decide on a course of action.

The processes are used in the description of an adaptive system that incorporates all of the stages described by Dieterich *et al.* (1993), however there is the potential for different stages to be performed by different agents. In this case, rather than splitting the processes amongst the agents, each agent would be required to perform all of the processes for the stage they were carrying out. As an example, if a user were to initiate an adaptation process, they would be doing so based on: (1) the collection of information about their requirements, (2) the comparison of those needs against the abilities of their technology to meet them in order to,(3) make a conscious decision to ask for assistance. If that request was received by an automated adaptation system it would have to (1) understand the nature of the user's problem, (2) compare it against existing forms of assistance and (3) propose a number of alternatives which can either be presented to the user or further compared at the decision stage.

The following sub-sections will discuss the three potential agents (User, Human-mediated, System) and the factors affecting their suitability for controlling different stages of the adaptation process.

2.4.1 User Responsibility

The more stages of the adaptation process that the user is responsible for, the lower their reliance on external assistance. However there are several potential problems that reduce the suitability of this option.

Self-Determination

The main reason to give responsibility to the user, is to provide them with an element of control. One of the central tenets of medical ethics is the right of the individual to self-determination. Although this does not imply that the user should be solely responsible for all stages of the adaptation process, it does recognise the importance of consultation, especially in the initiation and decision phases. One of the most problematic areas in user acceptance of automated assistance is the perception that it will add to, or disrupt the user's workload. As an example, despite its potential advantages, Microsoft's Office Assistant ultimately failed due to its tendency to divert users from their principle goals, which resulted in a reduction in productivity (Schaumburg, 2001).

Self-Efficacy

Although unsolicited support may be met with contempt, there may be a potential problem with the ability of some users to recognise that they require it. Self-efficacy is a measure of a person's belief in their own abilities (Bandura, 1978). As a multi-faceted construct, self-efficacy can be used to predict a user's ability to accurately know when they require assistance and how they are likely to react when assistance is suggested. Self-efficacy can be measured at both a general and domain-specific level. "Computer Self-Efficacy" (Compeau & Higgins, 1995) is a measure of a user's self-efficacy in technology-related contexts. As a constituent of the general measure, specific self-efficacy is both dependent on, and contributes to, a person's general self-efficacy (Marakas *et al.*, 1998). This means that a person with a generally low self-image is likely to doubt their ability in tasks they are yet to try.

Computer self-efficacy has been described as a function of previous IT experiences (Moores & Chang, 2009). Previous poor experiences can result in a low self-efficacy and a tendency to under-estimate or doubt one's abilities; conversely good experiences are more likely to result in increased self-efficacy and potentially lead to over-confidence. In addition to this, inexperience has been linked to both anxiety (Karavidas *et al.*, 2005) which reduces self-efficacy and over-confidence, owing to a lack of understanding with regards to the lack of deficiency that is present (Ballantine *et al.*, 2007).

However the user should not be discounted completely given that, as mentioned in section 2.3.3, the user may be able to observe factors or events that are imperceivable to a computer-based system. The example that has already been given is that of the use of 'no-tech' solutions that may be able to mask the existence of impairments. Similarly, research into recognition of emotions is still in its infancy, even though they have been identified as a potential source of information about the accessibility of a system (Peter & Herbon, 2006; Agarwal & Meyer, 2009).

Lack of Experience

When using adaptable technology there are very few times times when it is favourable for a user to be burdened with the task of obtaining and configuring support themselves, especially after a decision has been made to select a particular adaptation (Dieterich *et al.*, 1993). There are however two notable exceptions. Firstly, when external assistive technologies are used, the potential for automation reduces; unless facilitated by human-mediation, execution has to be performed by the user. Secondly, when it is the intention to teach the user, by completing an action themselves (with appropriate instruction) they are more likely to be able to replicate it if they find themselves in a situation where it cannot be automatically provided.

When considering the proposal and execution stages, the general consensus in the relevant literature points to a lack of experience in knowing what support is available or how to configure it (Trewin, 2000). As mentioned in section 2.2.2, disengagement is linked to the motivation of the user to use effort staying up-to-date with current technology. For example, in one survey of older people in the UK, although a variety of sources had been used to learn about the Internet, none of the non-users were able to name a formal Internet course (Olphert *et al.*, 2005). In addition the belief was held that cost would be a prohibiting factor if a course was found.

2.4.2 Human Mediation

The interest in human mediated assistance within the social literature is due at least in part to its prevalence. As motivated as users may be to keep up to date with available support, there are many benefits from receiving some form of external support. There are a variety of sources of human-mediated assistance:

Family Parents, Children, Grandchildren (Rettie, 2002).

Social Groups Friends, Colleagues (Olphert *et al.*, 2005).

Specialised/Mediated Groups Online Forums, Research Community Groups (Forbes *et al.*, 2009).

Experts Medical Professionals, Helplines, Taught Classes (Olphert *et al.*, 2005).

Experience

Expert evaluation is also the prominent model in the provision of assistive technology in healthcare. Lenker & Paquet (2003) identifies six models that have been developed to measure the success of assistive technology, all of them require an expert in order to facilitate their use. The reliance on experts in the medical domain is due in part to the complex nature of the domain. Medical aids such as hearing aids and glasses require assessment and configuration that the user is unable to undertake themselves.

Older and disabled people both often rely heavily on other people for all four stages of adaptation process. However a 2011 study of Occupational Therapists working in the Mental Health domain demonstrated the need for additional training to be provided with regards to the technology based solutions that are available for those with cognitive impairments Gitlow *et al.* (2011). A tendency was observed for therapists to suggest low or no technology solutions due to a gap in ongoing training, leading to lack of awareness with regards to technology-based assistance. The comparatively low number assistive technologies that are available for cognitive impairments was highlighted in 2.2.1 and the relevance of Gitlow *et al.* (2011) may be limited owing to the articles focus on a specific group. This however highlights the potential for experts to have a similar lack of awareness as users of the latest developments.

In a study investigating the ability of experts and non-experts in a particular method of assessing the accessibility of web-pages, Yesilada *et al.* (2009) unsurprisingly found that experts performed better than non-experts. Experts took less time, rated themselves as more productive and confident, and were more effective at the task with a higher level of reliability than non-experts. Although the example concentrates on a single method of web-page evaluation, it highlights the advantages of seeking expert help in the initiation and decision stages. This is due to their greater abilities to produce appropriate models of user performance and make decisions based on those models.

Motivation

Although not related to the stages of adaptation, human-mediated assistance is preferable over the user or technology controlled methods due to its inherent ability to provide social interaction. Forbes *et al.* (2009) describes the use of a curated social group in Dundee as a means of engaging older people. Although most members of the group had a technology based problem that triggered their attendance, it is the social aspect of the group that is seen as a key driver of its success. The friendly atmosphere motivated learners to continue participating in the group after their initial problem had been solved.

The experiences of the older people in the social group in Dundee were mirrored in the wider population of older people surveyed in Damodaran *et al.* (2013). In answer to a survey question asking about the most important thing to help them use digital technologies successfully, 25% of respondents identified human help and support, i.e. friends, family and tutors. The importance of face-to-face support was also confirmed, with participants being more likely to seek help from a person than any other external source.

Expense

Where human mediation is provided by friends or family members it is unlikely to result in any monetary cost to the individual. All forms of human-mediation do however result in a cost being incurred. Help-lines and tutor led courses both have costs associated with them in terms of labour and equipment and where support is provided by family and friends there is a cost associated in terms of the time they provide. This identifies the potential problem that human-mediators have with the collection of data, as they are more expensive (at the time of use) than the other agents, their ability to provide longitudinal data collection is limited.

2.4.3 Technology-Mediation (Automation)

So far the methods of providing assistance that have been discussed have relied on some form of human intervention, either from the user or a third party. This section will now discuss the use of automated and semi-automated support systems.

In a review of the use of automation in complex systems, Parasuraman (2000) identifies four functions that automation can be applied to: information acquisition, information analysis, decision (action selection) and action implementation. The first three of these map directly to the processes identified at the start of this section. The paper however identifies technological feasibility and cost as the primary criteria for applying automation to a new system; further, automation is only a valid option if there is no detrimental impact to the user. This makes sense given the desire to retain human control in section 2.4.1 and the potential costs identified in section 2.4.2.

Trewin (2000) summarises five reasons that can lead to a lack of configuration of accessibility related features in computer systems, all of which have been identified in the previous sections.

1. Lack of confidence in performing configuration.
2. Lack of knowledge of how to change configuration.
3. Lack of awareness of the available options.
4. Difficulty in identifying what problems exist and therefore the appropriate solution.
5. Lack of control over the unconfigured interface.

Trewin (2000) acknowledges the success of Accessibility Wizards in exposing available options, responding to the first two point and providing a partial solution to the third. Another approach (first proposed by the AVANTI project (Fink *et al.*, 1996)) is to use a questionnaire to gather user requirements with appropriate assistance provided based on user responses. This provides only partial automation, responding only to the first three points and in addition as described in the previous sections, problems still surround the quality of the information that the user provides and the willingness of the user to spend time on configuration. In response to this, points four and five require a proactive configuration method, delivered by way of automation.

2.4.4 Summary

As with previous summaries, the first key message to acknowledge is the variety of agents that are both able to collect and process information in order to identify appropriate accessibility solutions for an individual. Four different stages (initiation, proposal, decision and execution) which can all be accomplished through the use of three processes (afference, inference and efference). Three different agents were then identified (the user themselves, human-mediation and technology-mediation).

Responsibility for each stage of the adaptation process can legitimately be placed on any of the agents depending on the situation. In a similar way, responsibility for the different processes (e.g. adapting different components of the same system) may be shared between different agents, potentially within the system. Whilst the three agents (user, human-mediation and system) use approach the provision in different ways, they all follow the same process of afference, inference and efference.

Aim 5 (Variety of Agents Producing and Using Data)**.** *Recognise that there may be multiple entities that are both providing and wishing to use data. Data should be stored in a format that encourages transportability and is able to deal with the potentially distributed nature of its provision and use.*

As a result of the variety of agents producing and using data, data may be of varying levels of quality and its use should reflect this.

Aim 6 (Variability of Data Quality)**.** *Recognise the issues surrounding the variability of the quality of data and provide a mechanisms for both acknowledging the existence of, and mediating between, data of different levels of quality.*

2.5 Modelling Accessibility for the Provision of Assistance

The previous sections have discussed what accessibility is, the people that face accessibility barriers and how they are affected, the solutions used to achieve or improve accessibility, and the agents that may take responsibility for ensuring that appropriate solutions are identified and used. As a result, a number of aims have been identified that describe the diverse nature of the modern accessibility landscape. Rather than adding to the complexity of the landscape by suggesting the need for additional definitions, barriers, solutions or agents, the aims instead suggest the need for synergies to be identified in order for the complexity of the landscape to be reduced.

Modelling has a wide variety of uses within Human Computer Interaction and the study of accessibility. This section will discuss the use of modelling as a means of identifying and facilitating those synergies. Ritter & Young (2001) provides examples of three ways that cognitive models can been applied to HCI (taken from (John, 1998)).

Predictive models are descriptive in nature, using a corpus of stored data to create a mathematical representation of expected performance. They can then be applied to a situation (e.g. an interface) to make predictions about the amount of time taken to perform selected tasks, or the likely number/form of errors that will occur.

Substitute (replacing) models specify the information needed to perform a task, the way the information is processed and the steps taken to perform the task. This allows the model to be used to simulate user behaviour, acting as a surrogate for testing of new designs.

Assisting models simulate how a task is performed by actually performing it themselves. They can then be used to guide the user, providing tailored support based on their knowledge of the task being attempted.

Ritter & Young (2001) focused on the use of modelling to aid developers during the design stage through the use of substitute models. More recently, there is currently significant interest in the use of all three types of models to provide real-time support (Peissner *et al.*, 2012b). By *predicting* a particular user's behaviour, a *substitute* model can be created to identify the barriers that they are likely to face and *assistance* then provided to improve the accessibility of their interactions.

2.5.1 Recent Research – Demonstrating a Need for Interoperability

There have been two recent projects that have produced research in the areas described above. They both demonstrate a desire to unify the disparate accessibility landscape, and are approaching the problem from different points of view.

Global Public Inclusive Infrastructure

According to Vanderheiden *et al.* (2013) there is a "looming crisis" that may result from a disparity between an increasing requirement for people to have access to ICT while access solutions are either not available or cannot be delivered in a cost effective way. In response, a number of connected projects are attempting to develop a Global Public Inclusive Infrastructure (GPII)[14], which aims "to ensure that everyone who faces accessibility barriers can access and use the Internet". The GPII is an architecture for ensuring universal accessibility and is comprised of a large number of components, of which user modelling is only one element.

As seen in figure 2.5 the Cloud4All project is leading the development of the accessibility profiling and matching activity. The project aims to use user profiles

[14]http://gpii.net/

to transport needs and preferences, in order to facilitate the provision of personalisation and assistive technology across a range of devices. This is achieved through a number of components including (Vanderheiden & Treviranus, 2011a):

Personal Capture Mechanisms: For identifying the preferences that the user has initiated.

Preference Storage Mechanisms: For storing user preferences in the form of a profile, allowing them to be accessed across a range of technology.

Remote Identification Mechanisms: For recognising users and allowing their profiles to be accessed.

Fitter/Matchmaker: To use existing profile data to suggest the adaptations and settings that are appropriate for the user's current device.

Tailor/Launcher-Settings Handler: To retrieve and invoke appropriate adaptations when they are not natively available on the device in question.

In addition, of particular interest to this thesis is the project's creation of a central repository to hold a list of common terms, describing user needs and preferences. As a technology focused project, the GPII uses a vocabulary that is based on technology-focused preferences and the central repository is a definitive list of the types of assistance that are currently available. As new preferences and types of accessibility solution are developed, they can be added to the repository and subsequently used by each of the components listed above. The use of an online repository to collect and share data is not a new approach and its successful use can be seen within the open source community who use versioned repositories to share code.

The development of the GPII highlights how the provision of assistance is dependent on the availability of user data and all of the components within the Cloud4All project to either produce, store or consume user data in the form of profiles. Local profiles are created using locally defined terms which are then mapped against the terms contained within the central repository. Needs and preferences can be collected from each of the devices that a person uses and be used to speed up the configuration of any new devices the person wishes to use.

Virtual User Modelling and Simulation (VUMS) Cluster

The Virtual User Modelling and Simulation (VUMS) Cluster is a collaboration between four different research projects with the aim of increasing the compatibility between their individual virtual user models (Peissner *et al.*, 2011). The cluster

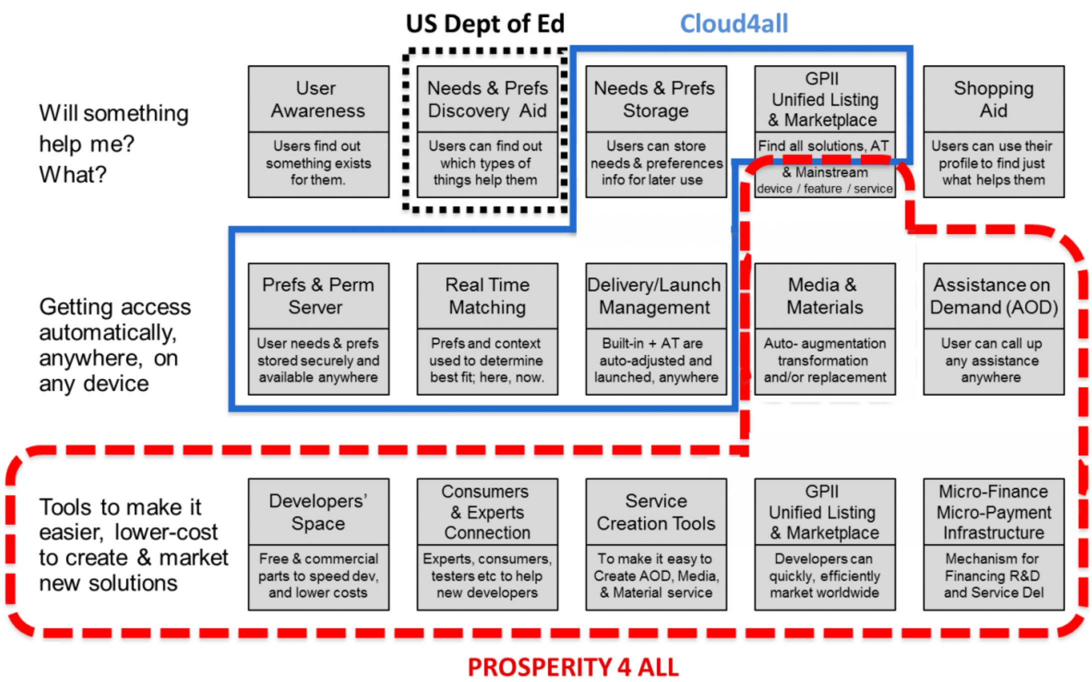


Figure 2.5: The elements that constitute the Global Public Inclusive Infrastructure (GPII) (Vanderheiden *et al.*, 2013).

is focused on modelling people with disabilities (including the elderly) and like the GPII each of the projects is investigating a different element of the design and use of user models for improving access to (and the accessibility of) user interfaces.

VICON: Virtual User Concept for Inclusive Design of Consumer Products and User Interfaces. – The production of virtual user models through the use of observational studies and their use by designers (http://vicon-project.eu/).

VERITAS: Virtual and Augmented Environments and Realistic User Interactions To Achieve Embedded Accessibility Design. – Using simulation-based and virtual reality testing to ensure the accessibility of future products and services (http://veritas-project.eu/).

GUIDE: Gentle User Interfaces for Elderly People. – Creating a toolbox to aid the development of interfaces for elderly users with mild disabilities providing multimodal interfaces that can be adapted to suit an individual (http://www.guide-project.eu/).

MyUI: Mainstreaming Accessibility through Synergistic User Modelling and Adaptability. – Providing adaptive personalisation of interfaces for real users, taking into account environmental context (http://www.myui.eu/)[15].

[15]The MyUI project is listed as a related project on the Prosperity4All website.

As a user-centred project, the VUMS cluster uses a vocabulary that is based on human capabilities and bio-medical data. The central exchange format is comprised of a superset of the variables used in each of the individual project models, meaning that each additional local model only required a single converter to translate between it and the central format. As the individual models represented subsets of the central model, they may not have contained all of the variables and when converting from the central to the local formats some of the available data would not be used. Conversely, when converting from a local to the central format, as the local model may not have all of the data required to fill the central model, there is the potential for data to be missing when converted to another local format.

Biswas *et al.* (2013) describes how the synchronisation of modelling activities between the four different projects allows VUMS to research and demonstrate interoperability of user models. Mandating a common standard user model was impractical, given the different aims of the projects involved and their attempt to provide an approach that could be used in the wider field. Instead, each project was able to define its own local format to suit its modelling needs. Attempting to provide direct conversions between all of the different formats would have again been impractical. For each additional local model, a converter would be required that translated between it and all of the existing local models, creating a problem with a complexity that increased exponentially as the number of local formats increased (potentially an n^2 problem). As a result a central exchange format was used to provide a common format that allowed the transfer of profile data between projects as seen in figure 2.6.

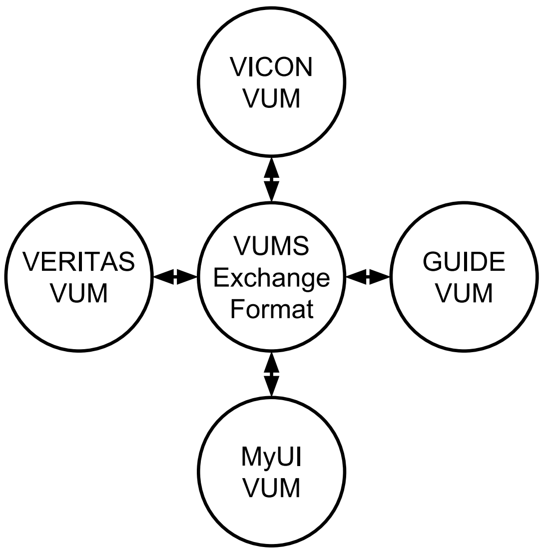


Figure 2.6: The use of a central exchange format within the VUMS Cluster (Biswas *et al.*, 2013).

2.5.2 Definitions – Profile/Model/Framework

As demonstrated in Biswas *et al.* (2013) there are a wide variety of standards and formats related to user modelling for accessibility and even the term "model" is not used consistently within the field of accessibility research. One of the initial aims of the VUMS cluster was to provide a common vocabulary, in order to avoid confusion amongst the many terms that are frequently used. Many of the proposed definitions were successfully used during a W3C WAI Research and Development Working Group (RDWG) Symposium on User Modelling for Accessibility (RDWG & W3C WAI Research and Development Working Group, 2012).

The initial distinction made was between models and profiles with the definition of a user model being taken from Benyon & Murray (1993), "*User models* are explicit representations of user properties including their needs, preferences, knowledge, as well as physical, cognitive, and behavioural characteristics." This makes a user model different from both the actual knowledge possessed by a user and the knowledge employed by a system designer. User models represent characteristics in the form of variables and are established by the declaration of these variables.

In contrast Biswas *et al.* (2013) defines *user profiles* as instantiations of user models, representing either a specific real user or group of real users. Profiles are used for the storage, transportation and provision of data in order to create *virtual users* which provide representations of users that can be used within simulations.

It is impractical to promote or enforce adherence to a single user modelling standard owing to the diversity of situations in which models may be used, however a desire to provide some form of unification has already been expressed. Given the definition of user models as a set of variables Biswas *et al.* (2013) proposes the concept of an abstract user model that is defined in terms of a series of parameters. The VUMS cluster uses this concept as a basis for creating *generic user models* which are used to provide the interoperability described previously.

The use of abstract models as means of guiding the construction of the models defined above is a popular one. Martin *et al.* (2007) describes a three-layered approach to modelling accessibility on the internet. The layers are defined in terms of their abstraction: from a high "meta-level layer" describing accessibility as a concept, to a "model-level layer" where the concepts are formed into models and finally an "application-level layer" where the models are instantiated. The higher layers provide general direction by allowing the 'definition of accessibility as a general concern which in practical terms equates to the identification of the method by which accessibility will be achieved (e.g. the use of compliance design or content ordering). Lower levels then provide more detail, such as identifying the particular set of guidelines (e.g. WCAG or Section 508). Finally, the lowest (application) layer identifies domain specific guidelines (e.g. eAdministration or Education related) that address the specific user-system interaction.

The example provided by Martin *et al.* (2007) demonstrates the use of gradually increasing specificity, with higher layers providing a basis on which lower layers are able to improve. Returning to the concept of generic user models, abstract models have also been used to guide the construction of more specific lower layer ones. The Universal Access Reference Model described in section 2.1.4.2 provided an abstract "reference model" which guided the construction of more concrete representative Common Accessibility Profiles (Carter & Fourney, 2004a). Alternatively Ackermann *et al.* (2012) proposes a "framework" to support user interface adaptability that is based on the CC/PP and uses a technology-focused vocabulary in contrast to the UARM/CAPs human-focused one.

As demonstrated in section 2.5.1, rather than focusing directly on the creation of models it would appear that recent research has instead focused on higher layer abstract models, or frameworks providing the necessary elements for lower level models to be created. In response to this trend, this thesis will also focus on providing a contribution to the creation of abstract models. In order to further disambiguate the vocabulary provided by Biswas *et al.* (2013), three definitions are provided below:

Framework: A high-layer structure that dictates the composition of models through the provision of a generic structure.

Model: A representation of a real person, object or situation and has been created in the form specified by the framework.

Profile: An instantiation of a model containing data relating to the subject of the model.

2.5.3 What Models are Required?

As well as identifying the differentiation between models and profiles Biswas *et al.* (2013) also differentiated between the different subjects of potential models used for modelling accessibility. So far the focus of this discussion has been on user modelling with the aim of creating a virtual user that can be used for simulating the accessibility of a real user. Benyon & Murray (1993) identified the importance of the provision of an explicit user model that can be displayed and edited separately to the logical processes used for identifying their accessibility. This allows—amongst other things—information about the user to be easily removed and replaced by information about a different user in order to assess their accessibility. If the user model can be identified as a distinct element within a larger accessibility model this then raises the question of what other models are required in order to assess accessibility. A short discussion of the common elements will now be presented.

The presence (or absence) of accessibility issues between a user and a piece of technology can be predicted by simulating their interaction using representative models and profiles. However, given the variety found within users, technologies and interaction paradigms, there is no universally accepted method of modelling for accessibility. Biswas *et al.* (2013) uses the term "simulation" to describe the process that enables the accessibility of the interactions of a virtual user to be assessed and it involves the use of a number of different models:

User Model: A set of user characteristics required to describe the user of a product.

Device Model: A formal machine-readable representation of the features and capabilities of one or more physical components involved in user interaction.

Environment Model: A set of characteristics used to describe the environment in which an interaction is taking place.

User Agent Capabilities Model: A representation of the capabilities of a user agent (any end user software that can render application content).

Application Model: A representation of the states, transitions and functions of the application.

***User* Interaction Model:** A representation of the interaction behaviour of an application.

Context Model: A representation of information that can be used to characterize the situation of any person, place or device that is relevant to an interaction.

The number of models that have been identified is due in part to the range of modelling applications for which the VUMS Cluster is attempting to cater. As an example, separate models are identified to describe an application, user agent and device, however only a single 'user' model is identified. Although they will be described further in the next chapter, two definitions will now be presented. "Levels of abstraction" describe the way that higher-level tasks can be built up from lower-level ones (e.g. hand-eye co-ordination). "Layers of abstraction" describe the way that data can be viewed in multiple frames of reference; with lower layers providing structural information and higher layers storing specific contextualised measurements.

This can be understood in relation to the concept of fidelity that is used to describe the level of detail at which a simulation (and its resulting models) are aimed. Biswas & Robinson (2010) describes the fidelity and the resulting trade-off between low-level (high-fidelity), and high-level (low-fidelity) models. A similar problem is also observable when contrasting between low-layer (low-fidelity) and high-layer (high-fidelity) models. The greater the fidelity of a simulation, the more accurately it is able to represent the situation under investigation. This accuracy is however gained at the expense of transportability, resulting in an inability to reuse the simulation to model other situations. Reducing the fidelity of the simulation increases its transportability, whilst decreasing its ability to accurately represent a situation.

2.5.3.1 ISO 24751

ISO 24751 deals with the provision of individualised adaptability and accessibility in the context of e-learning. It is a three-part standard, which describes the user and application profiles separately. Currently under revision, it is based on a fixed vocabulary of technology-focused needs and preferences and is being used by the GPII. Preferences refer to device or software settings that can be specified to improve accessibility (e.g. font size). As they are device specific, they provide an accurate representation of the needs of the user in terms of the technology-focused preferences required in order for a device to be accessible. As above however, this specificity reduces the transportability of a user's profile owing to a need for mappings between all related preferences across devices.

2.5.3.2 ISO 24756

ISO 24756 uses Common Accessibility Profiles (CAPs) to allow direct comparison between the needs and capabilities of users, systems and their environment (Fourney & Carter Fourney & Carter (2006a)). It views accessibility in terms of channels of communication that are facilitated by human-focused capabilities. CAPs are constructed from a series of Interacting Components which are able to either input or output via a fixed vocabulary of modalities (e.g. visual acuity). Unlike preferences, the approach provides a static vocabulary against which any device setting may be mapped (Atkinson *et al.* Atkinson *et al.* (2010)). Human capabilities provide the transportability required to compare a single user profile against multiple device profiles. Their generalisability, however, comes at the expense of the extra effort required in terms of describing device settings and adaptations in terms of their corresponding human-focused capabilities.

2.6 Aims, Objectives and Scope

Throughout this chapter, background and other relevant literature has been discussed in order to identify the scope within which this thesis will operate. This section will pull together the discussion in order to identify the contribution that this thesis will make.

2.6.1 Scope

The breadth of topics covered thus far is indicative of both the complex and multi-disciplinary nature of the problem being faced. There are many facets to the problem of providing universal accessibility and this thesis cannot reasonably be expected to address them all. The scope of this thesis is therefore constrained to the process of identifying accessibility barriers and providing a description of the support required in order to address them. Although this is a problem to which a technical solution can (and will) be proposed, the existence and importance of non-technical considerations has however been consistently acknowledged throughout this chapter. This theme will therefore be carried throughout the rest of the thesis and the development of the technical solution will be supplemented with investigations of the wider issues that surround its use and implementation.

There is also diversity in the literature regarding the level of technology employed by each of the elements identified in the process under investigation. Section 2.3 identified the use of both technology-focused as well as human-focused approaches for providing people with accessibility. Section 2.4 then identified the use of both technology and human-focused approaches to identifying their accessibility requirements and provide appropriate solutions.

This can be further demonstrated through recent UK campaign "Race Online" (2010—2012) which aimed to increase the number of people who regularly used the internet. The campaign took the approach that increasing availability and raising awareness of associated benefits would increase internet use. Although it was able to claim success in terms of its specified aims, the approach could be described as short-sighted in that it concentrated solely on technology-focused barriers to entry, rather than the wider human-focused barriers that surround sustained use. Having been initiated in a reportedly sustainable fashion (Capgemini, 2012), attention is shifting to other interventions and a follow-on campaign ("Go ON UK"[16]) has been launched which focuses instead on improving people's online

[16] http://www.go-on.co.uk

skills.

The two approaches appear to produce a dichotomy, either placing the burden on upgrading the technology in order to accommodate the user or 'upgrading' the user to accommodate the technology. The two approaches can however be considered as complementary (as described by the functional model in section 2.1.3) and there is therefore a need for both to be considered as this research progresses.

2.6.2 Thesis Aims, Their Derived Contributions and Resulting Objectives

Throughout this chapter, a number of aimss have been identified, which define the different elements of the problem that this thesis addresses:

Aim 1: Functional Assessment of Accessibility.

Aim 2: Variability Between and Within Individuals.

Aim 3: Variety Within and Interaction Between Accessibility Barriers.

Aim 4: Variety Within and Interaction Between Accessibility Solutions.

Aim 5: Variety of Agents Producing and Using Data.

Aim 6: Variability of Data Quality.

While all of the aims have been at least partially addressed in existing research (as will be discussed in chapter 4), existing approaches generally deal with the aims either in isolation, or as a subset of this list. The overarching aim of this thesis is therefore to develop an approach that is able to address all of the aims simultaneously. Rather than promoting a rigidly prescriptive format, the required functionality will be better provided by a 'framework' that guides both the creation of profiles and their comparison within subsequent of models of accessibility. A solution of this kind will provide a number of functional contributions:

- Identify existing approaches used to represent people or technology and develop a format that is suitable to store data for use on different platforms and devices.

- Investigate the ability of older people to provide data and develop a mechanism for identifying and describing data from different sources with the aim of understand the confidence with which predictions can be made.

- The format should enable descriptions to be made at different levels of granularity, in order to facilitate comparisons in different contexts.

- Develop an approach for comparing between profiles that is able to quantify the accessibility of an interaction and provide a description of the assistance that is required if a barrier is identified.
- Enabling speculative augmentation to depict different technology configurations or the use of different forms of assistance.

2.6.3 Objectives

In order to achieve the aims and associated functional contributions, the following objectives will be used to guide the research:

- A search of relevant literature will be used to better define the problem and identify existing approaches that provide partial solutions.
- The potential solutions will be combined to create a novel approach from which a design can be developed.
- An investigation of the ability of older people to provide the data required to drive the approach will be conducted.
- The ethical implications of automated data collection will be explored, resulting in the identification of a mechanism to allow mediation between data from different sources.
- The design and data mediation mechanism will be integrated and a prototypical system will be implemented.
- Data collection and evaluation studies will be used to provide data from real users for use within the implementation to evaluate the effectiveness of the approach in solving the problem.

2.6.4 Wider Goals

This research is designed to form part of the wider body of research and a list of wider goals is therefore presented that reiterate the wider scope of the work. This list is

- Supporting the ongoing technology inclusion of older and disabled people, in order to prevent disengagement and the related negative effects.
- Support the identification of accessibility barriers and provision of support to overcome them.
- Support for both legacy and future technology.

With the aim of the thesis identified, the following chapter investigates existing approaches that provide partial solutions and the potential for their combination in order to provide a unified approach which addresses all of the aims specified above.

Chapter 3

Research Methodology

The previous chapter reviewed existing literature and identified the need for a new (unified) approach to accessibility modelling. In order to fulfil this need a methodology is required through which the research can be conducted. Freitas (2009) describes the use of "the scientific method" within Computer Science and identifies three distinct methodologies and their applicability to different domains. Similar to the development of a new product or a piece of software, the research within this thesis follows an overall methodology that is similar in form to a development life-cycle. While there are a number of different development life-cycles, which vary in terms of their focus and structure, one common feature is their use of a series of stages (or phases). This chapter describes how the methodologies identified in Freitas (2009) are blended through there use at different stages of the research's "development life-cycle".

Section 3.1 begins by providing an introduction, in terms of the overarching approach taken in finding a solution to the problem identified in the previous chapter. Section 3.2 contains a brief description of methods employed in the development of the aims and functional contributions and how they fit into the approach. Section 3.3 then continues by providing details of the iterative process used to develop the proposed solution. Following the design process, section 3.4 identifies the evaluation process that addresses the latter stages of the life-cycle. Finally, section 3.5 concludes by providing an overview of the interaction between the stages, their location within the contents of the thesis and the contribution of the methodology to the research.

3.1 The Overarching Research Approach

The potential breadth and complexity of the problem identified in chapter 2 results in the need for an overarching approach that is flexible enough to encompass methods from a number of different domains within Computer Science. In demonstrating the scientific merit of the field of Computer Science, Freitas (2009) identifies three different methods that are regularly used within the field.

Theoretical methods are used in the formation of data models and algorithms; providing the ability to codify knowledge and understand its uses and limitations.

Experimental methods are used to test " the veracity of theories"; subjecting them to scrutiny and/or providing data that can be used to support them.

Simulation methods provide the ability to "investigate systems that are outside of the experimental domain"; providing a query-able product that can be used to better understand or test a theory.

Given the complex nature of the problem identified in the previous chapter all three of the methods needed to be employed in the development of a comprehensive solution.

Firstly, the problem revolves around the need for a new unified approach that is suitable to satisfy all of the aims in a single framework. The development of the framework relies on theoretical methods to ensure that it is both representative and functionally appropriate. Secondly, areas of the domain within which the problem lies are still the subject of ongoing research themselves (e.g. the collection and use of user-generated data). Experimental methods are useful in providing more reliable data from those areas, improving the specificity of the requirements. Lastly, following its development the approach required testing to evaluate whether it actually met the requirements that define the problem. Simulation methods provide a vehicle for testing the ability of the framework to meet the aims by allowing the framework to be practically trialled.

The combined use of the methods in developing a new approach to solve the problem, mirrors the development of a piece of software. For this reason the overarching research approach that is used is that of a software development life-cycle. Ideally (for brevity) the research would follow a basic 'waterfall' development model, with requirements being used to drive development and then evaluation in one linear flow. However, as will be described in section 3.3, a more iterative approach (Larman & Basili, 2003) is required in order to address some of the intermediate findings identified in chapter 6.

3.2 Initial Requirements Gathering

The first stage of a project should involve the gathering of requirements, with the aim of developing a clear understanding of its intended outcomes. Within this research the literature review (detailed in chapter 2) was used to accomplish this stage. Literature was selected to both provide a representative overview of the domain within which the research will be conducted and identify a set of requirements that the research will aim to satisfy.

The potential breadth of the field made it difficult to bound the literature search in many of the usual ways (e.g. age or source). Instead the literature was initially chosen to provide a base on which the research could be built, e.g. contextualising the research. In order to enable literature to be chosen systematically, the review was structured as a logical progression from generic "first principles" through to the specific domain that the research will target.

As described above, the purpose of the literature review was to identify the aims of the research. Therefore, as well as providing a knowledge base on which to ground the research, the literature was used to build up a set of requirements to be addressed within the development stage. This stage was accomplished through critiquing the literature, resulting in a "gap analysis" as seen in traditional research projects. However, rather than simply identifying the problem to which a solution is required, the requirements gathering process also needed to provide: 1) a set of requirements suitable to drive the development of a solution and 2) a set of criteria suitable for use in the evaluation of the solution's effectiveness.

While aims are useful to inform the direction that the research should take, by providing a target that can be aimed at (the first need), they were too subjective in nature to satisfy the second need. To this end, towards the end of the chapter the aims were used to produce a set of functional contributions that were more suitable for use in the evaluation process. In addition a set of objectives were also developed in order to provide a set of interim goals that could be used to guide the research.

With the aims, functional contributions and objectives of the research identified, a methodology is now required to ensure that they are accomplished. In fitting with the software development life-cycle described previously, this is accomplished through the development and testing of an approach. The remainder of this chapter will be used to describe the process by which that approach was developed and evaluated.

3.3 The Iterative Design Process

With the scope and scale of the problem defined, development stages of the life-cycle were entered, during which an approach was developed that was suitable to fulfil the requirements. While the aims of this research were specific enough that a set of measurable functional contributions could be created, the design of a solution to meet the requirements still included components from a range of different domains.

For this reason an iterative approach was chosen to allow concerns belonging to different domains to be addressed separately. Within an iterative software development life-cycle a number of stages are performed multiple times to gradually build up a solution that meets all of the requirements. They may include: planning, analysis and design, implementation, testing and evaluation. As the aims of this research are concerned with the development of an approach rather than a piece of software, only those stages concerned with the design process were performed iteratively. While the inability of the approach generated by the first iteration to meet 'all' the requirements will be highlighted, formal evaluation of the approach was be conducted following the second iteration.

Within this thesis the design process consists of two iterations. Each iteration comprises a number of stages:

An information gathering stage: During which existing approaches are identified or evaluated.

An approach development stage: During which a solution is developed to meet the initial requirements.

A design development stage: During which the approach is used within the development of a design for use in the evaluation process

The first iteration focused on the production of an approach that met the requirements, based on information gathered exclusively from the theoretical areas of the modelling community. The information gathering stage involved another literature survey, focusing on existing approaches within the field of accessibility modelling. As the field is currently biased towards the creation of models that are structurally and semantically appropriate, the two development stages were concerned with the core functionality of representing and comparing between individuals. The approach development stage involved the identification of synergies between the existing approaches and their subsequent combination. The new approach then results in a design being produced that is theoretically able to meet the requirements, although unsuitable for use in the real world.

The second iteration focused on the introduction of requirements that were generated by the uncertainty associated with data from real people. The suitability of the approach to meet these additional requirements was therefore determined by repeating the stages undertaken in the first iteration with this new focus. The information gathering stage consists of a study that examines the ability of older people to accurately provide the data required by the approach. The study is described in chapter 6 and full methodology is provided at the beginning of the chapter. Within the approach development stage the results of the study were addressed through the investigation of number of ethical considerations. As in the first iteration, the updated approach was fed into the design development stage, this time resulting in an updated design.

At the end of the development phase a new approach should have be produced, which is suitable to meet the aims identified in chapter 2. The approach is accompanied by a design that is suitable for implementation as a practical demonstrator of functionality that should embody the functional contributions. With the development stages complete, the research turns to assessing the appropriateness of the previous stages' outputs.

3.4 The Evaluation Process

The latter portion of most life-cycles consists of a number of processes that are dedicated to quality assurance (through the use of testing) and the identification of any lessons that may have been learnt (through evaluation). Once a suitable design was produced, a series of stages were used to support the evaluation of the approach. In section 3.2, the aims were identified as being unsuitable for use in evaluating the result of the development process. Instead a series of functional contributions were created, which provided a more objective set of criteria through which the research outputs could be evaluated.

As in a typical life-cycle, once the development process was completed the research turned its attention to evaluating the new approach that was developed. The subjectivity of the aims means that the approach itself was not directly testable, rather, the ability of the design to satisfy the functional contributions was investigated instead. Figure 3.1 depicts the relationship between the aims, contributions, approach and design. If evidence could be found to suggest that the design had satisfied all of the function contributions, the evidence could be used to support the use of the approach to provide a solution that achieved the aims of the research.

In order for the evaluation to be performed, the ability of the design to deliver against each of the functional contributions needed to be demonstrated. Rather

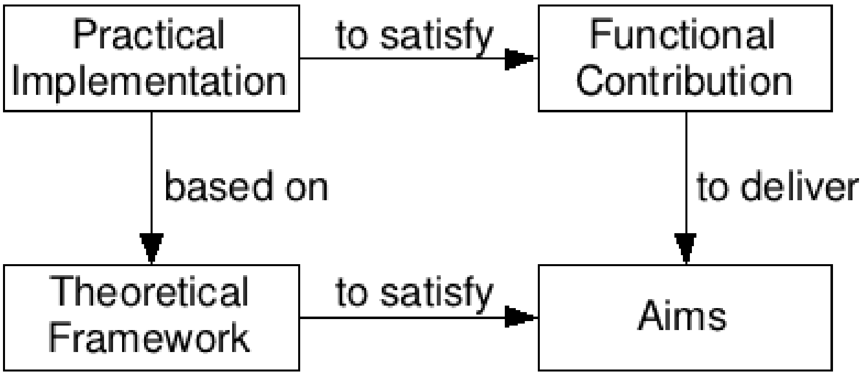


Figure 3.1: The use of aims and functional contributions in evaluating the approach.

than a theoretical analysis, a practical example of the approach being used was required. To this end the design was implemented as a prototypical system, which was capable of simulating interactions and providing feedback on their accessibility.

As well as providing an artefact for use within the evaluation stage, the implementation stage was also used to provide evidence to support the practical applicability of the approach for use in real developments. The technologies and techniques used within the prototype is be documented, along with the development process and the structure of the resulting system. Together this is used to suggest the feasibility of incorporating the approach into future systems.

In addition to developing a prototypical system, the validity of evaluation was also improved through the use of real world data. For this reason the evaluation incorporates a data collection study to generate data from typical users. The data is then used to drive a practical evaluation of the prototypical system, by demonstrating its ability to deliver against the functional contributions. Chapter 9 provides further details of the methods used, both for the data collection study and the evaluation of the system.

3.5 Summary

This chapter has described the overarching methodology that followed in order to conduct the research required to provide a solution to the problem identified in the previous chapter. This section will now provide a summary of the methodology in terms of the different stages within the life-cycle and sign-posting to the location of the stages within the thesis.

Chapter 2 has already been used to generate the research question in term of a series of aims, functional contributions and objectives. These form the requirements used to drive the development of a solution, the criteria that used to evaluate the solution's success in meeting them and a high level method through which the solution is obtained.

With the requirements identified, an iterative development approach is used, consisting of information gathering/planning and subsequent development stages. The first iteration is described in chapters 4 and 5; they include the identification of existing approaches, their used in informing the development of a new approach and resulting theoretical design. The second iteration is then described in chapters 6 and 7; which identify the complexities relating to the collection/use of real world data and attempt to develop both the approach and resulting design to take account of them.

While the iterative process is used to develop a solution to the problems identified in chapter 2, chapter 8 describes the development of a practical implementation of that solution as a prototypical system. The prototype supports the evaluation of the approach by both providing data relating to the feasibility of its use in real systems and a system for use within an evaluation study. The evaluation itself is described in chapter 9 and includes both a data collection study to provide real user data, and the use of that data within the prototype to identify its ability to satisfy the functional contributions.

By describing the overarching methodology used to conduct the research in this thesis, the chapter ensures compliance with the scientific method by this thesis. While the methodology takes the form of a development life-cycle, it includes three different methods that are widely used within the research community. The use of those methods contribute to the validity of the research by providing a documented process that is coherent with the scientific method; identifying (and quantifying) a problem, investigating existing approaches and proposing a novel solution, developing the solution and finally testing it using objective measures.

Chapter 4

Developing an Approach

This chapter will investigate two general themes that are common within the field of modelling for accessibility. Section 4.1 will discuss the need for standardisation to allow comparison to take place between representations of different actors. Section 4.2 then discusses the need for abstraction within representations to allow for variety in the detail of the information that is available. The benefits and disadvantages of the themes will be discussed through their use in existing approaches and their ability to satisfy the objectives identified in the chapter 2.

Section 4.3 will draw links between the approaches, and describe how they complement each other to suggest a need for the themes to be combined. The framework—that will be developed in subsequent chapters—will then be introduced by discussing how the themes and approaches will be integrated.

As described in the previous chapter, accessibility can be defined in terms of the functional match that is apparent between a user and a piece technology. The suitability of the match can be measured in terms of the quality of interaction they are able to sustain, depicted within an interaction model. Figure 4.1 depicts a basic interaction model that contains information about a user and a piece of technology. Throughout the chapter the model will be expanded, with links being drawn between existing approaches and identification of how they complement each other.

4.1 Standardisation

In order for a match to be identified between a user and a piece of technology, it must be possible for a comparison to be made between them. Section 2.5 described how models representing a user and device may be compared within an interaction model. While there is currently no universally accepted format to describe how an interaction model should be constructed, there is a need for compatibility between

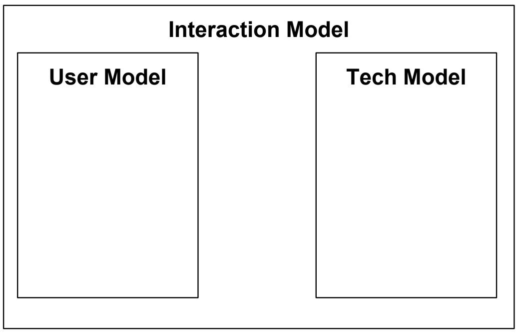


Figure 4.1: Interaction Model – A basic model including two actors.

its constituent parts in order for a comparison to be made.

Compatibility is often achieved through the use of standards (Peissner *et al.*, 2011) which dictate a common format that must be adhered to. While they ensure uniformity, the use of standards can be constrictive as demonstrated in section 2.3.2. The Shannon-Weaver Model of Communication (Shannon & Weaver, 1949) is an example of a very loose de-facto standard that has been used as a means of modelling interaction in a predictable format. The flow of information from a transmitter to a receiver—via a channel of communication—provides a standard representation, allowing a comparison to be made.

This section will focus on the use of standardisation within the interaction model. As well as standardising the model in terms of its components, it is also possible to standardise the structure of the components themselves. Figure 4.2 shows the potential for standardisation within the components in a generic interaction model. Firstly, a standard vocabulary can be used to bridge the gap between the two models allowing a channel to be created. Secondly, the models themselves can be standardised, both in terms of the structure used to represent their contents and the external interface used to expose their contents for comparison.

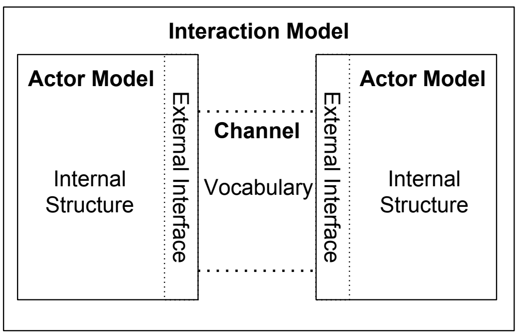
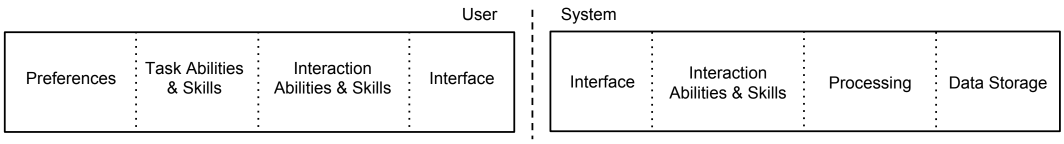


Figure 4.2: The elements of the interaction model that can be standardised.

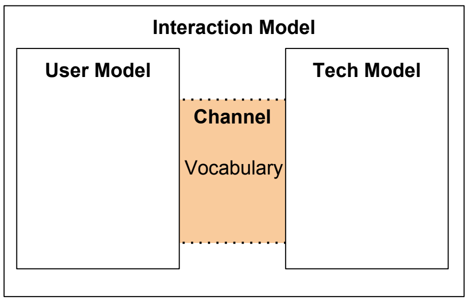
4.1.1 Common Vocabulary (for Comparison)

Accessibility occurs when a user has the abilities required to access the functionality provided by a piece of technology, and the technology is able to provide functionality that is sufficient for the needs of the user. Although people and technology are different, in order for the accessibility of their interactions to be assessed, a comparison must be made between their abilities and requirements.

In section 2.1.4.2 the Universal Access Reference Model (UARM) was used to describe interaction between a user and a device (or combination thereof). As seen in figure 4.3a, a user is able to employ a number of "Interaction Abilities and Skills" to complete various "Tasks" in accordance with their "Personal Preferences". A piece of technology takes the information it receives through its "Interaction Abilities and Skills" and uses "Processing" to interpret it in terms of the information available within its "Data Storage". By placing the point of comparison in the space between the user and device the UARM implicitly chooses a vocabulary based on low-level interaction abilities.



(a) An interaction continuum according to the UARM.



(b) An interaction model within the framework.

Figure 4.3: Two standard vocabularies that can be used to model interaction.

It is however possible for the point of comparison to be moved, resulting in a change to the vocabulary used. As there are a finite number of points along the interaction continuum at which a comparison can be made, standards and vocabularies can be classified using the point at which they are based. Table 4.1 is representative of four foci that are regularly used and the standards that are associated with them. The standards listed were designed for different purposes but have all been used to provide vocabularies allowing the assessment of accessibility, albeit in different situations.

It is also possible for a standard to contain elements from several categories. For example, ETSI ES 202 746 is able to store human-focused disabilities and abilities, technology-focused preferences as well as various other pieces of personal information. Each of the foci has its own benefits and drawbacks which will now be discussed in turn.

Table 4.1: Standards Demonstrating Vocabularies Based on Different Points of Comparison

Human-Focused Disabilities	Human-Focused Capabilities	Technology-Focused Tasks	Technology-Focused Preferences
WHO ICD	WHO ICF	WHO ICF[1](Billi Extension)	ISO 24751[3]
	ISO 24756	ISO 9999[1]	CC/PP
		ISO/IEC 40500:2012[2]	EN1332
			ISO/IEC 29138-1

[1] Mappings between the two standards are provided by Bougie & Heerkens (2009).
[2] Based on the WCAG 2.0 Guidelines; the Principles are ability-focused.
[3] Based on IMS AccessForAll.

4.1.1.1 Human-Focused Impairments

Human-focused impairments refer to named impairments that may cause a user to be disabled in a predictable way (e.g. low-vision, deafness or arthritis). Vocabularies based on this focus describe users in terms of the impairments they have, and technologies in terms of the impairments that they are designed to assist. The focus is therefore rooted in the medical model of disability (section 2.1.3) and its benefits/drawbacks are similarly inherited.

The advantages of an impairment-focused vocabulary revolve around its focus on identifying users' disabilities and the ability for assistance to subsequently be provided based on related stereotypes. Disability-focused vocabularies are most useful where help is required to solve a problem caused by a specific impairment. The BBC's "My Web My Way"[1] is a basic assistance recommender system that classifies adaptations based on the disabilities they assist. When a user arrives on the page, they are presented with the choices in figure 4.4 and making a selection leads to a list of possible adaptations and instructions for implementing them. As a user-facing service it has the advantage of not requiring the user to have any prior knowledge about the assistance they require; instead they only need to describe the problem they are experiencing.

[1] http://www.bbc.co.uk/accessibility/

Figure 4.4: Accessibility options offered through the BBC's "My Web My Way".

In terms of software development, users are regularly classified in terms of their impairments. Techniques such as personas provide developers with examples of real—if only stereotypical—users which they use to inform decision making (Quesenbery, 2013). Heuristic evaluations such as the barrier walk-through technique also make use of impairment-focused descriptions for similar reasons (Yesilada *et al.*, 2009).

It is the stereotypical nature of the focus however that also results in its disadvantages. While they are useful for situations where there is limited information about a user, stereotypes—by definition—do not accurately represent individual users. As described in section 2.1.3, the mapping between impairment and disability may be dependent on the user's situation. Impairments do not necessarily result in disabilities and it is also possible for a single impairment to lead to a user being disabled in several different ways.

This makes finding appropriate assistance problematic, as the assistance required by one individual may not be appropriate for another. Although technologies can be labelled in terms of the impairments they can cater for, without the appropriate mapping there is no inherent way to tell from a technology-focused preference what assistance it provides. In addition, impairments are negative, providing information about the barriers the user faces and things they are unable to do. As they do not provide any information about a users abilities, impairments cannot be used to identify the user's strength or suggest potential workarounds (Keates *et al.*, 2007).

4.1.1.2 Technology-Focused Preferences

Technology-focused preferences refer to hardware or software components that can be configured to improve accessibility (e.g. button size, zoom level or mouse sensitivity). Vocabularies based on this focus describe technologies in terms of the preferences they are able to provide, and users in terms of the preferences they require (potentially in order to perform a task). The focus can therefore be compared to the social model of disability (section 2.1.3) as it describes what technology must do to be accessible to a user.

The advantages of technology-focused vocabularies revolve around the ease with which they can be measured, stored and subsequently reused. The number of technology-focused standards listed in table 4.1 demonstrates the scale of their use and the specificity with which they can describe a user's requirements in relation to a particular piece of technology. As long as it is configurable, any technology preference can potentially be included in a profile used to provide personalisation for a user.

Profile creation involves an audit to expose the configuration settings for individual preferences, encoding of the settings in a transportable format, and storage to a medium which can be accessed by any technology wanting to re-use the settings. The Personal Portable Profile (Liffick & Zoppetti, 2007) was a USB flashdrive-mounted personalisation system, designed to capture system settings when plugged into a user's computer, allowing them to be invoked on another PC. While it was intended to provide transportability of system settings, the paper highlights the challenges associated with doing so, including: (1) identifying and isolating an appropriate accessibility setting and (2) translating and invoking the settings without requiring a system reboot. Although the project described future work intending to provide support for Mac OSX and Linux environments, no information can be found to indicate that even the Microsoft Windows version of the system was successfully implemented or that the project was continued beyond 2007.

Software-based preferences are natively stored as settings within program or system variables and there is no common method for their extraction. Application Programming Interfaces (APIs) can be used to expose settings, however their availability and implementation is variable. In addition, as described in chapter 2 there is currently no accepted format for the storage of settings once they have been exposed.

More recently the Global Public Inclusive Infrastructure (GPII) has developed a "Common Registry of Terms" which is proposed for use in updating ISO 24751. If implemented this will allow settings to be transferred more widely between devices, as local system preferences can be mapped to those in the registry. Although the registry will allow translation between existing technology-specific settings, its structure has been designed to be both flat and updatable. This means that whilst new preferences may be added, it will not be possible to suggest configuration values from those stored against existing preferences.

While settings from one device can be recorded and translated for use on another, technology-focused preferences have no inherent ability to provide translations between preferences. The discovery of new preferences relies on a deeper understanding of the effect that a preference has on a user's impairments. Preferences can be grouped based on their functionality and mappings used to suggest relationships, but this requires the use of vocabularies based on other points of comparison.

4.1.1.3 Technology-Focused Tasks

Like preferences, tasks are technology-focused, however rather than describing configurable settings, technology is described in terms of the functionality it is able to provide. The user's abilities are then described in terms of the tasks they are able to complete; for example clicking a button, moving a mouse or sending an email. This results in an approach which mirrors the functional model of disability, with comparison being focused closer to the point of interaction.

Unlike impairments and preferences, technology-focused tasks can be used to describe what the user is able to do. Where a mismatch between user and technology occurs, this focus allows alternative methods of communication to be found which are based on tasks that the user is able to perform. The TERESA tool (Mori *et al.*, 2004) provided an ability to model applications in terms of their constituent tasks, allowing interfaces to be designed using appropriate elements. Based on the ConcurTaskTrees notation, tasks can be broken down into their component subtasks allowing a degree of task transportability. Although the tool itself has been superseded by MARIAE it allowed the same application to be presented on different devices, using different interface elements.

As an example, a task requiring a user to provide an input could be achieved either through the use of the keyboard or mouse to manipulate a variety of interface elements. While similarities between the mouse and keyboard are limited to their use of buttons, interface elements can be grouped by the tasks required for their use (clicking, dragging, navigating or requiring alpha-numeric entry). The transportability of this approach is therefore limited to comparable interaction paradigms, and will require the definition of new tasks for future technologies (such as multi-touch gestures (Kammer *et al.*, 2010).

MARIAE is development environment for the MARIA language. MARIA (Paterno *et al.*, 2009) is a "Model based lAnguage foR Interactive Applications" which provides an incremental step over TERESA, by drawing from ConcurTaskTrees. MARIA provides the ability to move away from the description of individual user interface implementation to a higher level of abstraction. Through this abstraction, the intended functionality of an interface is described and additional interfaces can be developed for alternative devices.

The W3C's Web Content Accessibility Guidelines (WCAG) are the de-facto standard used to deliver online accessibility. They have been designed for use by developers and are therefore (mostly) technology-focused, describing the functionality that is required in order for web content to be accessible to a wide range of people. Automated testing tools are available to parse the underlying source code by using a series of rules to check adherence against each of the guidelines in turn. Some of the guidelines do require validation which cannot yet be defined in a script, however this is due to their use of human-focused sub-clauses[2].

4.1.1.4 Human-Focused Capabilities

Capabilities describe users and technology in terms of their human-focused interaction abilities. Like technology-focused tasks this again results in the potential for alternative methods of interaction to be identified when a mismatch occurs.

The human-focus of capabilities provides the additional benefit of a static vocabulary, which can be used to increase the transportability of profile data. The SUPPLE system (Gajos *et al.*, 2006) provides an example of an approach which selects appropriate interface components based on the abilities of the user.

The evolution of the human body is effectively static when compared against the speed at which new interaction paradigms are being developed. The interaction capabilities displayed by the current population is almost identical to those exhibited before the birth of the first computing engine. Fleishman *et al.* (1984) is a taxonomy of human capabilities that was still applicable 20 years after its creation (Balasubramanian & Venkatasubramanian, 2003) and can still be applied today. This makes human capabilities more resilient against the potential for change that is possible in all of the other vocabularies (Atkinson *et al.*, 2010).

While they have their advantages, human-focused capabilities are more difficult to measure than tasks, with current systems relying on either task-based assessments (Gajos *et al.*, 2006) or the user themselves for profile acquisition (Sala *et al.*, 2011). The inability to fully automate the assessment of WCAG 2.0 is due to its use of human-focused sub-clauses and it is for this reason that there is a growing appreciation of the need for human mediation in accessibility assessment (Kelly *et al.*, 2007).

4.1.1.5 Summary

All of the vocabularies can be used to assess the accessibility of interactions between users and technology. By using a common vocabulary comparisons can

[2]For example comprehension skills are required to determine whether captions provide an accurate summary of the tables they are describing (WCAG 2.0, success criterion 1.3.1, Technique H73).

be made, however this results in either the user or the technology being described in terms of the affordances, functionality or settings of the other. Impairments and capabilities place the point of comparison within the user, and preferences and tasks place it within the technology. While a technology-focus increases the ability to automate data acquisition—due to automation being provided by the technology itself—a human-focus provides a static vocabulary that can more easily incorporate new interaction paradigms.

The definition of accessibility developed in section 2.1 revolved around the compatibility of user and technology in terms of the quality of their interaction. Impairments and preferences provide a description of accessibility problems or solutions, but do not address the interaction itself. When a mismatch occurs, they have no ability to describe its underlying causes and as a result cannot be used to suggest alternative accessibility solutions. Tasks and capabilities provide a functional description of the abilities of users and technology to interact with each other. When a mismatch occurs, the gap that is apparent between their abilities inherently describes the assistance that is required for interaction to be possible. This will be discussed further in the next subsection.

As the benefits and challenges of the three approaches overlap, mappings can be used to combine two or more approaches, using the benefits of each to mitigate the challenges of the others. The VERITAS system (Kaklanis *et al.*, 2011) merges tasks and capabilities together, by building a series of low-level capabilities into a higher level task.

4.1.2 External Interfaces

The previous section demonstrated the use of a common language as a means of allowing a comparison to be made between profiles representing users and technology. Rather than the gap, this subsection focuses on the profiles themselves and the approaches used to expose their data. As seen in figure 4.5 the external interface of a profile is the point at which it interacts with other models, allowing comparisons to be made.

One element of the external interface is the vocabulary it uses, which ensures that exposed data is semantically comparable with that of other profiles. Another element is the format (or syntax) used to expose data, which ensures that it is syntactically comprehensible by comparison functions.

4.1.2.1 Semantic Standardisation

In order to compare two profiles against each other they must first expose information that is actually comparable. If a user requires text that is a certain size,

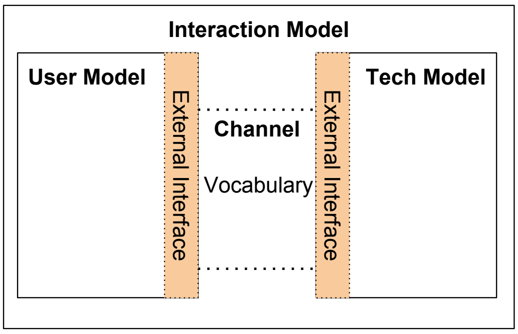


Figure 4.5: Interaction Model – Focusing on the use of standard external interfaces.

information about the colours that a monitor can produce text in is not useful. Therefore, rather than being concerned with the way that profile data is encoded, semantic standardisation ensures that the data that is exposed by each profile can be sensibly compared.

Abilities and Requirements The merits of using different vocabularies were debated in the previous subsection. The vocabularies were all described in terms of the way they exposed abilities and requirements; abilities were the qualities that a user or piece of technology had, and requirements were the qualities that they needed. For example, a person with poor eyesight (a low visual sensory ability) can be described as 'requiring' a large font size (a high visual presentation ability) to be able to read text. However it is equally true to say that a monitor (that is able to display text of a certain font size) may 'require' a user with a certain eyesight. There is an asymmetry between abilities and requirements, with the words being used to indicate the focus of attention of the vocabulary. Abilities are qualities that are contained by the owner of the profile and requirements are qualities that are being sought from other profiles.

The impairment and preference based vocabularies dictated whether it was the user or technology who was described in terms of their abilities. ISO 24751 is a three-part standard consisting of: ISO/IEC 24751-1 (2008) a preference-based vocabulary, ISO/IEC 24751-2 (2008) a description of user requirements, and ISO/IEC 24751-3 (2008) a description of digital learning resources' abilities to meet those requirements. Owing to the asymmetry between parts two and three, users are defined solely in terms of their requirements and learning resources in terms of their abilities. The standard is therefore restricted to comparing a single user against a single learning resource. In addition, profiles created using the standard have a limited ability to provide information that can be transported outside of their intended scope. They also imply that accessibility is provided solely by the learning resource, in keeping with the social model of disability. Both the user and learning resource are effectively black boxes and while the inclusion of assistive technologies can be reflected by changing the abilities of the learning resource, the actual cause of the change is hidden from view.

Transmitters and Receivers Rather than abilities and requirements, the Shannon-Weaver Model of Communication describes interaction in terms of transmitters and receivers, modelling information as it flows either user-to-technology or technology-to-user. Like abilities and requirements, there is an asymmetry between transmitters and receivers in that a transmitter is able to output information and a receiver is able to input it. However they also share a symmetry in that both describe the user or technology in terms of their abilities (the things they are able to do). This is similar to the approach taken by the task and capability based vocabularies described in the previous subsection as both described users and technology in terms of their abilities, either using a human or technology focus. While a user can have an ability to perform a specific task, accessibility is also dependent on a technology's ability to afford the task.

ISO/IEC 24756 (2009) is a standard based on the the Universal Access Reference Model (Carter & Fourney, 2004b), which is in turn based on the Shannon-Weaver Model. It describes the production of "Common Accessibility Profiles" (CAP) that allow direct comparisons to be made between the abilities of users, systems and their environment. By viewing accessibility in terms of channels of communication that are facilitated by human-focused capabilities, both users and technology are modelled as exposing their own abilities. Abilities are exposed through either Input Receptors ($\text{CAP}_{(IR)}$) or Output Transmitters ($\text{CAP}_{(OT)}$) which are both types of Component Features ($\text{CAP}_{(CF)}$). $\text{CAP}_{(CF)}$ describe capabilities to receive and transmit information and are defined in terms of a series of properties including:

Name: Allowing the the capability to be identified.

Direction: Whether information is transmitted or received (implicitly implied by the $\text{CAP}_{(CF)}$ being either a $\text{CAP}_{(IR)}$ or $\text{CAP}_{(OT)}$).

Modality: The type of channel over which the CF is able to operate (e.g. Visual, Auditory, Tactile, Olfactory).

Values: Describing the ability with which the CF is able to transmit or receive information via the modality.

Modality can be divided into a number of media types which can be used to specify the modality more precisely (see table 4.2). As a number of the media types are related to language there is also the potential to describe the language and script of the information being communicated. Description of modality can be taken from existing ISO standards and selection of the standards identified as potentially useful, or already integrated include:

ISO 14915-3:2002 Software ergonomics for multimedia user interfaces – Describing media types.

ISO 639-3:2007 Codes for the representation of names of languages.

ISO 15924:2004 Codes for the representation of names of scripts.

Table 4.2: Examples of Media (Fourney, 2007)

ModalityType	MediaType	Type	Example
Visual	TextWritten	Still image Moving Image	Static Text Marquee
Auditory	Audio	Music	Music
Tactile	Temperature ForceFeedback	Heat Pressure	 Braille
Olfactory	Odour	Odour	Smicons

The accessibility of interactions between two actors is then assessed by matching their abilities to transmit and receive information. The capabilities must share the same modality, be of opposite directions and have compatible values. Figure 4.6 demonstrates the comparison of two capabilities that are measured in terms of a range. Each capability is depicted as a vertical axis against which two arrows are used to identify the maximum and minimum limits of ability. Where the ranges overlap, successful interaction is possible. Where they do not overlap, either the transmitter or receiver is unable to operate, resulting in no communication being possible (e.g. a sound could be too quiet or high-pitched for a person to hear.

While a range measured against an axis is used here for simplicity, values could be provided in a number of different formats:

1. The default form is an upper and lower value with (where appropriate) units of measurement defined. This represents a range, however a single value can be expressed by setting an equal upper and lower value. ISO 80000 provides a list of well recognised units.

2. Where a capability is provided by a taxonomy or standard, it may also dictate the format of the value. For example sound can be measured using a frequency response graph or visual field via a two-dimensional array.

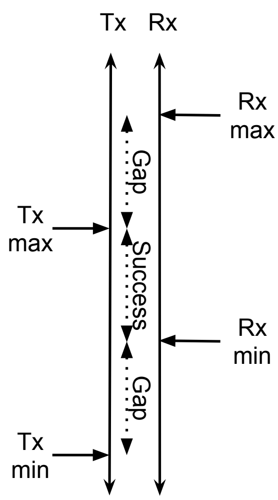


Figure 4.6: A diagram representing the use of standard interfaces to compare abilities between two actors.

The use of interfaces that expose the abilities of both users and technology in a semantically similar way increases the scope for comparing different combinations of profiles. As well as being able to assess interactions between a user and a piece of technology, interactions between two users or two different pieces of technology can also be assessed. If a user and a piece of technology do not have compatible abilities, an AT may be used as a go-between to relaying messages by receiving and re-transmitting them using its own abilities. This then provides the potential for ATs to be assessed against both people and technology, as seen in figure 4.7.

Rather than having two interfaces, the hearing aid in figure 4.7 has two abilities which could both by assessed against a single actor (e.g. a person talking to themselves). However, as the hearing aid has the ability to relay information from its receiving to its transmitting ability, by assessing its ability to interact with two different actors its suitability for use as a 'go-between' can also be assessed.

4.1.2.2 Syntactic Standardisation

In addition to the semantic element of the external interface, the syntax used to expose the data held within a profile can also be standardised. Syntax is the grammatical format used to express data. The VUMS cluster acknowledges that "the desirable common interface language has not yet been defined" (Peissner *et al.*, 2012b) but has successfully used XML in the creation of its central user modelling exchange format. Along with XML there are a number of other existing formats that are already used to provide syntactical standardisation within different profiles.

While the use of an appropriate syntax can improve the human-readability of

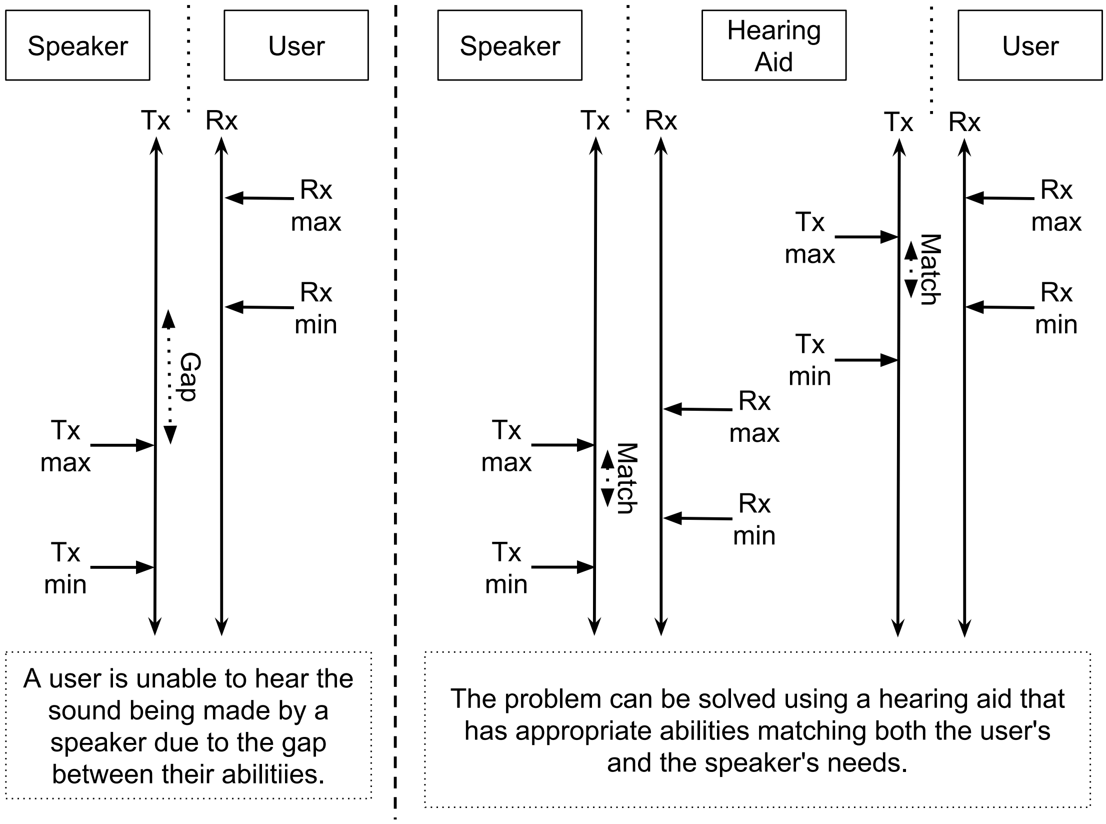


Figure 4.7: Fixing a communication gap using an Assistive Technology.

data by ensuring that it is structured in a logical way, by strictly adhering to the syntax of a technical language machine-readability can also be provided. When models are machine-readable, algorithms can also be written to allow them to be compared automatically as well.

Uniform Resource Identifiers (URIs) provide a method of naming and locating resources across a distributed network (e.g. the Internet). Given the distributed nature of mobile devices, URIs ensure that differing naming conventions and language barriers do not cause problems. Owing to their use within ISO 24751, URIs have been used by the Cloud4All project to identify preference property names in the GPII (Madrid *et al.*, 2012). URIs have also been adopted in the MyUI project through its use of RDF; of particular interest is the storage of all profile information against a URI that is used to represent the user (Wolf *et al.*, 2011).

URIs are not however guaranteed to share a one-to-one mapping with concepts and they have no inherent property for reconciling this issues. As it is possible for different URIs to be generated by different projects to describe the same concept, transferring information or matching between projects is dependent on either the provision of a standard set of definitions or mappings between related URIs. As

URIs can be generated to both name (URN) and locate (URL) resources, by providing information about their source (e.g. reg.gpii.net/common/volume [3]) rules can be generated, e.g. with URIs from certain namespaces being trusted.

Extensible Markup Language (XML) is widely used due to its ability to provide the structure required for machine-readability while retaining human-readability. This is achieved through the use of standardised predefined tags which are used to enclose content which represents the data itself with URIs able to be used for both. As a format, XML is able to store data within a simple semantic structure with schemas being created to describe the meaning of the structure and how the data inside it is to be used. This has resulted in its use as a data exchange format and suitability to provide portable accessibility profiles (Loitsch *et al.*, 2012). In particular it has been chosen as the language of the exchange format specified by the VUMS Cluster Peissner *et al.* (2012b).

The Resource Description Framework (RDF) is a collection of W3C specifications that provide a means of representing information by describing it in terms of its semantic relationships. Due to its transportability and structure RDF/XML is the most popular serialisation format, however, a variety of syntax notations and data serialisations are available (e.g. N-Triples, Turtle, RDFa[4]).

As a format itself it is gaining popularity through its visibility within the Semantic Web (Sosnovsky & Dicheva, 2010). RDF has been also been used for user profiling, with a number of projects taking advantage of the benefits it provides with regards to semantic representation (Ackermann *et al.*, 2012; Iglesias-Perez, 2010; Wolf *et al.*, 2011).

4.1.2.3 Summary

The use of a standard external interface for user, AT and technology models improves their comparability. From the point of view of the interaction model they are all effectively identical black boxes and any two models can be compared to assess the accessibility of their interactions. Anything that is able to take part in an interaction can therefore be referred to as an *actor* and the interaction model will be responsible for comparing different actors to assess their accessibility (as seen back in figure 4.2).

The ability for any actor to be compared against any other actor provides the potential for a single actor to be compared against multiple others with chains

[3] http://wiki.gpii.net/index.php/Discussion_on_Profile_Structure

[4] http://www.w3.org/TR/REC-rdf-syntax/ — http://www.w3.org/TR/n-triples/ — http://www.w3.org/TR/turtle/ — http://www.w3.org/TR/rdfa-syntax/

of actors being created. Figure 4.8 depicts a chain between a user, an AT and a piece of technology. If the AT is able to successfully interact with both the user and the technology it may be suitable for use as a 'go-between'. This is however dependent on the way that the two abilities are connected and as such, more information about the internal structure of actors is required for this to be fully possible.

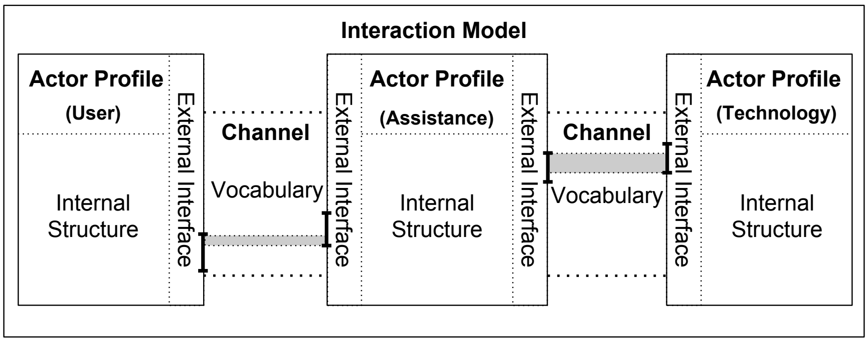


Figure 4.8: An interaction model including three actors.

4.1.3 Internal Structure

The standardisation that has already been described—a common language for comparison within a channel, and the external interface that exposes the information to be compared—are sufficient to allow comparison of the interaction between actors. This section will focus on the standardisation of the internal structure used to represent information within an actor model, as demonstrated by figure 4.9. By understanding the attribution of abilities to actors and the linkages between them the transfer of information between abilities exposed through the external interface can also be understood. As with the external structure, both the semantics and syntax of the internal structure can be standardised.

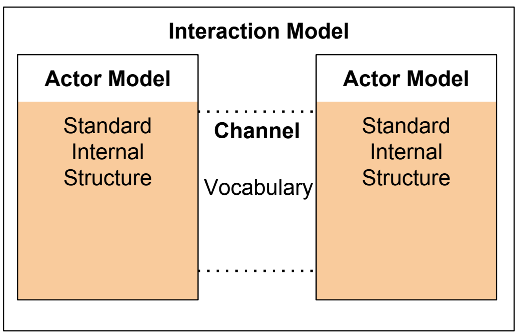


Figure 4.9: Interaction Model – Focusing on the use of standard internal structures.

Common Accessibility Profiles (CAPs) (Fourney, 2007) have already been discussed in terms of their ability to provide a semantically standardised external interface for assessing the accessibility of interactions between actors. They also provide a standard internal structure that can be used to attribute capabilities to actors. The interaction model is represented by an Overall CAP ($CAP_{(O)}$) which is constructed to represent the situation under investigation. The $CAP_{(O)}$ contains a number of actors to be compared, each of which is described using the structure seen in figure 4.10. Actors are represented as Interaction Components ($CAP_{(IC)}$) which have a series of Component Features ($CAP_{(CF)}$) that are described in terms of the properties listed previously.

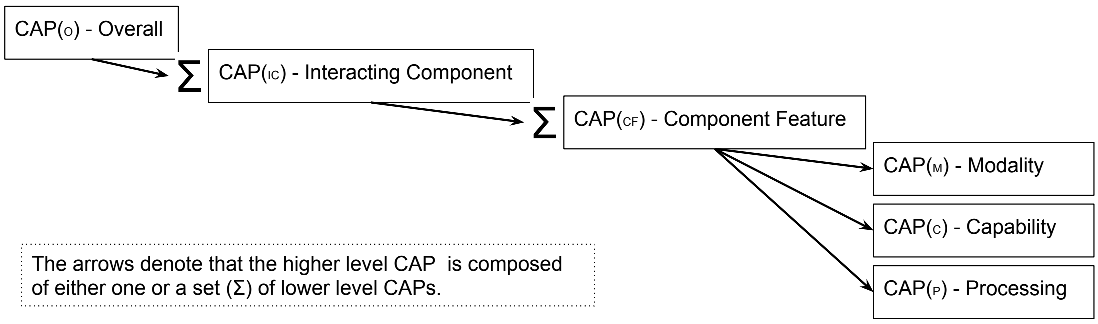


Figure 4.10: The fixed four level structure of the Common Accessibility Profile (Fourney, 2007).

The use of $CAP_{(IR)}$ and $CAP_{(OT)}$ allows information to be transmitted and received, however the ability to chain multiple actors together was also identified as dependent on the actor's internal structure. As well as attributing capabilities to actors, ISO/IEC 24756 (2009) also defines Processing Functions $CAP_{(PF)}$ which are used to provide connections between those capabilities. $CAP_{(PF)}$ are a specialised third type of $CAP_{(CF)}$ which identify a $CAP_{(IR)}$, a $CAP_{(OT)}$ and a transformation that is performed on the data as it is passed between them. Four basic transformations are identified by Fourney (2007) which describes transformation in terms of a change (or not) in modality. It is however also possible that a change could be made to the information while leaving it in the same modality, e.g. increasing the size of text or decreasing the volume of a sound.

1. No change/Pass through ($A \rightarrow A$) [pass-through only];

2. Modify the input ($A \rightarrow B$) [transform only];

3. Modify and pass-through the input ($A \rightarrow A, B$); [both];

4. No output ($A \rightarrow NULL$) [none].

Ontologies – The Web Ontology Language (OWL) is an extension of RDF that provides a syntax for specifying formal semantic relationships and creating ontologies that can be stored and manipulated through the internet. Rather than a single language, OWL is actually a family of W3C standards that are based on RDF and there are different syntaxes available that provide different levels of functionality.

In computer science an ontology is a formal, explicit specification of a shared conceptualisation (Gruber, 1993). "Conceptualisation" refers to an abstract model of a phenomenon that identifies relevant concepts along with their attributes and relations. "Explicit" means that the type of concepts used and the constraints on their use are explicitly defined allowing instances to be identified and recorded. "Formal" refers to the machine readability of the format used to store the ontology. Finally, "shared" reflects the notion that the knowledge captured within an ontology is representative of the views of a group, rather than an individual. Ontologies are used for knowledge representation and storage. By describing both data and the way it is structured they allow it to be exposed in order for comparisons to be made.

Razmerita *et al.* (2003) describes OntobUM, an ontology-based user modelling architecture that demonstrates the application of ontologies for user modelling. The system was aimed at personalised document delivery within a Knowledge Management context and structured around the Information Management Systems Learner Information Package specifications (IMS LIPs) which has subsequently been used to form the basis of ISO/IEC 24751-2 (2008). The architecture uses three ontologies that describe the user, the domain (personalisation applications) and a log of the level of user-application interaction. Data is explicitly captured from the user through a profile editor which provides a guided method of ontology creation. Intelligent services are then used both to monitor the user by feeding information into the log ontology and to provide personalisation by choosing appropriate documents based on their classification within the domain ontology.

A more extensive approach was presented in Heckmann *et al.* (2005) and used to create the General User Model Ontology (GUMO); based on UserML it is an XML user modelling mark-up language. GUMO was developed in reaction to the need for decentralised user models and the problem this created in terms of the potential differences between functional layers used in different models. GUMO attempts to augment traditional user modelling ontology statements with situational statements that provide the additional details necessary to use concepts across ontologies at different functional layers.

Feature-based models use name-value pairs to identify the features that a user has and quantify them and in the same way ontology-based models often use statements in the form $\{Subject, UserModelDimension, Object\}$. The $Subject$ indicates the person that the statement has been made about, the $UserModelDimension$ indicates the statement that is being made (the feature) and the $Object$ acts as a qualifier to judge the subject's adherence to the statement. Two examples would be:

- $\{User, RequiresZoom, 7\}$
- $\{User, AwarenessOfZoom, Low\}$

Although each of the statements provide information regarding a zoom adaptation, the $UserModelDimensions$ do not provide any inherent indication that they relate to the same concept. In addition no information is provided about how to use the objects, specifically with regards to the scale on which they are measured.

GUMO divides the $UserModelDimension$ component into situational statements of the form $\{Auxilary, Predicate, Range\}$. The $Auxiliary$ indicates the use of the predicate in describing the user, the $Predicate$ is the thing being described about the subject, and the $Range$ provides information about scales used for quantification by the $Object$. The previous examples could be expanded to:

- $\{User, \{hasPreference, Zoom, 0-10\}, 7\}$
- $\{User, \{hasKnowledge, Zoom, Low/Medium/High\}, Low\}$

This approach solves both of the problems highlighted in the previous paragraph. The relationship between the statements can be identified as they share a common $Predicate$ and the meaning of the $Object$ can be understood through the use of the $Range$. GUMO allows different statements to be made about the same user providing semantically different views of the same user modelling dimension. The ontology is intended to be extendible[5] in order to provide decentralised user modelling. This may however result in difficulty ensuring consistency and finding instances in the ontology to reuse.

4.2 Abstraction

Abstraction is used to describe the spectrum on which concepts can be placed to describe the level of detail that they contain. OED (Accessed 21/3/14) describes abstraction as "...the process of isolating properties or characteristics common to a number of diverse objects, events, etc...." As a common theme within computer

[5] At time of publishing an experimental version can be found at http://www.ubisworld.org/

science, abstraction can also be applied to the field of modelling for accessibility and used to describe the spectrum on which concepts can be placed to describe the level of detail that they contain. This section will describe two approaches that use the theme of abstraction: hierarchical levels and functional layers.

Hierarchical levels describe the way that higher-level tasks can be built up from lower-level ones. For example, sending an email is a task that is made up of a number of constituents, each of which is a task in its own right. The ability of the user to perform each of the lower-level tasks can be measured (and compared) separately.

Functional layers describe the way that data can be viewed in multiple frames of reference, allowing detail to be hidden. For example, three models were discussed in section 2.5 that described accessibility in different frames of reference. Data models presented a low-level view of the abilities of an individual actor, interaction models compared two actors to describe the accessibility of their interaction and conceptual models took a high-level view, describing how accessibility can be modelled.

Figure 4.11 demonstrates the relationship between these two approaches and provides an example of how they will be combined in section 4.3. Both the amount of time and the number of errors a user makes while typing a set text can be used individually as measures to describe how well they can type. Alternatively they can be combined or (as in the figure) used to produce a ranking which incorporates both measures in a single value. Three functional layers are shown, the structural layer provides information about the relationship between the measures, the data layer holds the scores associated with each measure and the contextual layer describes the context that each score was recorded in.

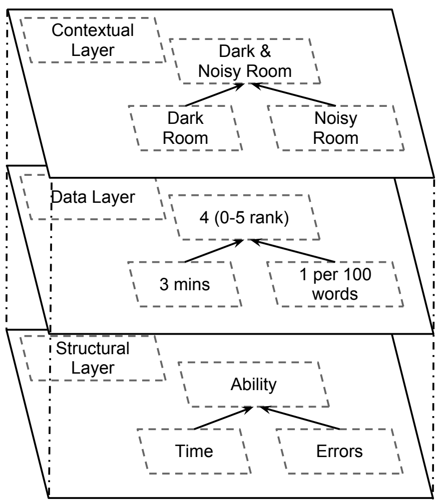


Figure 4.11: Combining hierarchical levels and functional layers of abstraction.

4.2.1 Hierarchical Levels

The use of hierarchical levels is an approach based on abstraction that allows classification by positioning entities in a treelike structure as shown in figure 4.12. Many—if not all—of the models identified in this thesis have some form of hierarchical structure due to the advantages that it affords them. Hierarchical trees provide an organisational structure that allows classification to be performed. Higher-level constructs can be compared based on their underlying (child) components and lower-level constructs can be grouped based on their higher-level (parent) components. By using both of these techniques together it is possible to infer missing data based on data that is available and make comparisons between similar but unrelated items.

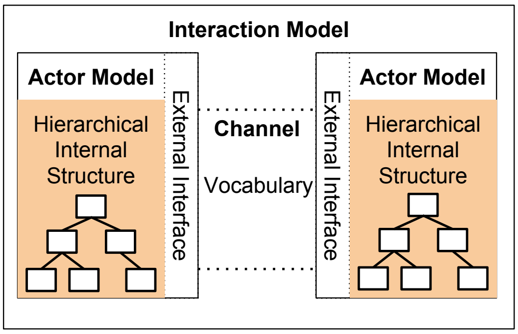


Figure 4.12: Interaction Model – Focusing on the use of a hierarchical internal structure.

4.2.1.1 Hierarchical Task Analysis

Crystal & Ellington (2004) describes Hierarchical Task Analysis (Annett & Duncan, 1967) as an incremental improvement on the basic technical premise. By allowing the decomposition of complex tasks into sub-tasks, a greater understanding of abilities and requirements can be obtained and a greater level of problem diagnosis performed. Hierarchical Task Analysis is described as useful for interface designers due to its ability to model task execution, providing goals, tasks, sub-tasks, operations and plans essential to users' activities. Its strengths are linked to its system-centric stance, viewing tasks as an abstract collection of interlinked goals, resources and constraints. This allows designers to develop performance related metrics against which their designs can be evaluated. It is however restricted in its application as it provides "no systematic way for dealing with the rich social and physical context in which activities are embedded". In addition, although success or failure can be attributed to the components of a system-related task, the approach provides no way to analyse system flows and dynamics.

Rather than a simple comparison, conceptual models attempt to represent the various cognitive processes that are involved in completing a task. The Model Human Processor (Card *et al.*, 1983) provided an ability to connect tasks by modelling them in terms of three interacting systems: perceptual, motor and cognitive functions. Any action taken by a user can be measured in terms of the time taken to perform each element of a task, with the sum being the time taken to complete the task.

Although not a conceptual model itself, the Model Human Processor formed the basis of the GOMS (Goals, Operators, Methods and Selection) family of models. The GOMS model enables a designer to simulate the sequence of actions of a user while undertaking a task by decomposing the task into goals and sub-goals. A number of variants have been created including KLM (a simplified version), NGOMSL (a structured language version) and CPM-GOMS (exploring parallelism in user's actions) (Biswas & Robinson, 2010). Despite their improved ability to model interaction in terms of both predicting and explaining user behaviour, the GOMS models do not consider people with disabilities or non-expert users in detail (Peissner *et al.*, 2012b).

Crystal & Ellington (2004) describes the development of task analysis as mirroring the evolution of HCI research from technical, to conceptual, and finally work-process models. Using this view, Hierarchical Task Analysis becomes an attempt to model the ergonomic, cognitive, information-processing and activity theory's contextual facets. The usability of task analysis is questioned, with complex cognitive models seemingly at odds with simpler practitioner friendly techniques. This results in a trade-off between efficiency (encompassing complexity and usability) and effectiveness (encompassing quality, depth and breadth of output).

4.2.1.2 Fidelity

Biswas & Robinson (2010) poses the issue of fidelity, discussing the hierarchical level of detail that an accessibility model should use. The lower the level of the model the more detail can be provided, however, the more difficult it is to generalise results for use in other applications. In contrast, modelling at too high a level may result in improved transportability at the expense of missing the detail of interactions between underlying components. This dichotomy can be demonstrated in terms of each of the focuses discussed when trying to find a common language for comparison in section 4.1.1. Each of the languages provided a series of benefits that were tied to the level at which they operate and the challenges they face stem from the their use at a different level.

The use of disability-focused impairments and technology-focused preferences provided a means of categorising assistance based on high-level labels. While the labels could provide a general description of the solutions that may be useful for a particular type of user, in reality their application at lower levels resulted in a lack of the descriptive capabilities necessary to describe how solutions should be practically applied between devices. The measurement of interaction-focused abilities however suffered from the opposite problem. Low-level description of abilities and technology-focused tasks provides a method of comparing users and technology based on specific features as their variation provides information about the fit between actors. As the measurement of these features and abilities is best performed using specific tasks in a specific context, this could reduce the applicability of the data collected to other situations, due to their potentially limited use in a different context.

It is possible to describe interaction at a number of levels within the same model and information can be translated between different levels. Biswas & Robinson (2010) suggests that with each level their is a higher chance that errors may be introduced. While this is a valid point, the desire to provide transportability between the models used by the different projects in the VUMS cluster demonstrates that there is a tolerance for reduced accuracy in return for transportability.

4.2.2 Meronomic (Whole-Part) Relationships

Mereology is the study of meronomic (part-whole) relationships and a meronomy is a specific type of hierarchy where higher-level constructs are composed of a series of lower-level parts, which are themselves composed in the same way. Mereology has three axioms, the part-of relation is[6]:

Transitive: "Parts of parts are parts of the whole" – If A is part of B and B is part of C, then A is part of C.

Reflexive: "Everything is part of itself" – A is part of A

Antisymmetric: "Nothing is a part of its parts" – if A is part of B and A != B then B is not part of A.

As a type of semantic relationship, meronomic relationships describes both the way that actors are made up from a series of components and the way that tasks can be built up from a number of constituent sub-tasks.

[6] http://www.w3.org/2001/sw/BestPractices/OEP/SimplePartWhole/

4.2.3 Functional Layers

The use of functional layers is an approach based on abstraction that allows the separation of concepts based on their description using different frames of reference. Hierarchical levels are used to describe the abstraction of different items that are semantically similar. For example a computer, application and interface element are all actors that may have interaction capabilities and using a mouse, pressing a button and flexing a finger are all capabilities that a user may have. Functional layers are positioned on an orthogonal-axis to hierarchical levels, describing abstraction using different semantic views of the same item.

The OSI model (ISO/IEC 7498-1, 1994) provides an example of the use of functional layers to hide the implementation details of technology-based communications systems. Semantically related functions are grouped together within a single layer which is able to provide services to the layer above it and receive services from the layer below it. A single communication can therefore be described in terms of the protocols used and the information contained in each layer. Lower layers are concerned with the way that data is transferred via: the hardware used, the protocols that specify how data is passed between components in a network and the routing of information through a network. Higher layers are concerned with the management of: sessions between applications, conversion between machine dependent and independent data formats, encryption and presentation of data to the user.

The use of abstraction can be observed several times in chapter 2. Firstly, in section 2.1.3 three models were identified that provided different views of disability. The use of each model resulted in a different focus being provided for the definition of accessibility and would have lead to a different solution being produced. Secondly, in section 2.3.6 the use of abstracted interface languages was identified that separated the description of required functionality from rendering information. This allowed functionality to be presented differently depending on the user, device and/or environment present when the interface was used. Finally, in section 2.5.2 a distinction was identified between profiles, models and frameworks; each functioned at a different (increasing) level of abstraction and provided structural guidance to its predecessor.

Although they may be composed of several functional layers, models used for determining accessibility are generally built to expose data at a single layer of abstraction (as demonstrated by both the VUMS projects and the GPII). The VUMS cluster projects were able to share their data due to the central Generic Virtual User Model (GVUM) exchange format. However, while the GVUM allows transportation of data between projects, it does not provide the ability to encode project specific information such as matches between user and task from VERITAS or environmental context from MyUI. There is also no semantic information available regarding the meaning of the structure of the GVUM as it is coded using a (light weight) XML format rather than a more informative (heavier) ontology.

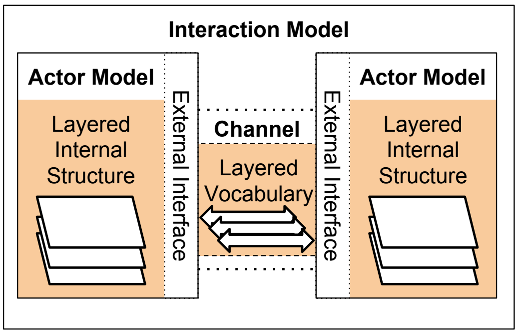


Figure 4.13: Interaction Model – Focusing on the use of functional layers.

4.2.3.1 Ontologies

Although already discussed due to their use for providing a standardised internal structure, ontologies have also been used as a method of organising data storage into multiple functional layers. With regards to ontologies for accessibility related modelling a number of solutions have been proposed.

- AccessOnto takes a high-layer view attempting to merge accessibility guidelines with requirements engineering (Masuwa-Morgan, 2008). It considers neither the capabilities of a user or their accessibility requirements and has no way of specifying or reasoning to find potential accessibility solutions.

- Karim & Tjoa (2006) presents a method of connecting ontologies to allow mappings to be created between a user's impairments and available interface characteristics. It only considers theoretical web interfaces and does not include a reasoning process.

- The projects ACCESSIBLE[7] and AEGIS[8] have both developed ontologies, but these are biased towards providing disability specific personas.

- The MyUI project uses a "context ontology" consisting of three parts: (1) user profile ontology, (2) sensor ontology and (3) application-specific ontology (Wolf *et al.*, 2011). The high-level dimensions proposed by GUMO are rejected in favour of interaction based variables that are measured on a series of ordinal scales which describe the level of impairment of the user.

[7] http://www.accessible-eu.org/
[8] http://www.aegis-project.eu/index.php?Itemid=65

Iglesias-Perez (2010) identified two trends in research related to assistive technology selection. Firstly, taxonomies and ontologies are being favoured over databases, this is likely due to the potential for the semantic linking of ontologies at different functional layers, allowing the connection of ontologies through structural links. Secondly, reasoning is either delegated to a programmed ad-hoc layer, or no reasoning is provided.

As an example, UserRDF is a centralised approach that defines a generic vocabulary for providing information about statements using metadata (Abel *et al.*, 2008). Metadata is data about data and can be used to affect how it is used. As functional layers are characterised by their semantic similarities, the metadata provided by UserRDF can be used to identify common concepts between functional layers and dictate its use. UserRDF originally provided four blocks of properties, defined using GUMO as a base vocabulary.

Main: The content of the statement.

Explanation: Details of how the statement was created and information to calculate its accuracy.

Validity: Information about the expiry including the expected time to live.

Administration: As statements are never deleted, this block allows statements to be tagged as deleted with pointers to their replacements.

A query language (UserQL) was then created to complement UserRDF allowing each of the blocks to be used to dictate how a statement should be used. A UserQL query can be composed of Matching, Filtering and Controlling properties that allow statements to be returned based on their functional layer.

4.2.3.2 Separation of Concerns

By separating data into different functional layers different processing functionality can be performed separately and the complexity of data can be gradually built up through the addition of functional layers. Three layers were identified at the start of the section (structural, data and contextual) with each providing a different type of functionality. While each of the layers represents a type of functionality that can be observed in current models, the list is not meant to be exhaustive, but rather demonstrate the potential for different functionality to be implemented as needs arise. The use of multiple functional layers mirrors the “separation of concerns design principle” that advocates for the separation of data and processing into distinct modules.

4.3 The Resulting Approach

When used in isolation, each of the discussed approaches provides its own benefits and drawbacks that have been demonstrated by the various applications discussed throughout this chapter. When combined however, a novel approach is created that is able to build on the advantages of each of its constituent approaches while negating many of the drawbacks.

4.3.1 The Need to Combine Existing Approaches

As accessibility is a measure of the ability of two actors to interact, its measurement relies on the ability for comparisons to be made between models representing the two actors under investigation. With regards to aim 1 (Functional Assessment of Accessibility), functional comparison is required which identifies either that a match is possible, or the nature of the gap that is apparent.

Standardisation provides the consistency required in order to ensure comparability between models created and manipulated by a wide variety of sources. While this addresses the variety that is identified in aim 5 (Variety of Agents Producing and Using Data), the kinds of rigid standards that have been identified do not allow the scope to represent the variety acknowledged in aims 3 (Variety Within and Interaction Between Accessibility Barriers) and 4 (Variety Within and Interaction Between Accessibility Solutions).

As an alternative to standardisation, abstraction provides the variability required to target comparisons to the appropriate level for the interaction being assessed. The ability to de-construct higher level tasks into their lower level constituents through the use of hierarchical task analysis addresses the variety that is identified in aims 3 and 4. Similarly, the ability to compare data in terms of different functional layers allows the suitability of data for use to be assessed in line with aim 6. Unfortunately, in providing this functionality existing models have had to strike a balance between fidelity and transportability.

4.3.2 Recursive Hierarchical Structure

In response to these problems, a new approach is proposed for assessing the accessibility of interactions that retains the advantages of abstraction while retaining the transportability facilitated through standardisation. The problem with current models that use hierarchical task analysis (e.g. GOMS or VERITAS) is that they rely on a fixed series of hierarchical levels. The use of merenomic relationships implies that although higher level constructs may be composed of a series of lower ones, each of the lower level construct is a construct in its own right.

The combination of abstraction and standardisation results in an approach that is recursive in nature (displayed in figure 4.14). The use of a series of standard elements that are hierarchically related to each other using standard meronomic relationships is proposed, both as a means of modelling communication and as a method of storing data from different functional layers. By modelling users, technology and their capabilities to interact with each other in identical recursive structures, profiles can be created and their potential for interaction compared at multiple different hierarchical levels, depending on the data that is available.

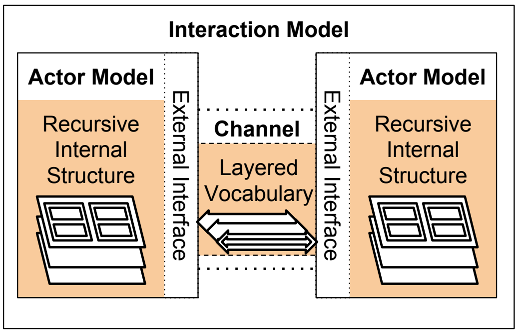


Figure 4.14: Interaction Model – Focusing on the use of a recursive internal structure

4.3.2.1 Modelling Communication

Initially the approach is intended as an extension to basic Communication Theory that will underpin the framework. In chapter 2 the use of channels of communication for assessing the functional ability of people and technology to interact was identified. Multiple channels of communication were identified as existing in order to allow information to be carried in different directions, via different modalities.

The approach uses a recursive series of channels to represent interaction between a user and a piece of technology which are modelled using the same symmetrical recursive structure. Taking the high-level view that accessibility simulation is a comparison between user and technology, there is little to differentiate between their resulting profiles. Rather than modelling them differently, they are both considered as ‘actors’ taking part in an interaction. Various elements within the interaction model can be standardised and a vocabulary based on interaction capabilities allows accessibility to be described using a functional assessment of the abilities and requirements that are being compared. The abilities and requirements are stored in a common ‘internal structure’ and exposed via a standard ‘external interface’.

If a match is observed between the capabilities of the user and technology then interaction is possible using those matching capabilities. When a mis-match is observed, a description of the required assistance is provided in terms of the gap that must be bridged and the capabilities that are available to do so. Rather than relying on the presence of specific technology-based preferences or assistance targeted at a specific labelled impairment, a solution could involve re-routing interaction via alternative (matching) capabilities.

4.3.2.2 Functional Layers

As will be discussed further in the next chapter, an element of recursion will also be applied to the functional layering of data that is stored within the framework. Context can be described in terms of the actors and capabilities at the time data is stored.

4.3.3 Focus of the Approach

Rather than providing another modelling format, the approach will be designed in the form of a framework that aims to guide the provision of modelling formats and profile creation. In this way the approach will initially be provided at a relatively high level of abstraction, an evaluation will be performed by subsequently moving down to lower levels by providing a practical implementation.

4.4 Summary

By combining two approaches that are widely used modelling, a new approach is proposed for modelling interaction in order to assess the accessibility of interactions between users and technology. The new approach is proposed in response to the inability of existing approaches to fully satisfy the aims identified in chapter 2. The approach proposes the use of a series of standard elements that are organised in a recursive structure and represented using multiple separate functional layers.

In the next chapter a theoretical design will be developed in order to demonstrate the potential applicability of the approach to the problem domain and allow the thesis to subsequently identify any potential problems that may hinder its implementation.

Chapter 5

Design

With the approach developed in chapter 4 in place, this chapter will now develop a theoretical framework design which will be used as the basis for the provision of an implementation which will be described in chapter 8. The framework provides a series of particles (components) and relationships—presented in three functional layers—which can be used to create profiles (describing users, technology and assistance) and matching algorithms to be defined allowing the accessibility of interaction between two actors to be modelled. The use of a standard series of components which are connected together by a standard series of relationships means that profiles are scalable and able to represent actors at an appropriate granularity. This provides an ability for accessibility to be modelled at the appropriate hierarchical level of abstraction for the situation under investigation.

This chapter forms the basis of an extended abstract that was included as part of the W3C Research Report on User Modelling for Accessibility (RDWG & W3C WAI Research and Development Working Group, 2012) and subsequently accepted as an invited paper that will be presented within a parallel session on the same subject as part of the 8th International Conference on Universal Access in Human-Computer Interaction (Bell *et al.*, 2014). The components and their relationships will be gradually introduced throughout the chapter in terms of a series of functional layers. Section 5.1 introduces the metaphor on which the framework is based. Section 5.2 will then provide a low-layer structural description showing how the framework facilitates the creation, manipulation and comparison of profiles. Section 5.3 takes a mid-layer view describing the storage of data against the structure and allowing more information to be provided about the quality of the match. Section 5.4 takes a high-layer view describing the use of context to suggest the appropriateness of the matches. Finally, section 5.5 provides a summary of the chapter and describes the other considerations that are needed in order for an implementation to be produced.

5.1 Particles, Atoms and Instances – A Metaphor for Hierarchical Levels and Functional Layers

The design will be described within this chapter using a metaphor which is based on the construction of matter from particles, as used in the natural sciences (e.g. physics). Metaphors such as that of a pipe for communication flow and valve for impairment have previously been used by the Universal Access Reference Model(UARM) (Carter & Fourney, 2004b). Metaphors are used as a means of providing a frame of reference within which new ideas can be placed — in order to make them more easily comprehensible. They are most useful where no common frame of reference is available. A number of metaphors are used within the field of Computer Science, two of the most prominent being those of the use of a Desktop and Windows within a Graphical User Interface (GUI). While they were introduced as concepts to identify the core functionality associated with their technology-based counterparts, they both break down if observed too closely. Additionally, it could be argued that as computers have become more common the metaphors became less needed, owing to GUIs becoming more commonplace.

While similarities can be observed between the framework and the Object Oriented (OO) programming philosophy, the use of a separate metaphor is proposed to avoid any confusion that could be caused through the reuse of vocabulary with subtly different definitions. The OO paradigm introduces the concept of parent and child classes, which provide hierarchical structuring and inheritance, and the differentiation between a class and and an instance, which allows generalisation. It does not however reflect the relationship between parent and child atoms or the way that an instance of a capability can be used in more than one higher-level capability. The framework uses an atomic (or particle-based) structure for the description and storage of data, mirroring the use of hierarchical levels identified in chapter 4. Like all metaphors, the one proposed here provides a frame of reference against which the concepts can be introduced. The efficacy of its use should therefore be measured in terms of its ability to introduce structures and relationships that can be built upon, rather than as a definitive blueprint to be copied verbatim.

In the physical world, matter is made up of a series of particles; basic building blocks that can be combined (using bonds) to create structures and then compared (based on their properties) to identify the similarities and differences between them. Within the framework, the real world is described in terms of a series of particles; discrete pieces of data (or facts) that can be combined to build up profiles representing users and technology. Those profiles can then be compared—in terms of the properties of their underlying particles—to assess the accessibility of their interactions with other profiles.

There are a number of different particles within the physics-based model of the world: (1) atoms are constructed from (2) hadrons (protons and neutrons), which are themselves made up of (3) quarks. The different particles can be described using a hierarchical tree structure that moves from atomic down to sub-atomic particles with each level. Within the framework, a series of basic particles will be defined that will be used to represent different types of data (actors, capabilities, modalities, context etc.) with higher layer particles being comprised of a series of lower layer ones. Unlike the physics-based model however, several of the particles in the framework also have a recursive structure of their own (for example a higher level actor may be decomposed into a number of lower level constituent actors).

The approach described in chapter 4 also identified a series of functional layers. In the physics-based model, atoms and hadrons are particular types of particle; there are then a variety of different types of atoms (hydrogen, helium etc.) and hadrons (protons, neutrons). In the framework there is also variety within particles—as will become apparent when they are described—and the word 'atom' is used a generic term to describe a specific example of a particular particle. Higher-level atoms can be related to lower-level atoms of the same particle type (in the same way that atoms are combined to make molecules in the physics-based model).

Finally, in the physics based model, instances of particles can be identified in order to model their interaction in different situations. Similarly within the framework there may be a number of different 'instances' of any particular atom. As an example the hydrogen and oxygen atoms in a cup of water will be different instances to the atoms in a cup filled with ice. While both the water and steam atoms can provide similar functionality (conducting heat for cooling purposes or hydrating a person) they do so at different speeds. Although the instances are interchangeable (the water and ice could be swapped) their differences may make them unsuitable for replacements for each other. Table 5.1 provides an example of the use of functional layers to differentiate between particles, atoms and instances within the framework.

5.1.1 Representing Particles and their Relationships Within the Framework

Like any other particle, actors are understood through their relationships to other particles. Throughout this chapter each particle will be identified by describing its

Table 5.1: Examples of Particles, Atoms and Instances Within the Framework

Particle	Atom	Instance
Actor	User; Laptop	Mr A; Ms B; Mr A's Laptop; Ms B's Mobile Phone
Capability	Sight; Producing Sound	Mr A's Visual Acuity; Ms B's Phone's Ability to Ring

role within the framework and the semantic relationships that have been provided to link them to the other particles that have already been defined. The framework will be built up gradually using two graphical representations. The first "graph-based view" is designed to explicitly represent the framework in its true form, a series of nodes (particles) connected by edges (relationships). The second "box and graph-based view" has been developed to aid readability by using a series of nested boxes to represent the two most common relationships and reduce the number of edges that need to be drawn.

5.1.2 The '*type*' Relationship

The framework is used to create profiles, which are constructed from a series of particles connected by a series of relationships. The '*type*' relationship allows the designation of data as being part of a given set and as a result it could therefore initially be used to discriminate between particles, atoms and instances. As an example [$data_P$ *type particle*] would identify the piece of data "$data_P$" as being a particle and [$data_A$ *type atom*] would identify "$data_A$" as an atom.

Throughout the chapter a series of different particles will be defined and examples of atoms will be given. As described above, atoms are discrete examples of particles and the '*type*' relationship can also be used to identify a piece of data as being of a particular particle type. Continuing from the example above [$data_A$ *type* $data_P$] would identify that "$data_A$" was of type "$data_P$" which (given the previous definitions) would make "$data_A$" an atom.

5.2 The Structural Layer – Providing Basic Matching

The structural layer is the lowest functional layer of the framework. It provides a series of basic particles and semantic relationships that can be used to represent people, technology and the different forms of assistance that they use to achieve accessible interactions. By the end of this section it will be possible to create basic profiles and compare them in order to identify the potential ways that they are able to interact with each other. The profiles are structures to which more detailed data can be subsequently added in order to allow more informative (higher layer) reasoning to be performed.

The structural layer defines four particles which are used as the basis for representing people and technology. Each particle represents a different element within the interaction model developed in chapter 4 and can be connected to the others by a series of semantic relationships allowing profiles to be constructed.

Actors represent people and pieces of technology that are able to take part in interactions;

Modalities represent the ways in which interaction can take place by describing the media via which information travels;

Capabilities represent the methods by which actors take part in interactions; and...

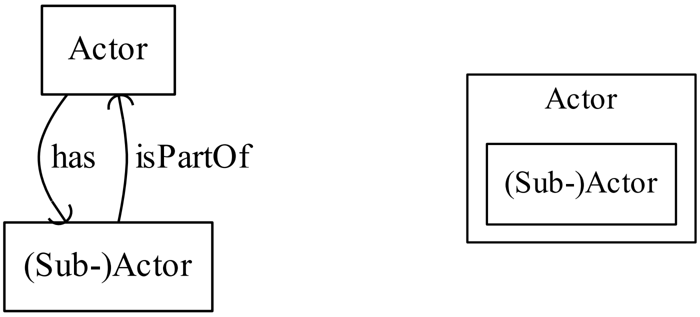
Processing Functions represent connections between capabilities and allow actors to act as assistive technologies by transforming, translating and relaying information.

5.2.1 Actors

Within the framework an 'actor' is a particle that forms a focal point around which profiles can be based. Actors are used to represent entities such as users, technology or assistance and may possess a series of lower-level sub-actors, describing their components. Sub-actors are actors in their own right and can be viewed either as autonomous entities, separate to other actors at their level, or as constituents of higher-level actors. As they are actors themselves, sub-actors may possess a series of sub-actors of their own, demonstrating the recursive nature of the framework.

5.2.1.1 The '*has*' Relationship

The framework defines the '*has*' relationship between two actors to imply that one is a component (sub-actor) of another (a computer '*has*' a screen). All of the relationships within the structural layer are uni-directional and as such, their direction is important. Each uni-directional relationship has an inverse, which represents the nature of the relationship in the opposite direction. The inverse of the '*has*' relationship is the '*isPartOf*' relationship that implies that an actor is a component of another (higher level) actor (a hand '*isPartOf*' a user). The '*has*' and '*isPartOf*' relationships are identical to the meronomic relationships identified in the previous chapter and are represented graphically in figure 5.1.



(b) A box-based representation.

(a) A graph-based representation.

Figure 5.1: Two methods of graphically representing actor particles and their relationships.

When displayed graphically, actors will be drawn using rectangles and the '*has*' relationship is represented either using a black arrow with a curved head or through the positioning of actors within each other. As each of the two relationships can be inferred from each other, only the '*has*' relationship will be drawn unless specifically stated, and the labels will be omitted to avoid clutter within diagrams.

Using the graph-based view, figure 5.1a represents two actors as nodes that are connected by two directed edges. One actor is displayed above the other and while they are both actors, their relationship is such that the lower actor is a sub-actor of the upper one. This representation highlights the links between actors within the framework and emphasises its hierarchical nature, with higher-level actors being displayed above lower level ones. Figure 5.1b depicts the alternative box based view and highlights the meronomic nature of the framework, with the sub-actor displayed inside of its higher level parent actor.

The recursive nature of actors mirrors the real-life construction of users and technology, with both being representable at different hierarchical levels of abstraction. While a device like a desktop computer may appear to be a single entity, it is actually a composite of several sub-devices such as a screen, keyboard, mouse and speakers. Each sub-device can be purchased separately and removed for use within different computer systems. When connected to a computer system a sub-device becomes part of the system, and all of its abilities are attributed to the system as a whole.

An actor and its sub-actors use the same meronomic relationship, with the higher-level actor exposing the abilities of its constituent sub-actors as if they were its own. Sub-actors can then be removed and/or replaced; representing the removal and/or potential upgrading of the sub-devices they represent. The ability to upgrade components of a computer system mirrors one of the methods of providing accessibility described in section 2.3. As an example, it may be beneficial for many people with poor hand dexterity to supplement the mousepad on their laptop with a mouse. While they both perform the same task (indirect manipulation) the mouse and mousepad have different characteristics. The improved speed and accuracy of the mouse (Hertzum & Hornbæk, 2010) imply that it has a lower dexterity requirement and the introduction of a mouse would therefore be equivalent to the use of any other traditional assistive technology.

5.2.2 Modalities

A 'modality' is a particle that is used to describe the medium (or method of interaction) via which information is sent and is equivalent to the vocabulary part of the channel described in the interaction model. Modalities are therefore interaction-based and can range from taking a low-level human-focus (e.g. visual contrast or audio pitch), to high-level composites which resemble tasks (e.g manipulation of an interface widget like clicking a button or dragging a slider). They use the same recursive structure as actors, with higher-level modalities being comprised of a series of lower-level sub-modalities.

5.2.2.1 The '*infers*' Relationship

The framework defines the '*infers*' relationship to describe the ability of a sub-modality to indicate the presence of a higher-level modality. The inverse relationship '*inferredBy*' is also defined to describe the inference of a modality from a sub-modality. Unlike the '*has*' relationship, there is no box-based representation for '*infers*' and '*inferredBy*'. Instead, as seen in figure 5.2, they are depicted solely using a blue edge with a circular 'arrowhead' to denote direction. Similar to the actor's relationships however, modalities are drawn in a hierarchical fashion using pentagons, with higher level modalities displayed above lower level ones. Also, only the '*infers*' relationship will drawn unless specifically stated and labels will be omitted to avoid clutter within diagrams.

5.2.2.2 Inferencing Functions

Similar to the way that an actor can be described as the sum of their sub-actors, modalities can also be described in terms of sub-modalities, which represent their lower level constituents. However, unlike the attribution of sub-actors to actors,

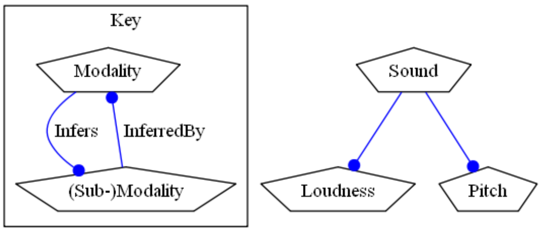


Figure 5.2: Representing modality particles and their relationships.

it is possible that a modality may not be equal to the sum of its constituent sub-modalities. As an example, sound is a modality that has a number of 'aspects' (components) including: loudness, pitch, brightness and bandwidth (Wold *et al.*, 1996). While a sound can be measured in terms of its aspects, they cannot necessarily be used to subsequently recreate it accurately. This phenomenon can be demonstrated in the difficulty faced by synthetic instruments when attempting to recreate the sound of the instruments that they are imitating.

A Human Computer Interaction (HCI) focused example can be provided using work derived from Fitts' (or Fitts's) Law, which describes the time taken to point to a target in terms of the relationship between the target's size and distance. Fitts' Law was developed based on Shannon's Information Theory, which also provided the basis of the Shannon-Weaver Communication Model identified in chapter 2. Multiple "Fitts' equations" have been proposed and as identified in Guiard *et al.* (2011) they all share a common core of variables in the arrangement: $\mu_t = a(\frac{D}{W})$[1]. 'μ_t' is the dependent variable 'time', 'D' is the distance to the target, 'W' is the target's width (one dimensional size) and 'a' is a constant that is used to describe the slope of the curve created by $\frac{D}{W}$ ($\frac{D}{W}$ is often referred to as the Index of Difficulty or 'ID'). For a fixed pointing performance (throughput 'TP') and a particular target (with a fixed ID), a trade-off has been observed between the speed at which the target is selected (measure in terms of μ_t) and the accuracy of the attempted selection (SD_x)[2] (MacKenzie & Isokoski, 2008). Given that an increase in speed results in a decrease in accuracy, pointing performance could be described as an equation in the (simplified) form: $TP = \mu_t \times SD_x$. This would therefore appear to imply that TP is dependent on both μ_t and SD_x and cannot be determined by either one on their own.

While this may be true, there is also a symmetrical relationship between the two halves of the equation $TP = \mu_t \times SD_x$, with an increase in the left hand side

[1]Three of the four listed equations perform a logarithmic function on $\frac{D}{W}$ and all acknowledge the existence of at least one additional constant, 'b'.

[2]MacKenzie & Isokoski (2008) describes SD_x as the standard deviation of selection co-ordinates over a series of trials, meaning that an increase in SD_x actually reflects a reduction in accuracy; in the same way that an increase in μ_t reflects a reduction in speed.

resulting in an equivalent (combined) increase in the right hand side of the equation. Guiard *et al.* (2011) acknowledges that a trade-off exists, but argues that Fitts' Law can be viewed "stochastically" as a mutual dependency between two random variables: movement time and relative variable error. Instead of describing a process that can only evolve in one way, stochastic processes allow multiple intermediate states and outcomes to be generated using a function that is dependent on time (Papoulis & Pillai, 2002, 285). Although the investigation and use of stochastic functions falls well outside of the scope of this thesis, they demonstrate the potential complexity involved in the use of inference for combining modalities and moving up or down a hierarchical tree. While a relationship could be defined to describe functions that are suitable for providing inferencing with regards to a given modality, there is the potential for different systems to use different inferencing methods and existing methods are also evolving. As such, the definition of inferencing functions will not be provided by the framework, however, this provides the potential for further research with examples such as ConcurTaskTrees providing a basis on which the hierarchical models of the VERITAS project were built.

5.2.3 Capabilities

A 'capability' is a particle which is defined by the framework to allow actors to interact with each other. Capabilities can be used to either transmit or receive information and are described at this layer of abstraction in terms of their direction and modality[3]. The capabilities of different actors can then be paired in order to allow directed channels of communication to be formed and investigated between actors. As such capabilities are the result of a desire for standard external interfaces within the interaction model.

As they are used to describe actors' abilities to interact via modalities, a number of different relationships can be defined; firstly as a to describe their use of modalities, secondly with regards to their recursive nature themselves and thirdly to describe their attribution to actors.

5.2.3.1 The '*inputsVia*' and '*outputsVia*' Relationships

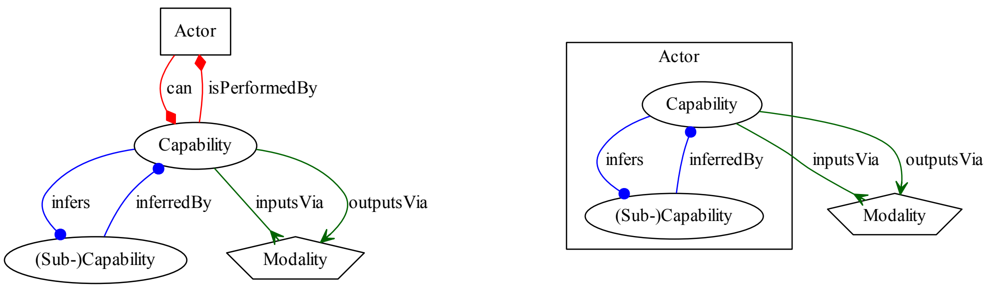
A channel can initially be created between two capabilities if they have the same modality and opposing directions. Although the quality of interaction that the channel can sustain is dependent on them having overlapping bandwidths and a shared context of measurement, given the structural nature of this layer this information is not provided until sections 5.3 and 5.4 respectively. Capabilities

[3]At higher layers of abstraction a measure of their bandwidth (ability) and context of measurement can also be given.

are therefore described at this level in terms of whether they are used to transmit or receive information and the modality via which they are used.

A capability's direction and modality are described simultaneously using one of two symmetrical relationships,'*inputsVia*' and '*outputsVia*'. While they are still uni-directional, unlike the relationships described previously *inputsVia*' and '*outputsVia*' are used to provide a link between two different particles, a capability and a modality. Although the two relationships are used symmetrically in the definition of a channel they are not the inverse of each other as they are both directed from a capability to a modality. The relationships describe a capability as transmitting or receiving via a modality and as shown in figure 5.3 are depicted using a green arrow with a "vee" style[4] arrowhead that point either towards or away from the modality.

The '*outputsVia*' relationship implies that a capability is able to output (or transmit) via a given modality (vibration '*outputsVia*' movement) and therefore has an arrow head that points away from the capability (towards the modality). As it is uni-directional its inverse relationship, '*isInputedViaCap*', can also be defined, however as demonstrated with previous relationships there is no need to provide a specific depiction for it. The '*inputsVia*' relationship implies that a capability is able to input (receive) via a given modality (hearing '*inputsVia*' sound). It has an arrowhead that is positioned next to the modality (to depict the uni-directional nature of the relationship) and points back towards the capability. Once again, its inverse relationship, '*isOutputedViaCap*', can also be defined, but no formal depiction has been identified.



(a) A graph-based representation.

(b) A box-based representation.

Figure 5.3: Graphical representations of capability particles and their relationships.

[4] http://www.graphviz.org/content/arrow-shapes

5.2.3.2 Reusing the '*infers*' Relationship

As they are linked to modalities, capabilities can be described as sharing the same hierarchical structure, with higher level capabilities being inferred through the presence of lower level constituents. This is depicted within the framework through the reuse of the same relationships that were defined for modalities, '*infers*' and '*inferredBy*' as can be seen in figure 5.4.

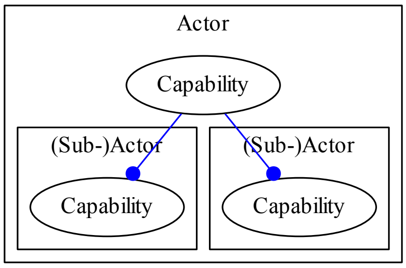


Figure 5.4: The graphical representation of capabilities between and within actors.

5.2.3.3 The '*can*' Relationship

As capabilities are used in order to describe actors' abilities to take part in interactions, a relationship is required to describe the attribution of capabilities to actors. An actor may have more than one capability, allowing it to take part in multiple channels of communication simultaneously in either one or both directions. The '*can*' relationship is defined to imply that an actor possesses a capability (a microphone '*can*' receive_sound). It is another uni-directional relationship between two different particles and as seen in figure 5.3, can be depicted using either the graph-based, or box-based representations. Its inverse '*isPerformedBy*' describes the relationship in the opposite direction (sight '*isPerformedBy*' an eye).

Figure 5.3a depicts a graph-based representation, which uses a red edge with a diamond shaped arrowhead pointing from the actor to the capability. By displaying it underneath the actor (in the same position as a sub-actor) the capability can be seen as being attributed to the actor. It can be noted at this point that the sub-capability does not appear to be attributed to an actor. However the '*infers*' relationship can be used to infer attribution, given the higher level capability can be used to imply the existence of the lower level one.

While this provides a simple way of attributing multiple recursively linked capabilities to a single actor, the introduction of sub-actors and the need to identify which abilities can be performed by an actor versus its sub-actors could quickly

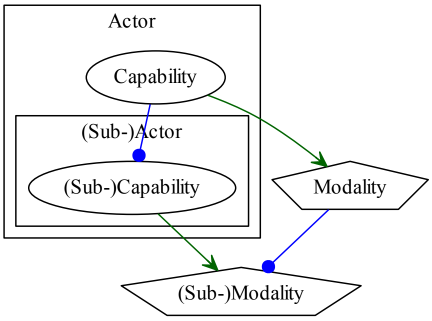


Figure 5.5: A box-based representation of sub-capabilities and their relationships.

result in overcrowding. Instead, using the box-based view as in figure 5.3b allows all of the '*can*' and '*infers*' relationships to be represented explicitly. The sub-capability is clearly within the bounds of the actor and is therefore attributed to it. If the actor had a sub-actor (as in figure 5.5) the attribution of sub-capabilities could be made explicit. For example, if the sub-actor were removed, the actor would no longer have the sub-capability (it could however be provided by another sub-actor).

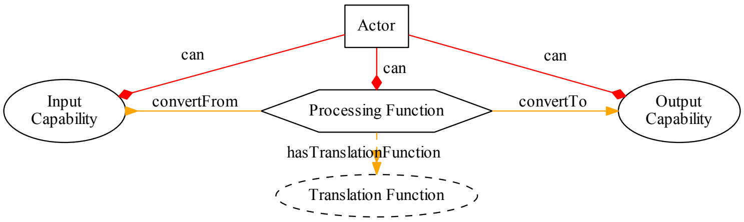
5.2.4 Processing Functions

Processing functions are used to provide links between capabilities, connecting them together and allowing an actor to translate or transform information from one capability to another. The framework does not distinguish between assistive technologies and any other kind of technology. In fact, as any actor with capabilities in both directions can potentially be used as an assistive technology this allows people to be modelled as providing assistance (e.g. moving a mouse in response to verbal commands).

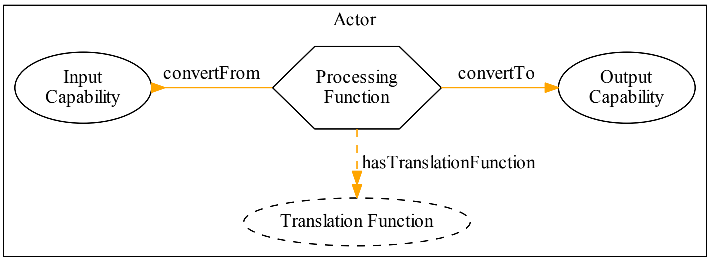
The example in section 5.3.1.2 described the way that the text produced by an application can be outputted by a device either via a screen (using a visual modality) or a speaker (using an aural modality). While the example is a translation from one modality to another, other forms of assistance (e.g. a hearing aid) provide transformations; receiving and retransmitting information using the same modality.

5.2.4.1 The '*convertFrom*' and '*convertTo*' Relationships

The framework uses a series of relationships to describe processing functions (as presented in figure 5.6a). As already described, the structural layer is not con-



(a) A graph-based representation.



(b) A box and graph-based representation.

Figure 5.6: The processing function particle and its relationships.

cerned with the actual information that is being processed, and as such the discussion of translation functions will be revisited in section 5.3.2.

As processing functions link two separate capabilities, two relationship are required, with each used to link the same processing function to a different capability. The '*convertFrom*' and '*convertTo*' relationships are defined in order to identify the capabilities that the processing function is linking between. The use of two separate uni-directional relationships highlights the directional nature of processing functions (and communication), with information travelling in via a receiving capability and out via a transmitting capability.

As seen in figure 5.6 both relationships are depicted using a yellow directed edge starting at the processing function. The '*convertFrom*' relationship uses a 'normal' arrowhead point back from the input capability and the '*convertTo*' points towards the output capability. Like the '*inputsVia*' and '*outputsVia*', the '*convertFrom*' and '*convertTo*' relationships are symmetrical and while their inverse relationships could be defined, the existing relationships are sufficient for the purposes of this work. The inverse relationships (once named) could be be inferred and vice-versa from those listed here.

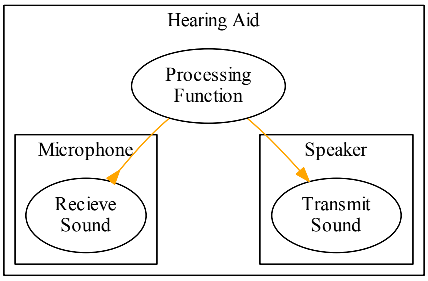
5.2.4.2 The '*can*' Relationship

As processing functions deal with capabilities, they should be attributable in the same way that capabilities are. For this reason the '*can*' relationship is reused, as demonstrated in figure 5.6. In the graph-based view (figure 5.6a) the use of multiple '*can*' relationships can be observed, owing to the way that each of the atoms involved in processing may be attributed to a different actor. Once again, the box-based view (figure 5.6b) is able to provide the same information without the clutter of the three '*can*' edges being drawn. Both the processing function and the capabilities are displayed as attributes of the actor.

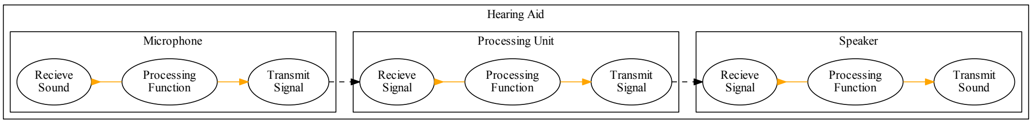
An example of a hearing aid is used to demonstrate the use of processing functions as a means of abstracting the internal workings of an actor and the importance of attribution. A hearing aid is a piece of assistive technology that takes in sound through a microphone, manipulates it, and outputs it again via a speaker. While it can be broken down into its constituent elements and described in terms of their ability to interact via electrical signals, a higher-level view could be taken which uses a processing function to connect directly between the microphone's input capability and the speaker's output capability. This situation is represented in figure 5.7a and shows how the input capability (receive sound), output capability (transmit sound) and processing function are all attributed to different actors. If either of the sub-actors were removed (or damaged) the processing function would not have the necessary capabilities available for it to link between. Alternatively, without the processing function, the capabilities of the two sub-actors would not be linked and while the hearing aid would be able to interact using their capabilities, it would not be able relay information between them.

5.2.5 Summary

The structural layer provides the elements necessary to allow profiles to be constructed, which represent different users or pieces of technology. Through the use of common underlying capabilities, the ability of a user to use one piece of technology can be used to imply infer similar abilities in other pieces of technology. Figure 5.8 shows three different top level actors, their sub-actors, and capabilities, which are all reliant on the same, common, underlying capability.



(a) A graph-based representation.



(b) A box and graph-based representation.

Figure 5.7: The use of processing functions inside a hearing aid.

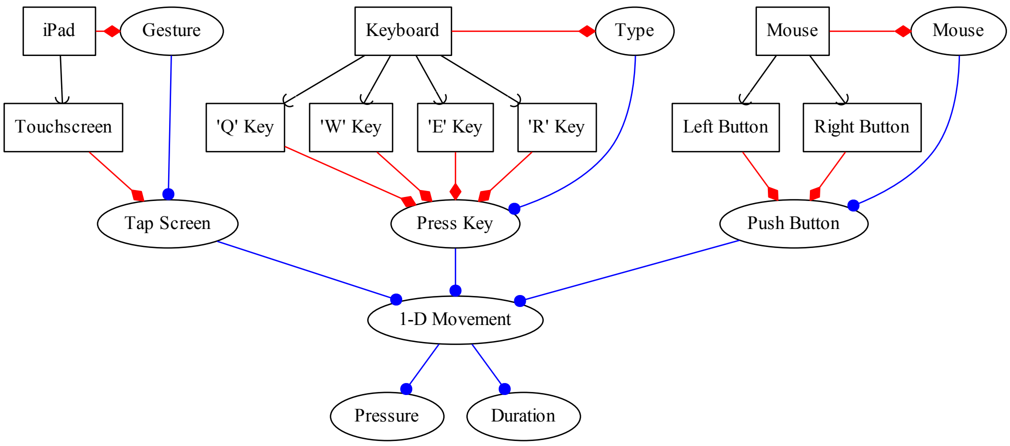


Figure 5.8: The attribution of capabilities to actors at multiple hierarchical level including inheritance.

5.3 The Data Layer – Describing the Quality of a Match

The data layer is the second functional layer defined within the framework and deals with the attribution of data to profiles. Through the particles and relationships provided within the structural layer it was possible for basic profiles to be created that could be compared in order to identify the existence of channels of communication. Section 5.2.3 suggested that one or more channels of communication could be constructed between two actors if they have capabilities with (1) the same modality and (2) opposing directions. From this information it was possible to identify the potential modalities by which two actors could communicate, based on their matching capabilities. While this basic information indicates the potential for communication to take place, it does not provide any information about the quality of the communication that a channel can sustain.

The quality of the communication that is possible between two capabilities can be measured in terms of the bandwidth of the channel constructed between them. The bandwidth of that channel is then based on the amount of overlap between the abilities of its two capabilities, as shown in figure 5.9. The figure shows a channel of communication being constructed between two capabilities, which are depicted in the same way as the standard interfaces in figure 4.6. Each capability has a bandwidth within which it is able to operate, the upper and lower bounds of which are indicated by the arrows; the wider apart the arrows the wider the potential range within which the capability can operate. Interaction is only possible where capabilities ability ranges overlap (e.g. the blue shaded region). While the wider 'pipe' is labelled as a channel, only the blue shaded region is usable for the transfer of information between the capabilities.

The greater the overlap the greater the bandwidth of the channel. The greater the bandwidth of a channel, the more information it can potentially carry; the more resilient it is against noise; and the higher the quality of interaction it can sustain. The provision of ability data therefore allows more informative matching to be performed.

Once potential channels of communication have been identified (as in section 5.2), the comparison of ability data can be used to provide a measure of the quality of communication possible via each channel. As the amount of information that can be produced about a match increases, so does the layer of abstraction at which the match takes place. For this reason, while the data in this section builds on that provided in the previous one, it has intentionally been separated in order to allow processes to be provided that are specialised for the functions of each layer.

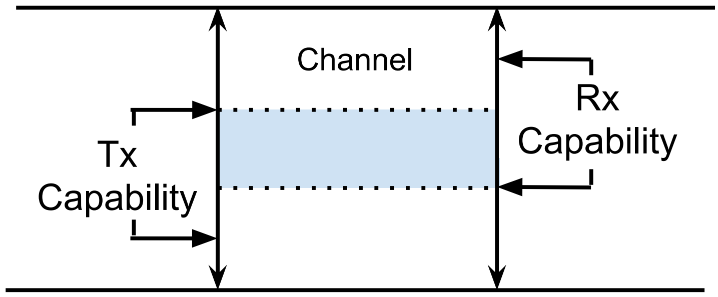


Figure 5.9: A channel of communication between two capabilities.

5.3.1 Data Items

A capability is a skill that allows an actor to transmit or receive information and was described at the structural layer in terms of its direction and modality. At the data layer, that description is extended to include a quantitative measure of the actor's ability, which can be used to provide more information about the quality of the match. Data about actors' abilities is stored within the framework using 'data item' particles and each capability may have one or more data items attached to it, in order to represent the variance in an actor's abilities in different situations.

5.3.1.1 Data Item Construction

Rather than dictating a single standard for measuring abilities, the framework has been designed to be extendible, allowing both existing and future measures of ability to be incorporated. While the hierarchical nature of the framework allows the use of a static low-level vocabulary that is based on human capabilities. Owing to the rapidly evolving technology landscape, the framework also has the ability to accommodate a more dynamic high-level vocabulary, based on technology-focused tasks.

Unlike structural-layer particles, due to the variety that exists between the different capabilities that may be measured there is the potential for variation in the construction of data-layer particles. Each data particle is therefore constructed as a self-contained atom; incorporating both a quantitative measurement of ability and the information necessary for the measurement to be interpreted and used appropriately. Information can therefore include the scale against which the ability was measured (including the measurement units) and the type of measurement that was taken. Information regarding its accuracy and context of measurement can also be stored, however as that is part of the contextual layer it will be described further in section 5.4.

5.3.1.2 The '*hasData*' Relationship

Like actors, the data item itself is a collection of lower-level particles which are connected to the data item identifier by a series of relationships. The '*hasData*' relationship has therefore been defined to link a capability to a data particle identifier. The relationship is uni-directional and as a result its inverse, '*isDataFor*' can also be defined to imply that a data particle provides data about a given capability.

5.3.1.3 Scales of Measurement – The '*scale*' Relationship

In order for a capability to be measured, a scale is required to allow the measurement to be expressed with regards to a shared point of reference. Modalities have already been defined as a point of comparison, allowing transmitting and receiving capabilities to be compared against each other. Although a capability will be measured in terms of its modality, a single modality can be measured on multiple scales or levels of measurement.

Field (2009) identifies five levels of measurement split into two categories. With each additional level of measurement more meaningful comparisons can be made to produce more informative results.

Categorical

Binary: A special case of nominal data with only two categories.

Nominal: A series of categories with no inherent ranking.

Ordinal: A series of categories with a logical order.

Continuous

Interval: Equal intervals on the scale represent equal differences in the property being measured.

Ratio: The same as an interval scale with the ability for meaningful ratios (which is reliant on a meaningful zero point).

Different levels of measurement can have have different functions performed on them. The structural layer was able to describe a channel using binary (existence and direction of a capability) and nominal (modality of a capability) data. From this, only basic matching could be performed to generate binary (existence of individual channels) and nominal (what modalities) information about the accessibility of any particular channel. The aim of the structural layer is to increase the level at which channels are measured in order to provide more informative assessments of their accessibility.

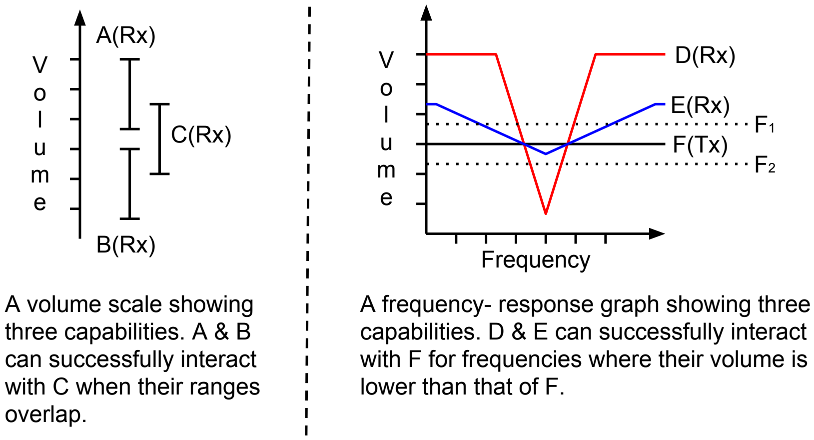


Figure 5.10: The comparison and measurement of capabilities using different measurement scales.

As an example, a user's hearing and a speaker's ability to produce sound are both capabilities that can be inferred from a number of sub-capabilities, as described in section 5.2.3. Those sub-capabilities can be measured via the audio-based modalities described in section 5.2.2 (e.g. loudness, pitch etc.) using their associated scales (loudness/volume for example is measured on a logarithmic scale). Although there are well defined single axis scales for measuring its sub-modalities, capabilities that transmit or receive sound tend to be measured on a (two axis) frequency-response graph. As a two dimensional representation, it could be argued that there is greater scope for interpretation with a frequency-response graph than numerical readings taken from a single axis scale. The measurement of higher level capabilities, where sound is only one element (for example the production or perception of speech, which combines both sound and language), then becomes even more subjective and difficult to measure.

Figure 5.10 shows six capabilities, three measured against a single axis and three against two axes. Capability $\mathbb{A}$ has the same size range as capability $\mathbb{B}$ and capability $\mathbb{D}$ has a similar area (if it is made into a triangle) to capability $\mathbb{E}$. While $\mathbb{A}$ and $\mathbb{B}$ cover different ranges of their ability spectrum, as they have the same size range their abilities could be described as being similar. They also overlap $\mathbb{C}$ by the same amount and so the likelihood of accessible interactions between $\mathbb{A}$ and $\mathbb{C}$ could be described as being the same as those between $\mathbb{B}$ and $\mathbb{C}$.

With $\mathbb{D}$ and $\mathbb{E}$ however, although their areas are similar, their profiles are dissimilar; $\mathbb{D}$ is able to hear sounds that are more quiet over a narrower range of frequencies and $\mathbb{E}$ being able to hear louder sounds over a wider range. Their comparison is therefore less straightforward, with both having different strengths and weaknesses. Again the accessibility of their interactions with $\mathbb{F}$ appear to be the same as they both have sufficient abilities over the same (central) range.

Interactions between $\mathbb{D}$ and $\mathbb{F}$ could be described as more robust than between $\mathbb{E}$ and $\mathbb{F}$ because $\mathbb{D}$ exceeds $\mathbb{F}$ by a greater amount than $\mathbb{E}$ (if a problem occurred which reduced the volume of $\mathbb{F}$ to $\mathbb{F}_2$). On other other hand however, $\mathbb{E}$ has a greater potential for increased accessibility, given that increasing the volume of $\mathbb{F}$ to $\mathbb{F}_1$ would result in a greater overlapping range than $\mathbb{D}$.

Rather than using a multi-axis measurement scales, higher level capabilities could be assessed through a more simplistic 'ability rating' Likert scale (e.g. from 0–5 where 0 is poor and 5 is good). The use of a more simplistic scale however introduces subjectivity, as a decision must be made as to what constitutes appropriate ability at each level of the 0–5 scale. Each additional degree of freedom that is produced by increasing the level at which capabilities are measured therefore increases either the complexity of the scale of measurement or the subjectivity of measurements. As a result the '*scale*' relationship can created to allow the identification of the scale of measurement from which a measurement was taken.

5.3.1.4 The '*dataType*' Relationship

In the same way that data may be collected in multiple levels of measurement, data can also be collected in a number of different formats. Returning to the example of sound, in the previous subsection it was suggested that a capability could be measured on either one or more axes depending on its level of measurement. At the lowest level of measurement, the existence of a capability could either be confirmed or denied and as its level increased, data was recorded in a greater level of detail. Similar to the level of measurement, a data item's data-type describes the amount of data that is stored.

Returning to the sound-based example from the previous subsection, a capability that is able to transmit or receive via the modality 'volume' can be stored in various ways. To get a realistic measurement of an actor's ability, a more informative data-type is needed, such as a frequency-response graph, however, in order to reduce the complexity of comparison functions, a lower order data-type could be used. A range of descriptive statistical techniques are available (e.g. $\bar{x}$ *and s.d.*, maximum and minimum) to provide more condensed snapshots of a more informative data set. Finally, the lowest level of measurement (a binary value) provides the lowest amount of useful data.

Binary: Single value as in section 5.3.1.3 [Does the actor have a given capability?].

Point: Single value at any other level of measurement [What is the actor's ability? (e.g. maximum, minimum, average, Fletcher-Munson curve)].

Range: Minimum and maximum values [What are the tolerance levels of the ability?].

Other: More complex or specialist data [Providing a more complete representation (e.g. Descriptive statistics, frequency response graph, visual field matrix (Humphrey Matrix))].

The framework defines the uni-directional '*dataType*' relationship to describe the attribution of a data-type to a data item. The data-type can then be used to identify both the format in which the value is stored and the set of functions that are appropriate for performing comparisons. An *AND* function could be used for binary values, equalities or inequalities used for points ($A > B$), the degree of overlap for ranges and whether a point is inside of range where there is a mix. More specialist comparison functions can then be created for more complex datatypes, e.g. statistical comparisons such as the the t-test for means, correlations for data sets (Field, 2009).

5.3.1.5 Units of Measurement – The '*units*' Relationship

In order for a comparison to be made between two values, they must not only be measured against the same scale, but must also be recorded in the same units of measurement. Multiple units may be identified for the same modality on the same scale, e.g. font size can be measured in points, picas, millimetres, or pixels, amongst others[5]. While approaches for describing units of measurement are available (Berrueta & Polo, 2011; Rijgersberg *et al.*, 2013), for the purposes of this discussion the critique of any specific approach can be considered out of scope. Instead it is the ability to identify units by providing a pointer to a description that is required. The uni-directional relationship '*units*' can therefore been defined to describe the relationship between a data item and the units of measurement it uses.

In addition to identifying standard methods of representing units, standards also exist to detail the units of measurement that are appropriate for different situations (ISO/IEC 80000, 2009). Where possible these should be adhered to, however, the higher the level of the capability (and therefore modality) the greater the likelihood that its measurement scale will suffer from some form of subjectivity.

5.3.2 Translation Functions

In section 5.2.4, processing functions were described as a means of linking two capabilities in order to model the ability of actors to translate or transform information. In the structural layer, the existence of a processing function between

[5] http://smad.jmu.edu/shen/webtype/measure.html

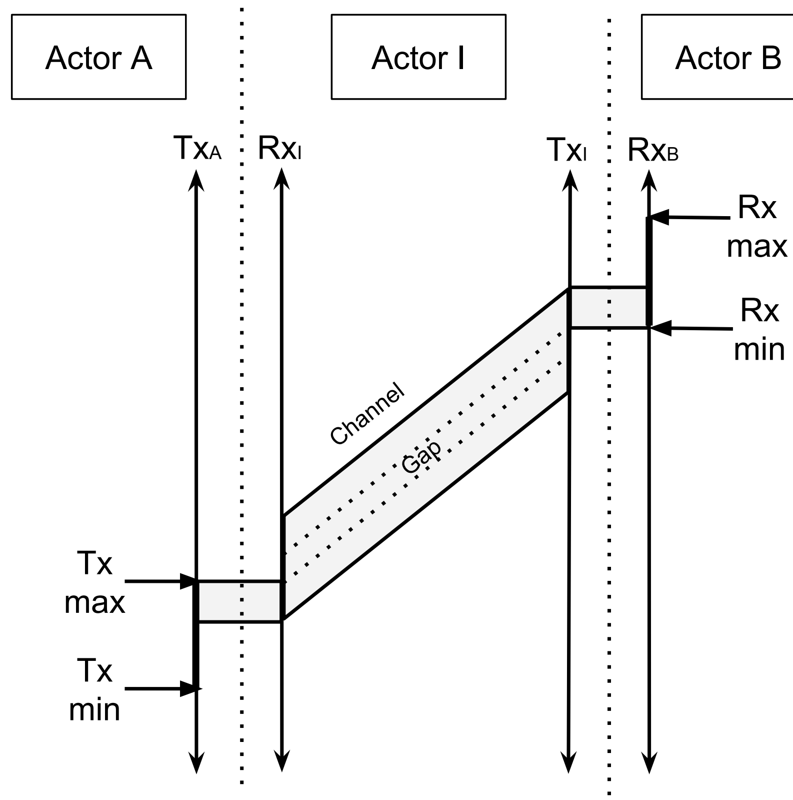


Figure 5.11: The use of a translation function within an extended channel of communication.

two capabilities not only implied that information could be passed between them, but also that any two actors that could interact with the capabilities, could interact with each other through the use of the actor containing the processing function. However, as demonstrated in figure 5.11 this may not be the case.

In figure 5.11, actor $\mathbb{A}$ is not able to communicate with actor $\mathbb{B}$ as their abilities do not overlap. Actor $\mathbb{I}$ has therefore been placed 'in between' them owing to the overlap of its capabilities with those of $\mathbb{A}$ and $\mathbb{B}$. As $\mathbb{I}$ provides a direct translation from its input capability (Rx_I) to its output capability (Tx_I), the range from which its potential input may come (the overlap between Tx_A and Rx_I) will not be translated to a value in the suitable output range (the overlap between Tx_I and Rx_B)

A function is therefore required to describe the nature of the translation or transformation between the two capabilities, providing a mapping from one to the other. There are two ways for mappings to be provided, either in the form of a table (similar to the approach used for routing network traffic), rules or functions (such as interpolations). Owing to the discrete nature of lower order data types (such as the mapping between pt to em when measuring font size) tables or rules are required. For most continuous data however, interpolation function can be provided. Where mappings resemble a function there are three common types can identified:

Linear: Transposes values from one range into another maintaining an equal ratio between equivalent points within each range.

Logarithmic: Transposes values from one range into another whereby a ratio of linear values on one scale is set equal to a ratio of logarithmic values on the other.

Inverse: Transposes values from one range into another, reversing its order.

5.3.2.1 Construction and Storage

Not only will different functions be required to represent the different mappings between different capabilities, but it may be possible for the mappings themselves to be dependent on external factors for which more complex translation functions could be created. It is therefore intended that translation functions are provided as separate modules in order to allow them to be imported and reused within the framework. Different functions could then be developed by experts in different domains (e.g. audiologists or technology manufacturers/developers) and provided differently depending on the system that uses them.

Local Storage and Processing: Basic functions such as linear interpolation which may be either be commonly used, or used as a fall-back function can be built in, being stored and used by the system locally.

Remote Storage, Local Processing: Alternatively, where functions are more specialised they could be stored by their creators and downloaded as needed.

Remote Storage and Processing: Finally, as the systems that use them may be unconcerned with their inner workings, translation functions could be provided as black boxs with values being entered and a result provided. If this is the case functions could be both stored and processed remotely.

5.3.2.2 The '*hasTranslationFunction*' Relationship

As the functions themselves may vary no attempt has been made to formalise their structure, instead a relationship has been identified describe the mapping between the two capabilities that are being linked. The '*hasTranslationFunction*' relationship is used to identify the translation function that describes the mapping and attribute it to the processing function. In figure 5.6 the relationship was depicted as an orange edge with a double arrow pointing towards the translation function. This is representative of both the link that the relationship provides and its use as a pointer towards the actual translation function itself, wherever it is stored. The function can then either request or expect, the data it requires and then return it again in an appropriate format.

5.4 The Contextual Layer – Appropriateness of a Match

As the highest layer of the framework that will be defined in this thesis the contextual layer deals with the accuracy and appropriate use of the data that is stored in the lower layers. Although the structural and data layers allow the creation of profiles and subsequent attribution of ability data with which to model interactions between people and technology, no facility has as yet been provided to understand the accuracy of those models or the data itself. As identified in chapter 2, the context surrounding an interaction can have a significant impact on its accessibility. The contextual layer therefore provides the ability for contextual information to be attributed to data items (in the form of context particles); potentially providing three functions within the framework.

As ability data may be collected from a range of different sources, contextual data provides a means of understanding the situation from which each data item was taken. The comparison of ability data is dependent on the similarity of the scales on which the data is recorded. If a person is described as being able to read text that is greater than size 12pt, the description of a screen as outputting text of size 14px is not immediately useful. While conversion functions are readily available[6], a data item that explicitly described the screen's text size in points (10.5pt) would be more easily comparable.

In a similar way, the appropriateness of comparisons between different data items is dependent on their contextual similarity (particularly with regards to the situation that is being modelled). The greater the contextual similarity of data items that are being compared, the greater the accuracy of their comparison is likely to be. The first function is therefore related to the ability of attributed contextual data to be used to identify the accuracy of any comparisons that are performed.

The second function builds on the first to inform the selection of data where more than one is available. Through the attribution of multiple data items to a single capability, the abilities of an actor can be represented in multiple contexts. As the accuracy of comparisons is calculated based on contextual data, where

[6] http://reeddesign.co.uk/test/points-pixels.html

multiple data items exist for a single capability, contextual layer data can be used to identify the most suitable data item for a particular purpose.

Finally, as interaction is dependent on the context in which interaction takes place, where no data has been collected for a particular situation, existing data could be with the expected change in ability being based on the change in context. For example, contrast sensitivity is reduced after exposure to high levels of environmental light (which is due to the reduction in pupil dilation). An actor's abilities with regards to a given light level can then be interpolated based on existing data, by identifying the difference in light levels between the given level and that recorded in the existing data item.

5.4.1 Categories of Context

In chapter 2 context was identified as encompassing a wide variety of factors that surround an interaction, all of which can influence its accessibility in different ways. Throughout this section, the interaction of two actors (a subject and its interaction partner) will be used as a running example. Each actor will have a series of capabilities which will have data items with both ability and contextual layer data appended to them. The framework identifies three different types of context (atoms), grouped according to their relationship to different areas of the interaction model: (1) the Subject (an actor for whom an accessibility assessment is being performed), (2) their Partner (the actor they are interacting with) and (3) the Environment in which the interaction (channel) is taking place.

5.4.1.1 Subject Context

Subject context describes contextual information relating to the state of the subject actor when a specific capability is measured. Ability data describes an actor's ability with regards to a single capability, however actors are capable of performing multiple capabilities. The ability (or inclination) to perform any one capability is dependent on the other capabilities that an actor has.

As an example, a person using a mouse and keyboard simultaneously is likely to demonstrate a lower typing speed than a person solely using a keyboard due to the overhead of switching between the devices. As a result higher typing speeds are likely to be recorded during periods where mouse related skills are reduced. ISO/IEC 24756 (2009) addresses this by allowing Component Feature CAPs to define the presence of one or more contiguous channels of communication. In terms of this example, a person with kinaesthetic abilities attributed to a single hand cannot use them to manipulate a mouse and keyboard at the same time.

As the framework is able to work at different levels of abstraction to ISO/IEC 24756

(2009), a different solution is required. Unlike the CAP, the framework allows the differentiation between individual capabilities that are attributed to different sub-actor components of an actor. By storing the state of other capabilities at the time an ability measurement is observed, a better understanding of the capabilities of an individual actor can be provided. The trade-off between mouse and keyboard ability is based on the limitation of the kinaesthetic abilities of an actor. Higher level capabilities may also suffer a trade-off that is based on the actor's 'cognitive' abilities. For example a user's ability to read may be reduced if they are also attempting to listen to music or spoken word and a device may run more slowly due to limitations caused by the speed of its processor or size of its memory.

5.4.1.2 Partner Context

Partner context describes contextual information relating to the state of the partner actors when a subject's specific capability was measured. Similar to the way that a subject actor's specific ability may be dependent on the abilities of their other capabilities; the ability may also be dependent on the partner with whom the subject is interacting. As an example when talking to person (or device) with impaired hearing, many people raise their voice, demonstrating an ability for producing louder sounds than during normal speech.

While ability data can be provided through self-reporting, any form of measurement by an external actor (e.g. human-mediation or autonomous agent as described in section 2.4) takes place through the observation of an interaction. Vision or hearing tests measure ability through the observation of a person's ability to perceive a series of images or sounds that are transmitted to them and typing speed can be observing the time taken or errors made by a person when entering a prescribed piece of text. While this provides an objective method of measuring ability, the ability data that is recorded is inherently tied to the context of the measurement method.

Returning to the example from the previous section, by requiring a user to occasionally interact with an interface using a mouse (through the attribution of appropriate receiving capabilities to the actor representing the interface) the interface presents a situation that results in a lower typing ability being exhibited by the user. Partner context therefore aids the understanding of ability data by providing information about the context surrounding its measurement.

5.4.1.3 Environmental Context

Finally, as well as the two actors involved in an interaction, the environment in which the interaction takes place can also have an effect on its accessibility.

Chapter 2 identified how environmental factors such as light and sound can disrupt channels of communication by introducing noise. In Weaver (1949) noise is described as interference that distorts information as it passes between a transmitter and receiver. Due to the potential for the message that is received to be different from the one that was transmitted, redundancy or error checking is required, increasing the size of the message. Over a channel with a fixed bandwidth, this has the effect of increasing the time taken for the message to pass through the channel.

In order to pass a noise resistant message at the same speed as its original message, an increased bandwidth is required. Either this can be achieved by increasing the bandwidths of the capabilities themselves or the interaction will be negatively affected. If capabilities are be measured through the observation of their use in interactions (as described in the previous section); the environment within which ability data is recorded will dictate its suitability for use in other situations. Ability data that is recorded in an environment that is conducive to interaction will not reflect the actor's abilities in a more aggressive environment.

Given the effect that it can have on channels and their constituent capabilities, environmental context could be described in terms of the modalities of the channels it affects and the size of the effect it has on them. The combination of modality, direction and a value is identical to that used for representing the capabilities that are stored in the other two types of context and as a result, environmental data can be stored in the same format.

5.4.2 Representing and Attributing Context – The '*hasContext*' Relationship

As they provide a description of the state of a subject actor and their interaction partner at the time that a piece of ability data was collected, actor and partner context are both constituted from a series of capabilities. Similarly environmental context can be described in terms of the capabilities of the environment and exposed in a format compatible with a user's (or technology's) abilities to perceive them. As they are all collections of capabilities, each of the three types of context can be modelled using the same structure and vocabulary as an actor. The recursive nature of the framework is then reiterated as by representing context as an actor, any functions that are created to compare capability data can be re-purposed for the comparison of contextual data.

Although three types of context have been identified, the extensible nature of the framework allows other types to be defined as they are identified. As an example, Brusilovsky & Millán (2007) identifies a spectrum of contextual dimensions which range from device to the user centred. As users and devices can be represented as actors within the framework, the actor and partner context already provide the ability for many of the items identified and similarly the environmental aspects are also representable within the framework. There are however other dimensions including social and task/goal based context. While they have not been considered in this chapter, the generic structure of actors, capabilities and modalities could be used as a means of representing any feature-based data.

5.4.3 Comparison of Contextual Information

While each of the types of context defined in this section will be described as being of '*type*' context, it will be a separate atom, representing data from a different source. As such, in order to compare different types of context, the retrieval of data from each of the sources will be required as described below and shown in figure 5.12.

Subject: The capabilities of the subject actor themselves.

Partner: The capabilities of the actor that the subject is interacting with.

Environment: The environmental conditions in which the interaction is taking place.

5.5 Summary

This section has provided a theoretical design for a framework that is based on the approach developed in the previous chapter. Rather than a mandating a static format, the framework defines a series of particles that represent the different types of information that may be required in order to describe the interaction abilities of people or technology. Particles can then be linked together through the use of a series of semantic relationship, which allow them to be organised into profiles.

The accessibility of interaction between two actors can be assessed by matching their capabilities and the potential to include contextual information allows a measure of the accuracy of the resulting prediction to be provided. Where higher-level capability data is unavailable for a match to be made, inference can be used to recursively find a lower level at which matching constituent capabilities are available. As contextual data is stored in the same format as users and technology, contextual matching can also use similar algorithms.

Where a match is not possible owing to incompatible capabilities, an alternative route can be identified. Alternatively, speculative augmentation of actors with

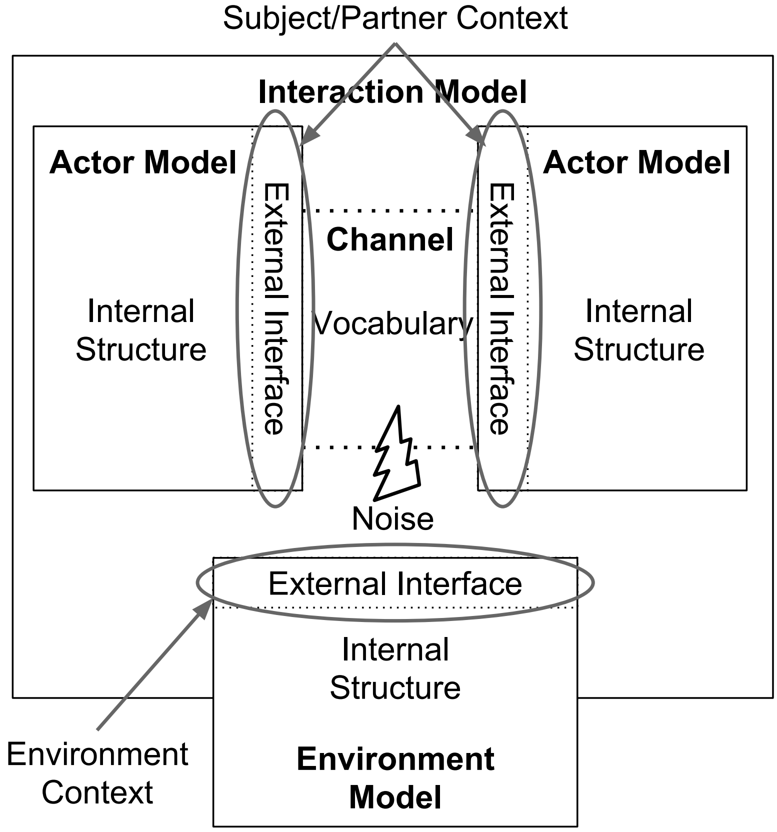


Figure 5.12: An interaction model highlighting sources of context.

representative sub-actors can used to identify the effect of different potential forms of assistance. As an example, both hearing-aids and text-to-speech functionality can be described in terms of the way they re-route and transform the existing channels of communication. Software can then be modelled on different devices by placing its highest representative actor within the actor describing each device.

5.5.1 Contribution

Through the production of a theoretical design, this chapter addresses several of the aims identified in chapter 2. The ability to dynamically create profiles that describe a range of people (responding directly to aim 2) and technology. The use of common interfaces to expose their interaction capabilities allows functional accessibility assessments (aim 1) to be performed and a range of impairments to be identified (aim 3. Through the ability to model the flow of information through different actors the effect of different accessibility solutions can then be identified, responding to aim 4.

The definition of a standard series of standard elements and relationships provides a common method by which profiles and information about their applicability to different contexts can be described in a standard way. While this

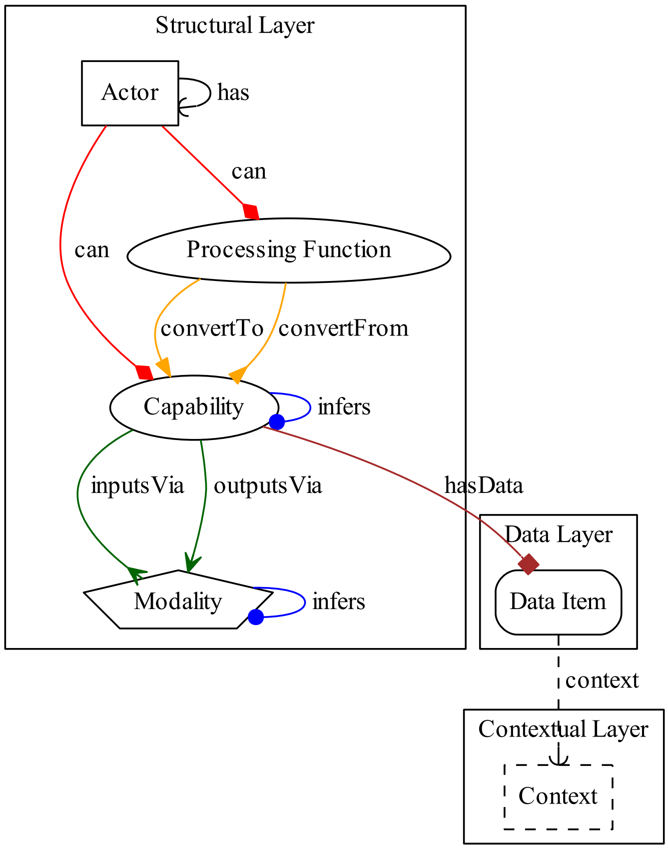


Figure 5.13: A generic graph of the particles and relationships.

partially responds to aims 5 and 6 it does not address the practical collection and use of data within a working system. As a result, the next chapter will investigate the collection of data for use in a system that is based on the framework.

Chapter 6

Self-Efficacy Study

With a theoretical framework in place, the question of data acquisition is raised. The previous chapter acknowledges that data may not be available at the appropriate hierarchical level for a match to be made and proposes the use of inference to re-purpose existing data. It does not however consider the variation in accuracy of data caused by the use of different methods of provision.

This chapter describes a study that has been performed to investigate the ability of older people to provide the data required to drive a model based on the framework. Section 6.1 provides the rationale for the study and develops a series of hypothesis which are tested in 6.3 using a methodology developed in section 6.2. Thirty-three participants between the ages of 52 and 88 ($\bar{x} = 66.15$, $SD = 10.64$) were recruited in three data collection sessions. After providing informed consent, participants completed a technology inclusion survey and a computer anxiety questionnaire. Participants' ability to use four different input devices was assessed using an objective measure which was compared against subjective judgements for the device and objective measures for other devices to investigate their relative predictive power.

Descriptive statistics, correlation analysis and linear regression were used to describe the data and identify relationships between variables. Comparisons between computer anxiety, technology inclusion, age and gender were all consistent with existing research, suggesting that the dataset was representative of the wider population. Comparisons were then carried out between subjective and objective measures for a single device and objective measure of different devices to assess their usefulness in predicting ability. Although statistically significant correlations were observed between subjective and object measures for three of the devices, a lack of consistency reduced their usefulness as predictors of actual performance. In comparison, correlations between objective measures from different devices provided consistently better statistical significance and higher correlation coefficients. Section 6.5 concludes that the use of objective measures of performance, as provided by automated or semi-automated agents are more appropriate for providing data than users' subjective judgements.

6.1 Rationale

This section will describe the rationale behind the study and describe how it fits into the thesis as a whole. After detailing its relation to the thesis' aims, existing computer self-efficacy literature will be used to build up a series of hypotheses for which a methodology will be developed.

6.1.1 Relation to Thesis Aims

Chapter 2 described how many people—especially older people—may benefit from some form of assistance in order to improve their ability to access technology. The problem that was identified was two-fold: (1) many older people are unable to identify that they require assistance and (2) as a result, they do not have the knowledge required to identify suitable assistance. Chapter 5 provided details of a theoretical framework that has been designed to identify the accessibility of interactions between a user and a piece of technology. While it addressed both the second problem above and many of the aimss identified in chapter 2, it makes the assumption that all of the data used within the framework is accurate at the point of provision.

Section 2.4 described three potential 'agents' that may interact with a system that is based on the framework: the user themselves, human mediators and (semi-)automated technology-based agents. Each agent had different advantages and disadvantages, with one of the factors dictating use being the quality of data they they were able to provide. In terms of the thesis' aims, aim 6 identifies how an appreciation of the variety between data sources is an important consideration. This chapter is therefore intended to respond to aim 6 by investigating the accuracy of different data collection methods. This will be achieved through the comparison of subjective judgements and objective measures in order to identify their ability to predict actual performance.

6.1.2 Self-Efficacy

As described in chapter 2 self-efficacy is the strength of a person's belief in their own abilities, for example to complete tasks or reach specific goals (Bandura, 1978). Rather than being a measure of actual ability, self-efficacy can be used to measure a person's ability to estimate their own abilities. This means that a person's self-efficacy may not accurately reflect their actual abilities with regards to a task; however as constructs, they are intimately linked. As shown in figure 6.1 self-efficacy is both dependent on, and able to affect, actual outcomes.

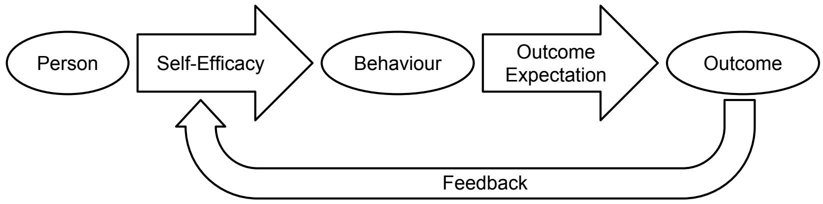


Figure 6.1: A diagram describing the effect of self-efficacy on outcome (adapted from Bandura (1978).

A person's self-efficacy judgements are formed in relation to a number of influences including: vicarious experiences (observing others), verbal persuasion (what they are told to expect) and emotional arousal (how they feel in response to situations involving the task). The strongest influence is however reserved for a person's own experiences, meaning that self-efficacy is largely dependent on previous experience. This means that where a person has previously had positive experiences and success in relation to a specific skill, they will have higher self-efficacy. Conversely where a person has previously exhibited low levels of ability, their self-efficacy is likely to be low, leading to an expectation that previous negative experience will be repeated if they re-attempt a task requiring the skill in question.

A person's self-efficacy can therefore dictate their behaviour (Bandura, 1995). The desire to avoid negative experiences (as predicted by low self-efficacy) will lead to avoidance or an unwillingness to attempt a task, based on the belief that failure is likely. As a result, low self-efficacy is reflected in behaviour; reducing tolerance to failure and removing a person's desire to engage with a task, leading to an outcome of poor actual performance. High self-efficacy, however, leads to an expectation of success, increasing engagement with an activity, resilience to setbacks and willingness to persevere through failure. This results in a person with high self-efficacy getting more practice and consequently improving their actual ability, reinforcing their self-efficacy judgements (Chang & Ho, 2009).

6.1.2.1 Predictive and Evaluative Self-Efficacy

Given that self-efficacy is partially dependent on a person's previous experiences (as described in section 6.1.2 and denoted by the feedback loop in figure 6.1) it can be inferred that the level of experience a person has will affect the accuracy of their self-efficacy judgements. This can be directly demonstrated by contrasting prediction (the process whereby a person attempts to guess how well they will perform in the future) and evaluation (an attempt to assess how well a previous action was performed). In the first instance, it would appear that predictive judgements should be less accurate than evaluative judgements, as an increased amount of experience is available to inform the latter. However a person's existing self-efficacy can result in behaviour which is more likely to lead to the predicted result (self-fulfilling prophesy) and unexpectedly positive or negative outcomes can lead to disproportionate evaluative judgements being created (through the feedback loop) Bandura (1978).

6.1.3 Computer Self-Efficacy

Self-efficacy is a broad construct which can be measured at both a general and task-specific level. Compeau & Higgins (1995) describes computer self-efficacy as a measure of a person's perception of their own abilities with regards to using computers and other technology based tasks. As with general self-efficacy, computer self-efficacy is a single construct that is comprised of three dimensions: magnitude (the attainable difficulty an individual perceives themselves as being capable of), strength (the person's conviction in their perceived magnitude) and generalisability (the breadth of situations to which the perception can be applied). Increased computer self-efficacy increases the likelihood of computer use and is therefore a significant predictor of older people's technology inclusion (as described in section 2.2.2).

Carroll *et al.* (2005) suggests that the link between self-efficacy and performance can result in self-efficacy judgements being used as a means of interrogating complex capacities where it is difficult to measure actual performance directly. In terms of capability measurement for use within the framework, this suggests that older people's self-efficacy judgements could provide an appropriate source of data. Existing studies examining the use of self-efficacy judgements for determining ability have however previously focused on tertiary level students. Sieverding & Koch (2009) for example describes a study in which students (with age $\bar{x} = 23.7$ and $S.D. = 4.2$) were asked to watch other people attempting to solve a home computer problem. Roughly half of the participants were psychology students and the other half were from other disciplines. Participants were then asked to rate their own ability to solve the same problem and the study (which focused on gender) observed a significant difference in the self-efficacy of men and women. A history of literature can be identified as proposing a link between computer self-efficacy and perceived ease of use, however the link is not empirically proven as ease of use does not necessarily reflect competency (Hong *et al.*, 2002).

Larres *et al.* (2003) suggests that self-assessment is often driven by lack of resources and this is certainly true of the motivation for asking users to self-report. In the study, objective and subjective measures of computer literacy were obtained from entry level undergraduate accounting students and used to assess the validity of self-assessment. Subjective judgement were found to be a poor predictor of students' actual ability. A later study by the same authors re-confirmed the result, however it did suggest uncertainty within the field and the regular use of self-efficacy to assess computing ability (Ballantine *et al.*, 2007).

Whilst self-efficacy may initially suggest a positive correlation with performance, overly high or low self-efficacy can be detrimental, owing to the need for adjustment based on outcome feedback Moores & Chang (2009). Ballantine *et al.* (2007) identifies a correlation between self-assessment accuracy and skill, with less-able students over-estimating their abilities. Given the stereotypical low abilities expected in the ageing population, it is likely that self-efficacy will be a poor predictor of actual ability for this age group. In order to measure this, a comparison between an objective (control) measure and a participant's predictive and evaluative (subjective) judgements will be used to investigate its suitability and this gives rise to the first hypothesis:

$H1_1$: No significant correlation is observable between older people's subjective judgements and the control measures of performance.

The framework is designed in chapter 5 on the premise that higher level capabilities can be compared in terms of their underlying sub-capabilities. Where the control is unavailable this implies that data from related objective measures can be used to predict the control measure:

$H2_1$: Older people's alternative objective measures of performance is positively correlated with their control measure of performance.

In another study investigating predicted ability, students in an introductory tertiary eduction business computing class were asked to predict their final grades at the beginning of the course (Baxter *et al.*, 2011). Self-efficacy proved to be a poor predictor of outcome compared to the objective measure (grade-point average) as in Larres *et al.* (2003). Given this information and the first two hypotheses this suggests that objective measures are preferable to subjective measures as a source of data:

$H3_1$: Older people's alternative objective measures provide a better prediction of performance than their subjective judgements.

6.1.4 The Use of Other Factors to Predict the Self-Efficacy Accuracy

Although self-efficacy judgements are potentially unreliable for use as a method of predicting performance, computer self-efficacy is linked to a number of inter-related constructs which may allow better predictions to be made. Baxter *et al.* (2011) identified how gender, age and previous technology training were significant predictors of grade outcome in an introductory tertiary level business computing class. In the context of an ageing population technology inclusion will be substituted for technology training as many older people have not had any formal training. Technology inclusion (the amount of technology that a person has been exposed to) is positively correlated with self-efficacy (Moores & Chang, 2009). This would suggest that the more technology a person has been exposed to the higher their perception of their own abilities. While this perception may not reflect reality—given the self-correcting relationship identified between self-efficacy and actual ability in the previous chapter—information about technology inclusion should improve the accuracy of self-efficacy judgements.

There is also a well researched relationship between self-efficacy and computer anxiety that partly explains the above relationship. While self-efficacy describes perceived ability, computer anxiety can provide an indication of a person's reluctance to engage with technology. If a person has high anxiety levels, they are less likely to engage with technology and more likely to have a lower level of inclusion – reducing the experiences on which they can draw and the subsequent accuracy of their self-efficacy judgements (Charness & Boot, 2009). Increased computer anxiety is therefore negatively correlated with self-efficacy. Due to the relationship between anxiety and technology use, people with higher anxiety levels are less likely to engage with technology, reducing the potential for practice, which could increase ability. As described in section 6.1.3 the feedback loop that is present between outcome and self-efficacy (based on behaviour) can create a self-fulfilling prophecy that is due in part to the negative effects of computer anxiety.

The above relationships between self-efficacy, computer anxiety and technology inclusion contrast with the finding in Ballantine *et al.* (2007) that people with lower ability levels tend to over-estimate their ability. Although at odds with the above relationships, the finding can be understood in the wider terms of the reduced accuracy that is inherent in low technology inclusion. If for whatever reason a person does not have especially high computer anxiety, it is possible that their ability estimates will suffer from a natural over-confidence that may be present either in their personality, or due to a favourable experience they have had. As well as reducing ability, a lack of inclusion reduces the experiences from which a person has to draw and therefore the potential for their inaccuracy self-efficacy to be corrected through the feedback loop.

Given the stereotypically lower technology inclusion of older people, the corresponding increase in computer anxiety should lead to a reduced self-efficacy being observed Olphert *et al.* (2005). Older adults tend to underestimate their abilities (Karavidas *et al.*, 2005). This may be due to a number of factors, however the popular view is that it is a combination of computer anxiety and low technology inclusion. There is however an issue with using chronological age as a measure, given the variability of individuals' ageing characteristics.

Finally, there is evidence to suggest that males tend to rate themselves more highly than women of equivalent skill (Henry & Stone, 1995; Hoffman & Vance, 2007; Sieverding & Koch, 2009). This would suggest that either men have a tendency to over-estimate, or women to under-estimate, their abilities. The finding could be explained by gender differences in the use and perception of technology, as described in Karavidas *et al.* (2005), however a later literature based study found difference to be less significant (Wagner *et al.*, 2010). Acknowledgement of this potential bias could be used to adjust self-efficacy assessments if its existence can be confirmed.

The above discussion leads to the adoption of the following three hypothesis. Given the importance of the related factors in determining computer self-efficacy, examination of the factors may lead to an improvement in the ability of self-efficacy judgements to reflect actual ability:

$H4_1$: Inclusion of participant factors will improve the ability of older people's subjective judgements to predict performance.

The introduction of participant factors may improve subjective judgements owing to their ability to better describe the participant. As alternative objective measures provide a measure of a participant's objective ability, the addition of participant factors should not have the same effect on predictive power:

$H5_1$: Inclusion of participant factors will not improve the ability of older people's alternative objective measures to predict performance.

With the addition of participant factors to subjective judgements, their combined predictive power will be increased. Given the previous results in Baxter *et al.* (2011) however, alternative objective measures are still expected to provide a better prediction of ability:

$H6_1$: Older people's alternative objective measures will provide a better prediction of the control measure than the combination of subjective measures and factors.

6.1.5 Rationale Summary

This section has presented the rationale behind the study and developed the six hypotheses ($H[1–6]_1$) to be tested. The hypotheses allow a methodology to be developed in order to collect data which will either support or contradict them. Each of the hypotheses will then be tested through the creation and evaluation of a series of null hypotheses ($H[1–6]_0$).

6.2 Methodology

In order to test the hypotheses a cross-sectional study is proposed that will investigate the ability of older people to predict and evaluate their own ability to input data to a computer. Subjective self-efficacy judgements will be compared against objective measures of ability when using alternative devices. This will allow the relative accuracy of each method of assessment to be identified and therefore appropriateness of data provided by users to be evaluated.

6.2.1 Variables

In order to test the hypotheses developed in section 6.1, a number of variables are identified and either controlled or observed. Self-efficacy is a measure of a person's ability to accurately judge their own ability. The hypotheses investigate the suitability of self-efficacy as a measure of ability by comparing it against objective measures and examining its dependence on a number of factors. Participant's self-efficacy is therefore identified as a dependent variable and technology inclusion, computer anxiety, age and gender are all independent variables.

In order to focus examination on the relationship itself, the existence of a number of external variables is recognised. The participants themselves, the tasks used to provide objective measures of ability and the environment in which the study is conducted are all therefore controlled.

6.2.1.1 Participants

Participants can be thought of as compound variables and described in terms of a number of attributes which are all variables in their own right. A range of participants were needed to provide a comprehensive snapshot of the population under investigation and allow the variety found within the ageing population to be represented in the data collected for the study. A number of constraints were also used to reduce the potential for biases created by external factors (confounding variables (Field, 2009)).

As identified in the previous section, computer self-efficacy is related to age, gender, technology inclusion and computer anxiety. The process of recruiting participants was therefore designed to provide variety in these variables, leading to a range of self-efficacy judgements being observed. Three different data collection sessions were used, allowing recruitment from different demographic groups to take place.

Mickleover is a village/suburb of Derby, recruitment took place within a church community and focused on the younger end of the ageing spectrum.

Rawmarsh is a large village in the Metropolitan Borough of Rotherham, recruitment took place within an older people's social group and focused on older people with low technology inclusion.

Dundee is a city in Scotland, recruitment took place within an older people's technology interest group and focused on older people with high technology inclusion.

In accordance with the best practice toolkit produced by the Sus-IT project, "gate-keepers"[1] were used to access each of the groups (Damodaran *et al.*, 2012) and ethical procedures were followed to ensure the safeguarding of participants. As a particular population was required, a purposive selective sampling method was employed to find participants. Each gate-keeper provided access to a community which represented a different part of the population under investigation. While selective in nature, the use of 'snowball sampling' was appropriate given the difficulty of accessing the desired population.

Task-Specific Computer Self-Efficacy Computer self-efficacy as a construct was been discussed in section 6.1.3 and can be understood in greater detail through an investigation of its three dimensions: magnitude, strength and generalisability. The 'generalisability' dimension of computer self-efficacy judgements weakens their use in terms of predicting the outcome of specific tasks and so task-specific judgements will be collected. Given that self-efficacy judgements are dependent on participants' previous experiences (as described in section 6.1.2), it is possible for judgements to change in response to feedback as experience is gained (Moores & Chang, 2009). This provides the opportunity for comparison between predictive and evaluative self-efficacy, where the former is a judgement made before attempting a task, and the second is one made after a task has been attempted. Both types of judgement were therefore collected and stored independently during the study.

[1] people who could facilitate access to a community under investigation

Predictive Judgements were measured before the participant attempted the task.

Evaluative Judgements were measured after the participant had attempted the task.

Age As the study focuses specifically on age-related issues a lower age boundary was chosen to differentiate 'older' from 'younger' people. The use of age as a measure of the ageing process is rarely informative (as described in section 2.2.2). Controlling for chronological age does however allow the exclusion of participants based on certain experiences that they may have shared, reducing the potential for additional extraneous variables. One of the reasons for a widespread increase in computer literacy is thanks to its use as an educational tool, with specific computer skills tuition provided by schools. According to (Rettie, 2002), people born after 1977 have experienced an increased exposure to technology during their younger years which may have influenced their acceptance of new technology. This may be due in part to the increased use of technology in schools. Restricting participants to those who are over forty will therefore remove the effect of school-based IT education (Bliss *et al.*, 1986; Molnar, 1997).

Further, while a fifty year old person may be considered to be leaving middle-age, they are likely to be computer literate, either owing to work-based experience or personal necessity. Although advanced in years, at fifty the ageing process will not have made a significant impact ensuring that the full spectrum of old age is investigated, from those about to enter the ageing process to those well into it (Hawthorn, 2000). For this reason recruitment was constrained to people over the age of fifty.

Ability, Technology Inclusion and Computer Anxiety The study is investigating ability, which is be dependent on a number of age related as well as other factors. The definition of disability identified in chapter 2 does not differentiate between disabled and non-disabled people allowing two discrete groups to be identified. The use of assistive technology by an individual could be used as an indicator of disability, however the ubiquity of hearing-aids and glasses would require an arbitrary differentiation between acceptable and unacceptable assistive technologies. The use of any form of assistive technologies could however reduce the objectivity of the study by varying the perceived difficulty of the task. It could also be argued that people requiring assistive devices are more accurately aware of their abilities, or lack thereof, owing to the process of finding and calibrating assistive technology.

These concerns are out of the scope of this study and there is therefore no benefit in comparing the scores of arbitrarily designated disabled and non-disabled people at this stage. In the same way that a lower age limit was enforced, a minimum ability criterion was also used, restricting participants to those with the physical and mental capacity to complete the data collection tasks. Participants were recruited based on their regular (even if not often) use of computers, providing both a simple inclusion/exclusion criterion and a base-level of ability. This avoided the need to compare ability using two different scales, by only measuring ability of completed tasks.

The criteria mentioned in the previous paragraph provided a minimum level of technology inclusion due its impact on the ability of participants to complete the task. No restriction was however placed on computer anxiety in order to produce the variance required to allow it to be included as a variable in the analysis.

6.2.1.2 Devices

Input devices sense the physical properties of a user (motions, touch, voice etc.) and convert them into predefined signals which can be interpreted by a computer (Taveira & Choi, 2009). Input devices vary greatly and people will be more familiar with their own equipment. Allowing the use of personal equipment for the study creates too much diversity for an objective measure to be created and places participants without their own equipment at a disadvantage. For this reason the same equipment was used for all participants, with factory settings applied.

An Asus One netbook and a first generation iPad (with a similar sized screen) were used. The equipment was chosen as it would likely be unfamiliar to most participants, given the tendency for older people to prefer larger equipment. This preference is due the ease with which larger equipment can been seen and manipulate compared to smaller equipment (Taveira & Choi, 2009). The lack of familiarity with the test equipment also meant that users' predictive self-efficacy judgements were more likely to be based on general technology inclusion rather than prior use of similar equipment. The reduced size of the equipment was also designed to emphasise differences in participants' abilities by increasing the difficulty of their use.

Four different input devices were chosen in order to represent a range of common computer input methods used by older people:

Keyboard: "The keyboard is the... most common computer input device" (Taveira & Choi, 2009) using a series of keys to input linguistic characters. A standard qwerty layout was chosen due to its common use within the English speaking population under investigation.

Mouse: The mouse "is the most commonly used non-keyboard input device with desktop computers" (Taveira & Choi, 2009). An indirect pointing device which is used to manipulate an on-screen cursor. A mouse is normally moved around through the use of a combination of arm/wrist/hand movements which are captured in two dimensions and translated to move the on-screen cursor. One or more buttons can then be pressed by fingers, allowing selections to be made.

Mousepad: Mousepads are indirect manipulation pointing device similar to the mouse. Cursor movement is produced through tracking the movements of fingers on a touch sensitive area, composed of electrodes which form a two-dimensional grid. Mousepads can be operated in either absolute or relative mode, relative mode was chosen owing to: (1) factory pre-sets being used, (2) increased comparability with the behaviour of a mouse, and (3) the small dimensions of the device can pose challenges for older people (Taveira & Choi, 2009) which could be amplified in absolute mode. While physical buttons are available for selection (similar to the mouse) selections can also be made by tapping on the touch sensitive area. In order to allow participants the freedom to interact in a way that suited them, both selection options were left enabled however all participants chose to use the buttons for selection. Although some modern mousepads are able to interpret multi-touch gestures, the mousepad available on the netbook only supported single touch inputs, mimicking a mouse.

Touchscreen: A direct manipulation device which includes both visual output and kinaesthetic input components. Like modern mousepads, modern touchscreens are designed to accept multi-touch gestures and have enabled different paradigms to the traditional mouse and keyboard interaction to be developed. As gestures require a specific knowledge-base and provide an enhanced interaction paradigm compared to that of the previous input devices, their use was deemed out of scope for the study. Instead only basic single touch/press functions were used.

Given the study is an exercise in comparing perceived ability against actual ability, for evaluation it does not matter if the participant is able to reach their full potential as the objective is simply to self-report. From a predictive viewpoint however the participant should be given every opportunity to prove that they are able to perform as they have predicted. Placing the hardware on a firm table will allow the inputs to be manipulated without having to steady them and a well positioned seat will allow the participant to sit at the table in a comfortable position.

6.2.1.3 Environment

The potential variability within the environment has the potential to affect participants' performance, influencing their ability (as described in the discussion on context in chapter 2). As a lab-based study, it was possible to control the majority of the environmental aspects. A practical approach was taken to ensure that results were not affected by altered environment, while ensuring that the environment can be recreated at each of the data collection locations.

In order to provide participants with the opportunity to reach their full potential, different environmental conditions were controlled to facilitate maximum performance:

Kinaesthetics: All of the input devices receive via kinaesthetic channels. Study equipment was placed on a firm table to allow input devices to be manipulated without the need to steady them. A chair was provided with adequate back support and height to allow participants to comfortably access the equipment on the table. It was also ensured that the room was at an appropriate temperature.

Vision: Lighting was identified as an important environmental variable that could affect the ability of participants to see the screen. The room was well lit with natural light which was supplemented by artificial room lighting and equipment was positioned to avoid screen glare. Visual distractions were also kept to a minimum.

Audio: While the study did not have an audio element, background noise is a potential source of distraction. Locations were chosen for their ability to provide a quiet environment and audio distractions were kept to a minimum.

Cognition and Self-Reporting: As described previously distractions were kept to a minimum to reduce the cognitive load[2] on participants and allow them to concentrate fully on each task. This was provided through the use of a private room, which emphasised the fact that participants were being tested. As the effect on testing could have resulted in the Hawthorne effect (Adair, 1984) and in order to reduce this, participants were reminded of anonymity of their data and the need for them to provide honest responses.

6.2.2 Instruments

In order to facilitate data collection, a number of different instruments were used. Questionnaires were used to gather data regarding: demographics, technology

[2]Specifically divided attention.

inclusion, computer anxiety and subjective self-efficacy judgements. Computer-based tasks were then used to provide objective measures of the participants' ability to use each input device.

6.2.2.1 Questionnaires

Sus-IT Digital Inclusion Questionnaire As the study was conducted in collaboration with the Sus-IT project it was able to take advantage of the project's digital engagement questionnaire. The questionnaire was produced in response to the lack of a definitive measure of technology inclusion in order to inform research on digital engagement and subsequent disengagement in older people. The questionnaire was developed through a series of versions with 'critical friends' being used to ensure its appropriateness of its delivery (Keith, 2010). The most up-to-date available questionnaire (v4) was used, with completed questionnaires being forwarded on to provide data to the wider Sus-IT research project (with appropriate ethical clearance obtained). Throughout the project, different versions of the questionnaire were distributed to a total of around 750 people (aged 50+) (Project, 2013) with the 323 responses to the final version (v6) published in Damodaran *et al.* (2013).

The questionnaire recorded quantitative data in the form of: demographic information, sources of assistance that they had used, the amount/types of technology that a participant had used and their comfort in using them. A series of qualitative questions also provided information on the technology-based problems that a participant had experienced and their reasons for any disengagement they had experienced. The questionnaire provided a measure of technology inclusion through recording the number of different pieces of technology that a participant had used. Included technology was selected to represent the variety of potential devices and uses that older people may encounter in order to understand the extent and effect of technology inclusion/disengagement.

A selection of the data was extracted from the completed questionnaires including:

Age: Measured in years at the time of participating.

Gender: Male/Female.

A Technology Inclusion Score: Measured in terms of the number of pieces of technology that the participant had used (between 0 and 33).

Computer Anxiety Rating Scale Many older people claim to experience some form of computer anxiety which is acknowledged as a main factor in the low computer literacy of the age group as described in section 2.2.2. The Computer Anxiety Rating Scale (CARS) (Heinssen *et al.*, 1987) provides a pre-validated measure of computer anxiety which has been widely used, updated and is still in recent use (Broome & Havelka, 2011). Appendix B provides a copy of the questionnaire that was filled out by participants.

The CARS is composed of 19 questions, 10 of which are anxiety laden and 9 non-anxiety laden providing a means of identifying and reducing acquiescence bias. Participants score each statement on a scale from 1–5 where 1 is 'strongly disagree' and 5 is 'strongly agree'. Scores for the non-anxious statements are then reversed before all the answers are totalled. Resulting scores are positively correlated with a participant's anxiety level with the lowest possible score being 19 and the highest being 95.

Subjective Self-Efficacy Judgements As described in section 6.1.2, self-efficacy describes a person's belief in their own abilities. While a number of general computer self-efficacy questionnaires are available, their length and more importantly lack of task specificity made them unsuitable for use in this study. As a result, a custom measure was developed from the guidance in Bandura (2006). The custom measure was designed to allow task specific self-efficacy judgements to be collected, and compared, both in terms of predictive/evaluative judgements and judgements between tasks.

Existing self-efficacy measures use a series of statements, phrased to elicit perceived capabilities from a participant. Bandura (2006) provides a guide for constructing self-efficacy scales and contains a number of examples. Another notable example is Compeau & Higgins (1995), one of the most used (and referenced) computer self-efficacy scales.

A differentiation is identified between predictive and general self-efficacy, with their statements using different modal verbs 'will do', 'can do'. Repeated questioning may get monotonous resulting in the participant paying less attention and responses becoming less accurate (Agarwal & Meyer, 2009; Field, 2009). Therefore the decision was made to keep the length of the questionnaire to a minimum. Three statements were used, corresponding to the dimensions described in section 6.1.3:

Magnitude How well do you think you can [use the input device]? — Expresses perceived ability to use the input device, in a general sense.

Strength How well do you think you [will be able to/were able to] play a game that was based around [the input device]? — Expresses perceived ability with regards to the actual task, providing a measure of the strength of conviction.

Generalisability How much would you AGREE with the statement: "The [input device] is my favourite way to input to a computer"? — Expresses the participant's perception of the input device and perceived ability in comparison with other input devices.

Responses were collected using a Likert scale which moved from one to ten, with one being lowest possible rating (Maurer & Andrews, 2000), allowing participants to simply circle their choice. Ordinarily a scale of one to five would be sufficient however as the sample population is generally inexperienced, responses will be biased towards the bottom of the scale. Increasing the range of the scale will give better definition to the data received. In addition, as they all contribute to the single unified construct of computer self-efficacy the responses were summed, reducing the likelihood of erroneous answers through the use of multiple (related) questions.

6.2.2.2 Objective Assessment Tasks

In order to assess the accuracy of subjective self-efficacy judgements, an objective ability measure was required. Object measures are impartial and measured without bias or prejudice against a scale in a way that is not open to interpretation or subject to personal opinion. In terms of measuring participants' ability to use input devices, a number of methods of measurement are available, two of the most common being the time taken to perform a task, and the number of errors made while performing it.

Common measurements of keyboard performance include words per minute (wpm) and errors per x characters (Arif & Stuerzlinger, 2009). In terms of pointing performance, throughput is a measure of ability comprised of the speed and accuracy with which a target can be acquired, and can be used to describe both direct and indirect manipulation devices. Throughput can be derived from the formula described in Fitts' Law and provides a more thorough measure of ability than either one alone, as demonstrated by the independence of the speed-accuracy tradeoff (MacKenzie & Isokoski, 2008). Although informative, throughput is a reasonably crude method of measuring task ability, a number of more sophisticated measures of pointing performance for example are described in MacKenzie *et al.* (2001). However, while these may be useful in future research to increase the granularity at which measurements can be made, the use of a measure that resembles throughput is sufficient for the uses of this study.

As described in chapter 5, the ability to use an input device is dependent on both the abilities of the user and those of the device/technology itself. Fitts' Law Fitts (1954) allows the identification of the time taken to acquire a target based on its size and distance. The Model Human Processor (Card *et al.*, 1986) uses extra information to allow the modelling of a selection task, by representing visual search and cognitive processing functions as well as the motor element of the task. While both Fitts' Law and the Model Human Processor provide benchmarks against which performance can be judged, they are naive in their presumption that the user is an expert. As both require calibration, neither are appropriate for the prediction of pointing performance in the context of this study.

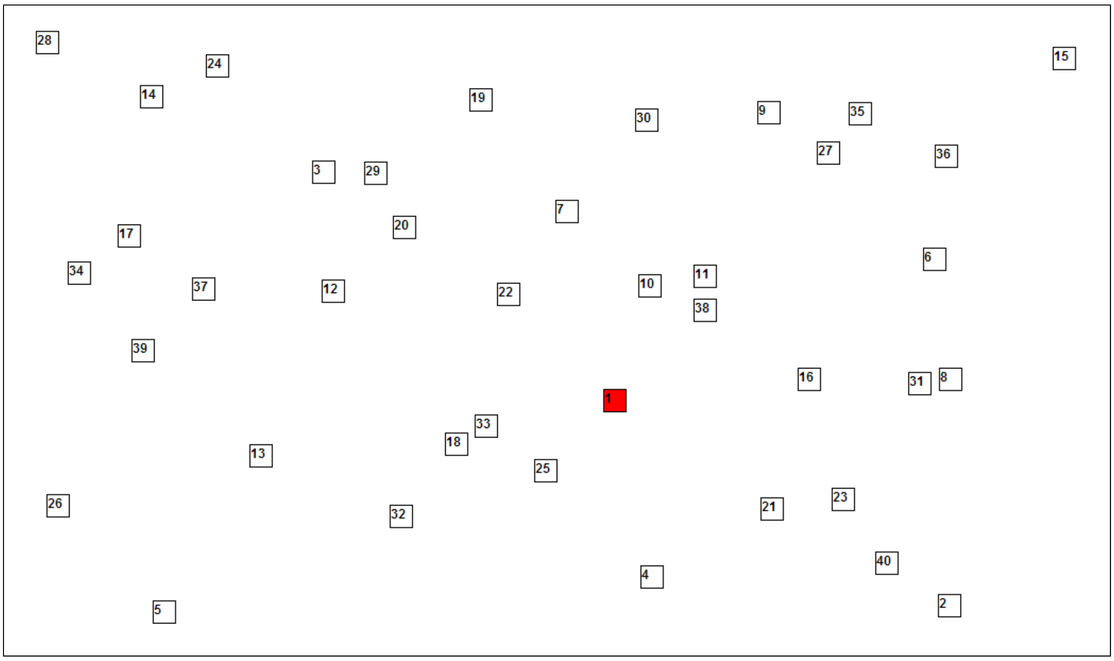
Standardised assessments are available to measure different aspects of the use of different input devices (ISO 9241-400, 2007). However, as with Fitts' Law (and the Model Human processor) the assessments provide a sterilised measure of performance, removed from the reality of a real task.

6.2.2.3 The Mini-Games

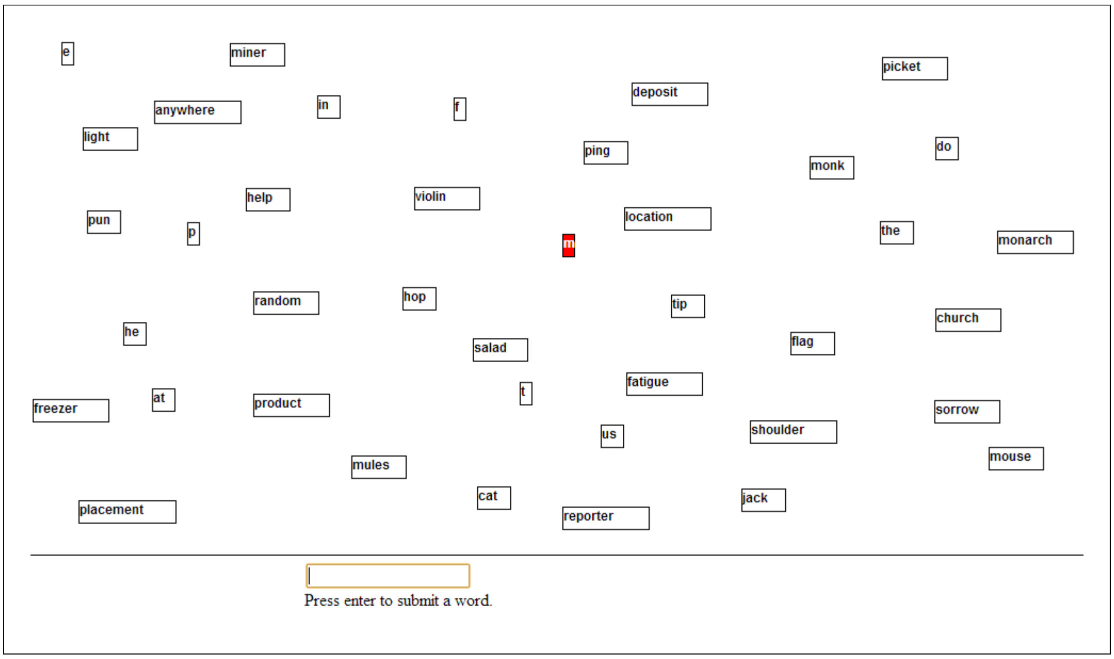
As a result a series of mini-games were produced, which provided a more rounded assessment of ability, through the use of a more realistic task. A series of multi-target selection tasks were used, which required the participant to both identify and 'select' a series of targets using each input device. A screen shot from the tasks is provided in figure 6.2 The task was customised to each input device, with variations being used to represent the use of different types of assessment, replicating the use of different sources of data within the framework.

Given the similarity of the devices, the mouse and mousepad tasks were designed to be identical; both tasks required the user to identify a target, then use the device to move the cursor to the appropriate target and use a single click for selection once it had been acquired. The touchscreen task was similar, with participants being required to identify and touch each target in turn. In order to accommodate the reduced accuracy that is inherent in the use of a touchscreen, targets were increased in size. Finally, the keyboard task was designed to reflect the typical usage of the device; requiring the participant to identify the appropriate target as in the previous task, with selection then being performed by the participant typing the characters that appeared in the target.

Each of the mini-games required the participant to locate and select the targets in the order that they were highlighted, by using the input device under investigation. The user was required to: move the cursor and click the mouse/mousepad, tap the target with a finger (touchscreen) or type the word contained within the target (keyboard). The mini-game collected the number of errors made (the number of clicks that were outside of the currently highlighted target) and the total time taken to complete the mini-game. Further information about the data collected from the games and a brief test procedure can be found in appendix B.2.



(a) The Mouse, Mousepad and Touchscreen games.



(b) The Keyboard game.

Figure 6.2: Example interfaces of the objective assessment games.

6.2.3 Test Procedure

The study's testing procedure was designed in accordance with the guidelines described in Damodaran *et al.* (2012), in an attempt to ensure both the comfort of the participants and the validity of the data collected. As described in section 6.2.1.1, gate-keepers were used to recruit participants. On arrival at the testing venue, participants were welcomed by the gate-keeper, who escorted them to the room that the study would take place in and introduced the investigator. This provided the investigator with a level of familiarity that was designed to speed up the trust building process and reduce the potential anxiety of the participant. To this end, although they were not present during the study, the gate-keeper remained at the venue.

The study was split into two sections, an initial introduction and background data gathering exercise, followed by the collection of subjective and objective ability data. Informed consent was gained by informing participants of the nature of the study and their freedom to leave at any time. The informed consent procedure was also used to build rapport and highlight the importance of participants' honesty when answering questions to reduce self-reporting biases. The digital inclusion and computer anxiety questionnaires were then presented verbally, giving participants a chance to get used to the data collection process and think about their use of technology.

The ability data gathering section consisted of the participant using each of the tasks; giving a prediction of their ability based on an initial description of the task, attempting the task and then providing an evaluative judgement. The order in which devices were used was randomly pre-determined to reduce the effect of a learning bias. A description of the test procedure is documented in Appendix B.

6.2.4 Methodology Summary

This section has presented a methodology for a study which investigates the ability of older people to provide data about their own abilities with regards to different input devices. For each device, an objective measure of performance is gathered in order to act as a control against which the accuracy of other measures of ability can be compared. Participants' subjective self-efficacy judgements and objective measures from alternative devices are also gathered and the suitability of each to be assessed.

In the next section the data collected using this methodology is analysed allowing the results presented in three stages:

- Suitability of the collected data for use in the analysis (in terms of usability and representativeness).
- Initial comparison of subjective and objective measures against the control.
- Ability of other factors to improve prediction of objective or subjective measures.

6.3 Results

This section will report the results of the study and compare them against the null hypotheses (H[1–6]$_0$) used to test each of the hypotheses developed in section 6.1. Thirty-nine participants ($\bar{x} = 67.1$, $SD = 10.7$) were recruited in three data collection sessions. Data collection relied on a stable internet connection, however due to intermittent problems with the test equipment six of the records were incomplete. To improve comparability between statistical tests, incomplete records were removed resulting in a dataset with n=33 between the ages of 52 and 88 ($\bar{x} = 66.1$, $SD = 10.6$).

To assess the reliability of the dataset collected, it was compared to the previous research highlighted in chapter 2. In addition an initial analysis using descriptive statistics was performed in order to assess the validity of the data for use in subsequent statistical tests. Subsection 6.3.2 describes the usability of the dataset; the extent to which it supports existing research and the subsequent impact on the ability to generalise the finding. Correlations were then examined to identify existing relationships within the data in order to validate it against existing research as a representative sample. Finally, regression analyses are performed to investigate the utility of different predictors of performance.

6.3.1 Differences Resulting From Reduction in Dataset Size

The loss of data reduced the dataset from 39 down to 33 records. The resulting change in the dataset can be described through comparison of the maximum and minimum values along with the mean and standard deviation of the dataset, with and without the removed records. An independent t-test was used to compare between the included and excluded records. The exclusion resulted in insignificant changes to Age, Computer Anxiety and Technology Inclusion as demonstrated by table 6.1.

The gender balance (Male:Female) changed from 15:24 to 14:19. Although this resulted in more females being excluded than males, this is acceptable as females still outnumber males.

Details of the removed records and resulting change in descriptive statistics are displayed in tables B.1 and B.2 in appendix B. Further analysis is performed using the restricted dataset.

6.3.2 Is the Dataset Usable?

The dataset comprises of 33 records and the data can be split into three types:

Table 6.1: Independent t-tests to Demonstrate Effect of Removing Records

	Group	N	Mean	SD	t	df	Sig (2 tailed)
Age	Included	33	66.15	10.64	-1.315	37	.197
	Excluded	6	72.33	10.25			
Computer Anxiety	Included	33	19.52	6.41	1.647	37	.803
	Excluded	6	14.83	6.34			
Technology Inclusion	Included	33	44.06	10.76	-1.496	37	.914
	Excluded	6	51.33	12.09			

Factors: Age, Gender, Computer Anxiety and Technology Inclusion.

Subjective Measures: Predictive and Evaluative Judgements for Touchscreen, Mouse, Mousepad and Touchscreen.

Objective Measures: Touchscreen, Mouse, Mousepad and Keyboard.

Initial analysis will examine the raw data in terms of descriptive statistics to identify its potential use for further analysis.

As the dataset is comprised of data from three collection sessions the validity of each session can be investigated. The different sessions were however designed to ensure a variety of participants were recruited, providing a representative holistic dataset. Variation between the sessions is therefore acceptable and explanations are provided where this is the case.

6.3.2.1 Factors

Age Participants ranged from 52 to 88 years old ($\bar{x} = 66.1$, $s.d = 10.6$). As it is measured on a continuous scale, age is appropriate for use in a range of statistical tests. The Shapiro-Wilks Test provides a quantitative evaluation of the normality of a distribution, with non-significant results ($p > .05$) indicating normality (Field, 2009, p 248). As age $D(33) = .918$, $p = .017$ is significant at the $p > .05$ level (see table 6.2), this indicates that distribution is non-normal and checks will be made for normality of residuals when using tests for correlations.

Although the distribution of ages is non-normal this result can be explained through the biases of the data collection sessions. As a younger population, Mickleover contributed towards the spike at the lower end of the age range. An artificial floor was created by restricting participants based on age, in effect only the 'upper' half of the expected normal distribution is present in the sample.

Collection in both Rotherham and Dundee was facilitated through older peoples' social groups. They therefore provided older participants, as demonstrated by the increase mean ages displayed in table 6.3. Although neither of the distributions are normal, they both demonstrate the characteristics expected of a normal distribution (an increased frequency in the middle of the distribution and two tails). The age ranges present in the two groups resulted in a second spike towards the higher end of the overall distribution.

Table 6.2: Shapiro-Wilk Test of Normality for Factors

	Statistic	df	Sig.
Age	.918	33	.017
TI	.936	33	.053
CA	.983	33	.876

Table 6.3: Age Profiles of Data Collection Sessions

Group	N	Mean	S.D
Mickleover	17	59.88	8.42
Rotherham	8	68.50	8.09
Dundee	8	77.13	7.14

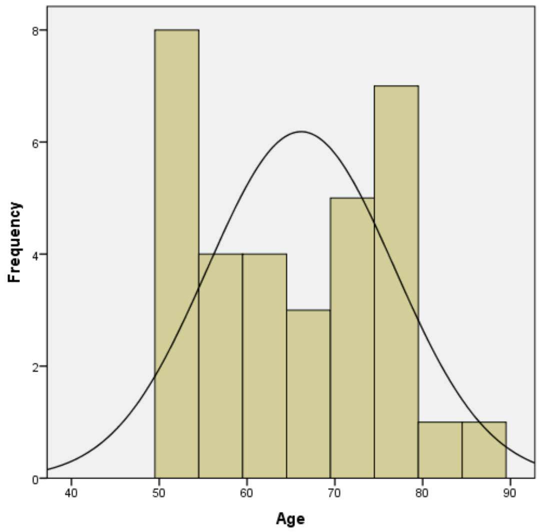
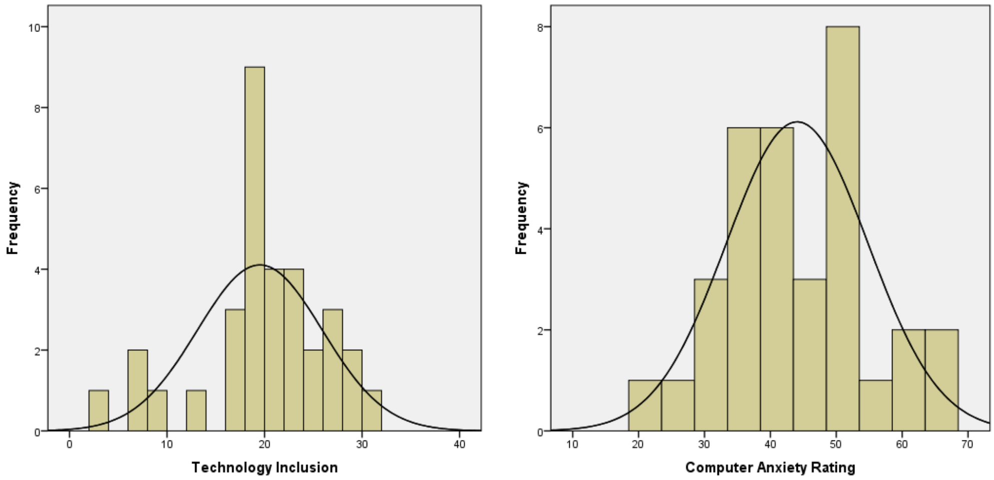


Figure 6.3: Histogram showing the distribution of participants' ages.

Technology Inclusion and Computer Anxiety The uneven distribution of Technology Inclusion Scores and the spike that is apparent in figure 6.4a can be explained in terms of the questionnaire design. The version of the questionnaire that was available for use described a limited set of technologies using a series of generic groupings. A number of newer consumer technologies were not present until later versions and as many people are exposed to a base level of technology this created a partial floor effect. The three of the four participants that scored below 10 were from the Rotherham group and may be representative of the financial barrier created by historical poverty in the area. The Shapiro-Wilks Test for normality is marginally above the $p < .05$ significance level (p=.053) indicating normality. The low adherence to a normal distribution however results in the need for checks to be made for normality of residual values when using tests for correlations (Field, 2009, p 248).



(a) Technology Inclusion. (b) Computer Anxiety.

Figure 6.4: Histograms showing the distribution of results for two factors.

The distribution of Computer Anxiety seen in figure 6.4b has a spike in the top half of the distribution. No explanation can be provided for this, however the troughs that are present either side of the spike suggest that some form of bunching occurred and the data is consistent with a sample taken from a population with a normal distribution. The Shapiro-Wilks Test of Normality is insignificant with p=.876 (see table 6.2 indicating normality of the distribution.

6.3.2.2 Subjective

The eight subjective measures can be described in terms of their grouping by device or measurement type (predictive or evaluative). The histograms depicting their distributions and a table containing their Shapiro-Wilks scores (table B.3 can be found in section B.4). Visual inspection of the self-efficacy measures for the Touchscreen and Mousepad show greater adherence than the Mouse and Keyboard. This is backed up by the Shapiro-Wilks scores. Both the Predictive and Evaluative judgements for the Touchpad and Mousepad being non-signficant ($p > .05$) whereas the Mouse Predictive and Keyboard Evaluative judgements were both significant at the $p < .05$ level.

Also of interest is the skew of the distributions. While the Touchscreen and Mousepad are evenly distributed, the Mouse and Keyboard distributions are skewed to the right (see table B.3). This skew is present in both predictive and evaluative judgements and can be explained by the difference in popularity of the devices (described in section 6.2.1.2). As participants already have experience of using the mouse and keyboard, their preconceived biases result in a right hand skew, indicating over-confidence. As participants have had limited exposure to the Touchscreen and Mousepad, their self-efficacy judgements are not biased by personal experience, resulting in a more even distribution.

As they are all measured on the same scale, the distributions are directly comparable and indicate the relative efficacy of the participants in relation to each device. The distributions have similar standard deviations, allowing the means to be directly compared. The paired t-test tests the similarity of two distributions, with significant results indicating dissimilarity (Field, 2009, p 325). Table 6.4 present the results of t-tests run between the predictive and evaluative scores for each of the devices. The predictive scores for both the Touchscreen and Mousepad were lower than for the Keyboard and Mouse. The evaluative scores however show the Mousepad being ranked below the Keyboard, Touchscreen and Mouse. This meant that on average participants provided significantly higher evaluative scores for the Touchscreen, $t(32) = -6.354$, $p < .001$, $r = .75$, and Keyboard, $t(32) = -5.090$, $p < .001$, $r = .67$. Both display large effect sizes ($r > .50$), increasing the validity of the reported result. No significant difference was observable between the predictive and evaluative scores for the Mouse or Touchscreen. The change in Mouse and Mousepad Self-Efficacy Judgements was not statistically significant.

Table 6.4: Comparison Between Predictive and Evaluative Judgements

	Mean	S.D.	S.E.	T	df	Sig. (2-tailed)
Touchscreen Predictive	15.45	6.063	1.005	-6.354	32	.000
Touchscreen Evaluative	21.67	5.035	0.877			
Mouse Predictive	21.64	5.349	0.931	-1.331	32	.193
Mouse Evaluative	22.45	4.549	0.792			
Mousepad Predictive	14.27	5.970	1.039	-1.048	32	.303
Mousepad Evaluative	15.12	5.770	1.004			
Keyboard Predictive	17.09	5.986	1.042	-5.09	32	.000
Keyboard Evaluative	20.36	5.482	0.954			

6.3.2.3 Objective

Unlike the subjective measures, the objective measures did not display normal distributions (see table B.4). The left hand skew can be explained due to the open-ended nature of the scale used for measurement. As the distributions are measured on identical scales they are directly comparable. The differences in test design between the three device types reduces the usefulness of this comparison and the lack of normality makes the use of a t-test impractical. Comparison of the means does however indicate that the Touchscreen was the easiest device to use ($\bar{x}$ = 58.48, SD=34.08) followed by the Mouse ($\bar{x}$ = 108.66, SD=95.04), with the Keyboard ($\bar{x}$ = 250.27, SD=115.29) and Mousepad ($\bar{x}$ = 260.08, SD=223.06) harder still. Given the existing research described in section 6.2 this is an expected result.

In order to improve the normality of the data a Logarithmic function was applied to all of the objective measures. This reduced the skew of all of the measures and resulted in the Mousepad and Keyboard producing non-significant results in the Shapiro-Wilks test, indicating increased normality. The lack of highly normal samples results in the need for checks to be made for normality of residuals when using tests for correlations.

6.3.3 Is the Data Representative?

In order to assess the reliability of the dataset, it will be compared against previous research, as described in 6.1. Previous research has focused on the relationships between the different factors. If the dataset matches the characteristics described by previous research, it will improve the validity of comparisons between subjective and objective measures.

The Pearson product-moment correlation coefficient is a measure of the strength of the linear association between two variables. Pearson's correlations assume that data is measured on either an interval scale or as a categorical variable with two categories. In order to assess the significance of a correlation, the sampling distribution should also be normal.

The Spearman rank correlations is a non-parametric test which describes whether a monotonic relationship exists between two variables. By performing a correlation on a ranked version of the variables, the test is able to detect non-linear relationships. As there is debate about the legitimacy of describing ranking data as interval (rather than ordinal), comparison between Spearman's rho and the Pearson correlation allows a more informative interpretation of the data to be made.

Both statistics provide a measure of the strength of the relationship between two variables and can indicate whether the relationship is positive or negative. Relationships between two variables can be described in terms of an outcome which is predicted by a mathematical model, e.g. outcome = (model) + error. For linear relationships this can be expressed as an equation (6.1), where Y is the outcome to be predicted by X_1, β_1 is the gradient of the relationship and β_0 is the intercept (if plotted on a graph). Correlations do not describe the nature of the relationship in terms of equation 6.1. They do however provide a measure of the strength of the relationship (r = Pearson and r_s = Spearman) and the probability that the equation is significant (p) i.e. not due to random chance.

$$Y = (\beta_0 + \beta_1 X_1) + \varepsilon \tag{6.1}$$

While Pearson describes the strength of the linear relationship between two variables, Spearman is able to cope with any non-linear monotonoic relationship. For linear relationships therefore the two measures will provide similar results. If the Spearman results are significantly better than Pearson results, a non-linear model may provide a better representation of the relationship and a better fit to the data. Although more complex models are available, their increased data requirement cannot be satisfied by the existing dataset. Later in the chapter, multiple linear regression will be used to provide additional support to the conclusions drawn from correlation analysis. Linear regression is based on the use of linear models to describe data and provide additional information on top of basic Pearson correlations. The observation of significance from Pearson results indicates that a linear model can provide an adequate level of predictive power to fulfil the aims of this study and the use of more complex models is therefore out of scope.

6.3.3.1 Technology Inclusion and Age

Technology Inclusion was significantly negatively correlated with Age, $r = -.42$ and $r_s = -.35$, both $p < 0.05$. The negative coefficient is understandable given the reduced technology exposure that currently characterises the the ageing population (Olphert *et al.*, 2005). The fact that the Pearson coefficient is slightly higher than the Spearman coefficient indicates a potential issue with homoscedasticity. A visual inspection using a scatter plot confirmed this and allows a better understanding of the relationship to be gathered.

Figure 6.5 confirms the general negative relationship but highlights that the Dundee group appears to have higher Technology Inclusion than other participants of equivalent ages. With the Dundee group excluded the relationship improves, $r = -.69$, $p < 0.001$ and $r_s = -.62$ $p < 0.01$. This observation is likely to be related to their involvement in the Dundee user centre. Participants may be drawn to the user centre by a general interest in technology or attendance may have resulted in increased exposure.

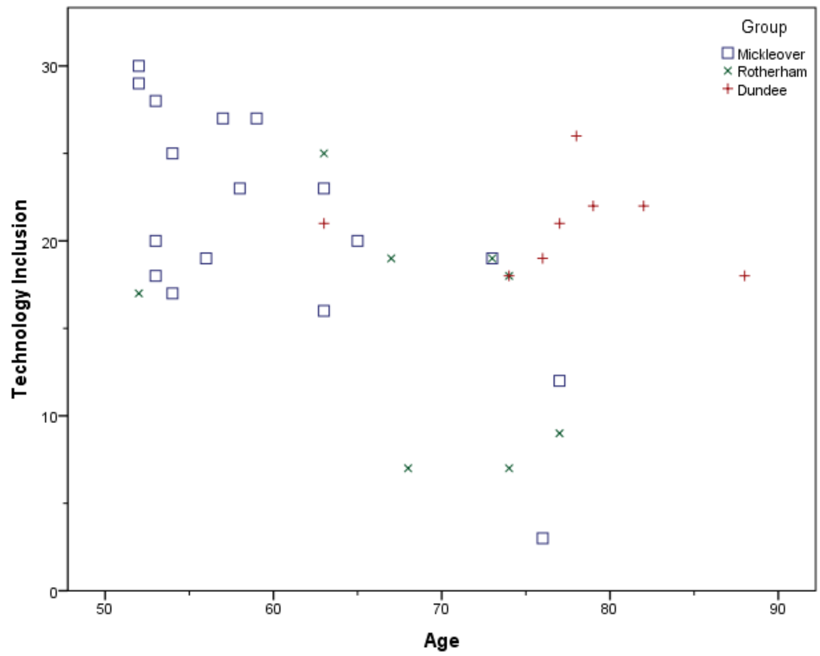


Figure 6.5: Scatter plot comparing participants' Technology Inclusion and Age.

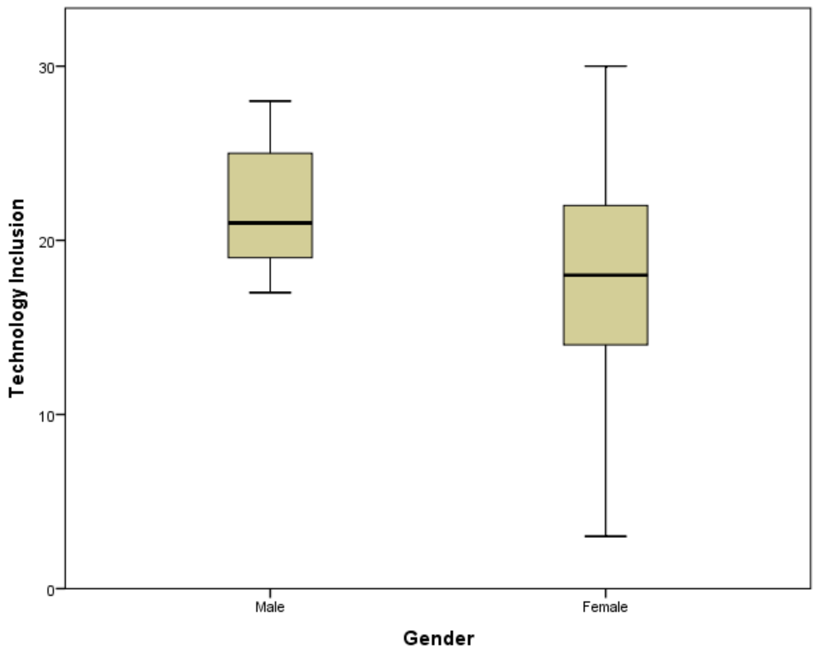
6.3.3.2 Technology Inclusion and Computer Anxiety

Technology Inclusion was also significantly negatively correlated with Computer Anxiety, $r = -.38$ and $r_s = -.43$, both $p < .05$. The comparable negative coefficients suggest agreement between the two measures, indicating a negative relationship as expected given previous literature. Chua *et al.* (1999) found a consistent negative relationship between computer experience and computer anxiety. This trend has been re-iterated in Olphert *et al.* (2005), which highlights the psychological barriers that are restricting older peoples' technology inclusion.

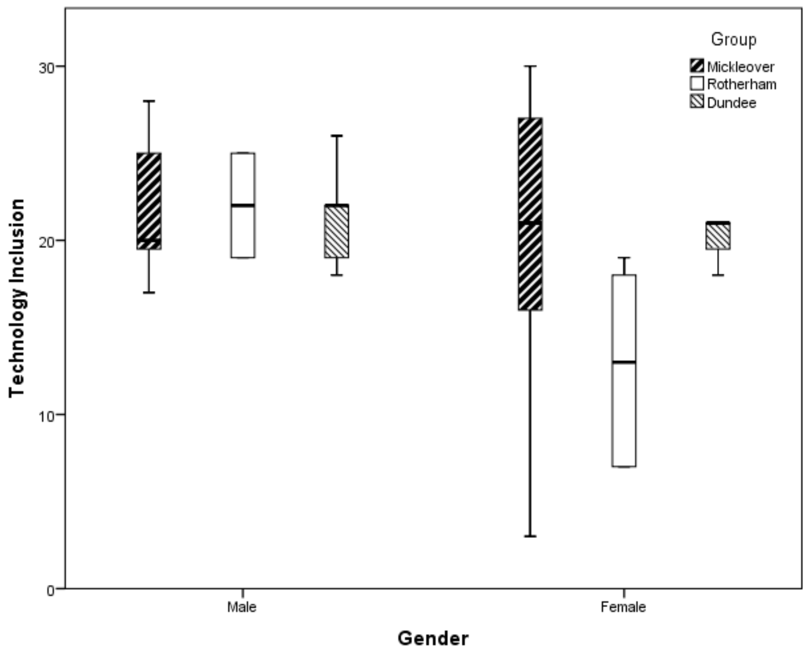
When the Dundee participants were removed both coefficients improved slightly, however the significance of the Pearson coefficient was reduced, $r = -.39$, $p > .05$ and $r_s = -.46$, $p < .05$. This indicates that although the Dundee participants tend to have higher Technology Inclusion than other people with equivalent ages, their Computer Anxiety scores are in line with that expected, given their exposure.

6.3.3.3 Gender and Technology Inclusion

In the literature the influence of Gender on computer use is uncertain, with studies finding both increased use in older males and no difference based on gender (Wagner *et al.*, 2010). A visual inspection of the data (as seen in figure 6.6) generally supports increased male use. No significant difference was however found between the Technology Inclusion of Males ($\bar{x} = 21.79$, SD=3.53) and Females ($\bar{x} = 17.84$, SD=7.56), $t(31) = -1.806$, $p > .05$.



(a) Whole dataset.



(b) Grouped by data collection location.

Figure 6.6: Box plots comparing participants' Gender and Technology Inclusion.

6.3.3.4 Gender and Computer Anxiety

The previous literature is inconclusive about the relationship between Gender and Computer Anxiety. Wagner *et al.* (2010) identified studies that report males to have a more positive attitude to computers, that computer anxiety may be more significant for males and that gender has no relationship with either measure. Conversely Chua *et al.* (1999), a meta-analysis focused on computer anxiety, found female under-graduates to be more anxious than males.

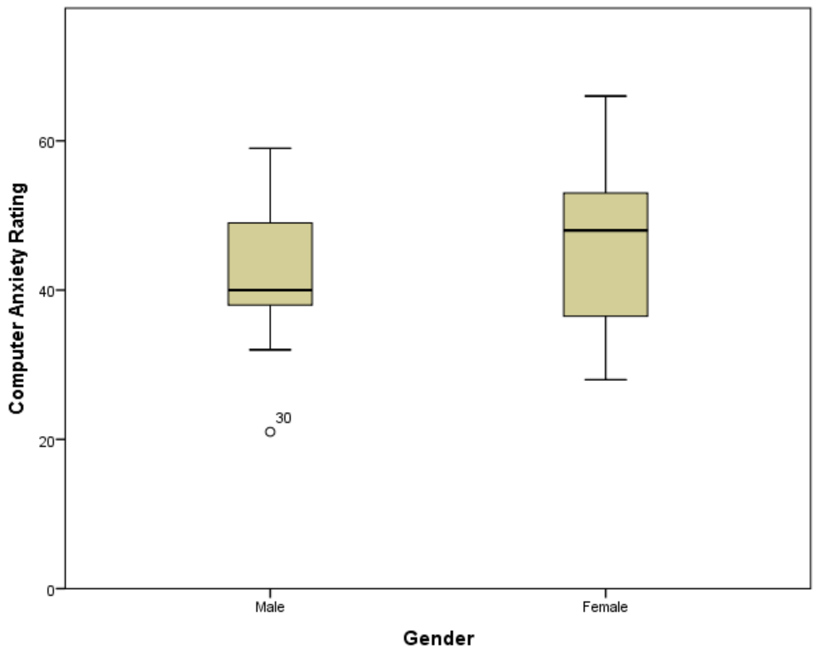
No significant difference was found between the Computer Anxiety of Males ($\bar{x} = 41.43$, SD = 9.76) and Females ($\bar{x} = 46.00$, SD=11.30), $t(31) = -1.215$, $p > .05$. Visual inspection of figure 6.7 is also inconclusive, with the Males and Females in the Mickleover group overlapping, Females being more anxious in the Rotherham Group and Males more anxious in the Dundee group.

6.3.3.5 Gender and Age

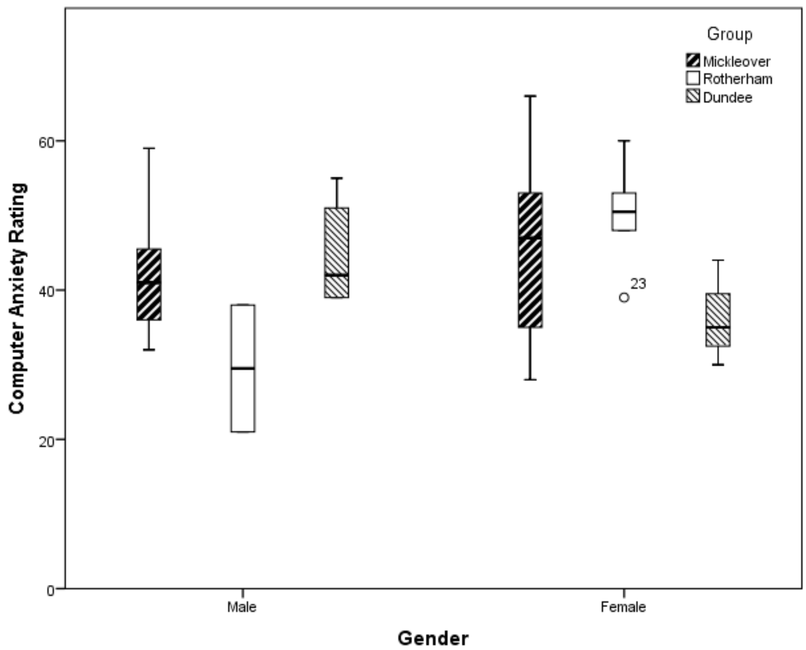
No significant relationship was found between Gender and Age $r = -.15$ and $r_s = -.16$, both $p > .05$. On average women have a longer life span than men(WHO, 2013), however this trend is not observable within the dataset. On initial visual inspection, the males within the study appear to be older than the females (figure 6.8a). Figure 6.8b however shows that this is due to a bias created by the Dundee group being comprised of more male participants.

6.3.3.6 Age and Computer Anxiety

No significant relationship was found between Age and Computer Anxiety $r = -.13$ and $r_s = -.23$, both $p > .05$. This appears to be at odds with the relationships observed between Technology Inclusion and Computer Anxiety and Technology Inclusion and Age. Chua *et al.* (1999) however found that "many studies find no relationship between computer anxiety and age". If the correlation were the result of the previous relationships, any change to one would be mirrored in the others. By removing the Dundee participants improves both the significance and strength of the correlation between Age and Technology Inclusion. As removing the Dundee participants had little effect on the significance and coefficient of the correlation between Age and Computer Anxiety, $r = -.16$ and $r_s = -.26$ (both $p > .05$), this confirms the observation made by Chua *et al.* (1999) and the acceptability of the Dundee participants in the dataset.

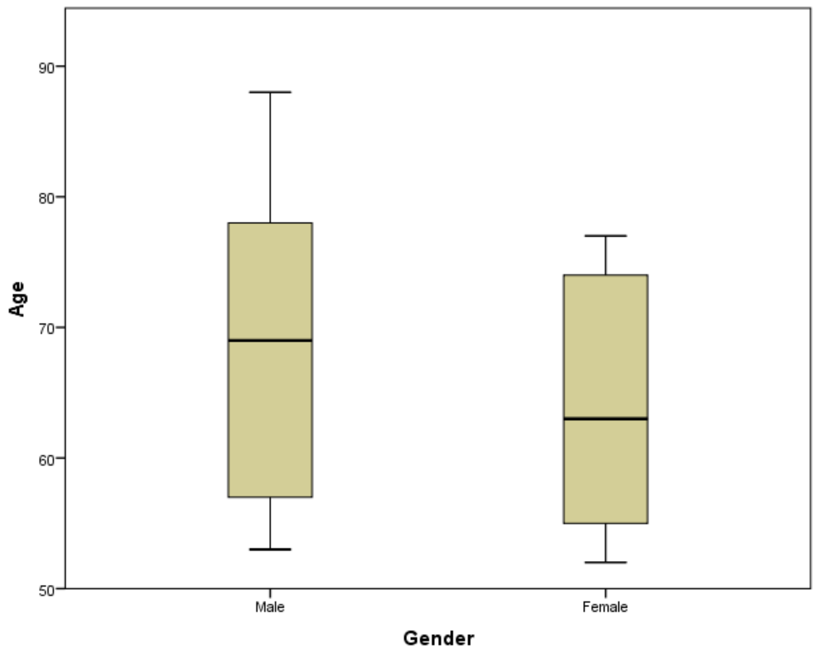


(a) Whole dataset.

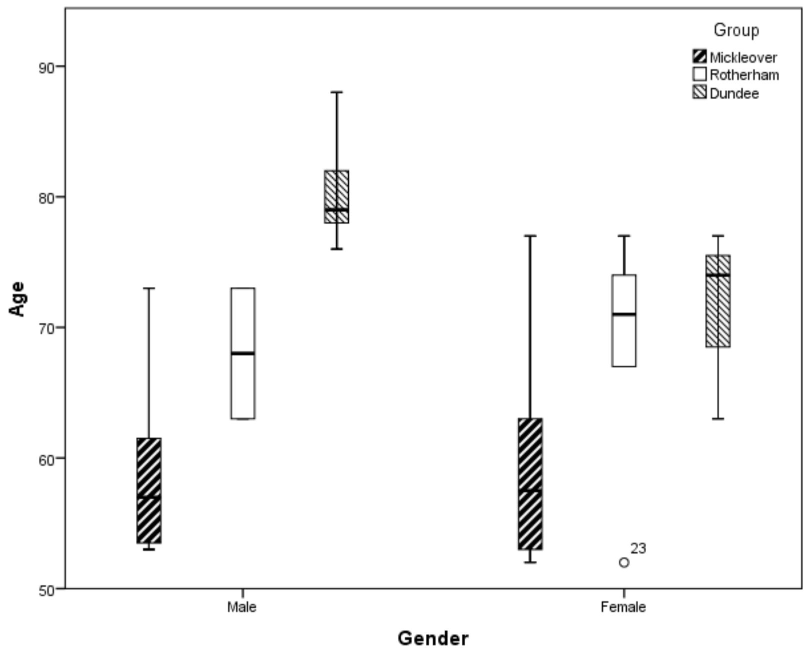


(b) Grouped by data collection location.

Figure 6.7: Box plots comparing gender and computer anxiety.



(a) Whole dataset.



(b) Grouped by data collection location.

Figure 6.8: Box plots describing the Gender and Age of participants.

6.3.3.7 Summary

On first inspection the dataset behaved as expected. The positive relationship between Ageing and Computer Anxiety and the negative relationship between Age and Technology Inclusion identified in existing literature were both visible in the dataset. Other relationships provided less significant results, which again mirrored those observable in the literature. Initial analysis of the dataset suggests that it is representative of the variety found in real computer users, especially given the general trend for increased technology use of older people in general, as identified in chapter 2.

6.3.4 Can People Predict or Evaluate Ability?

The ability of participants to predict or evaluate their ability to use different devices can be understood through an evaluation of the relationship between their subjective and objective scores. The strength of the correlation between the subjective and objective (control) scores provides a measure of the accuracy of participants' subjective judgements. The null hypothesis $H1_0$ is defined, suggesting that "a significant relationship will be observable between participants' subjective judgements and the objective control measure", in order to allow $H1_1$ to be tested. Similarly, the strength of the correlations between the control measure and the objective measures of alternative devices will provide a measure of their respective predictive accuracies. $H2_0$ is also defined, suggesting that "no significant relationship will be observable between participants' alternative objective measures of performance and the control measure for each device", allowing $H2_1$ to be tested. Finally, by comparing the accuracy of participants' subjective judgements against those provided by alternative objective measures the suitability of each measure can be identified. $H3_0$ is defined, suggesting that "either no significant difference in predictive power will be observed, or subjective judgements will provide a better prediction of performance than alternative objective measures", allowing $H3_1$ to be tested.

Table 6.5 provides a summary of this comparison. The upper rows show the result of both Spearman and Pearson correlations between the subjective (predictive and evaluative) and objective measures for each device. The lower rows show correlations between the objective measure of each device and the other devices. The results will be discussed in greater detail in the following subsections.

While bivariate correlation provides limited support to $H1_0$ and reject $H2_0$ (by indicating the strength of the relationships observed), another statistical technique can be used to provide further evidence and tackle $H3_0$. Multiple linear regression allows the estimation of the linear function which represents the re-

Table 6.5: Correlations Demonstrating the Usefulness of Subjective and Objective Measures for Predictive Performance

n=33	Objective Measures							
	Touchscreen		Mouse		Mousepad		Keyboard	
	r_s	r	r_s	r	r_s	r	r_s	r
Predictive	-0.222	-0.209	-.370*	-.510**	-0.111	-0.297	-.361*	-.455**
Evaluative	-.445**	-.524**	-0.220	-0.316	-0.086	-0.079	-0.221	-0.268
Touchscreen	1	1	.574**	.446**	.519**	.380*	.683**	.623**
Mouse	.574**	.446**	1	1	.755**	.821**	.731**	.721**
Mousepad	.519**	.380*	.755**	.821**	1	1	.679**	.658**
Keyboard	.683**	.623**	.731**	.721**	.679**	.658**	1	1

* = $p < .05$; ** = $p < 0.01$; otherwise $p > 0.05$.
r_s = Spearman rho; r = Pearson correlation coefficient.

Table 6.6: Mathematical Models Used in Multiple Linear Regression.

$$
\begin{array}{llllll}
\text{(a)} & Y = & \beta_0 & +\beta_1 X_1 & & +\varepsilon \\
\text{(b)} & Y = & \beta_0 & & +\beta_2 X_2 & +\varepsilon \\
\text{(c)} & Y = & \beta_0 & +\beta_1 X_1 & +\beta_2 X_2 & +\varepsilon
\end{array}
$$

lationship between a dependent variable and a number of explanatory variables. Equation 6.2 is a generalised expansion of equation 6.1 which describes a mathematical model where Y is the outcome expected given the explanatory variables $X_1 \ldots X_i$. For each explanatory variable a β is estimated which describes the regression coefficient of variable in explaining the outcome.

$$Y = (\beta_0 + \beta_1 X_1 + \cdots + \beta_i X_i) + \varepsilon \tag{6.2}$$

Through comparison between models consisting of different explanatory variables, the weighting of different βs can be identified. Table 6.6 shows three models, each with a different combination of explanatory variables. By investigating the difference between (a) and (c), X_1 will be isolated and its effect on the full model (β_1) can be estimated. If the difference between (b) and(c) is then investigated X_2 will be isolated. β_2 can then be compared against β_1 and the relative importance of the two variables can be identified.

As with previous statistical tests, the validity of the result is dependent on a number of assumptions being satisfied (Field, 2009), a selection of the relevant assumptions are:

Variables Types: Variables should be measured on an interval scale.

Non-Zero Variance: A lack of variance in the explanatory variables reduces the potential for a relationship to be identified.

No Strong Multicollinearity: If the explanatory variables are too highly related, it may result in inflation of β values.

No Externally Related Variables: The presence of other variables that can be used to predict the outcome will reduce the power of the model being investigated.

Normally Distributed Errors: If the difference between the model and the observed data is not small and random it is likely that there is an unexplained variable (or constant) biasing the model.

6.3.4.1 Touchscreen

No-significant relationship is observable between the Touchscreen Predictive and Objective Judgement. Evaluative Judgements however are significantly negatively related to the Objective Judgement, $r_s = -.45$ and $r = -.52$, both $p < .01$. The low predictive correlation and subsequent improvement in the evaluative coefficient is understandable given the inexperience of the participants with the device.

In comparison, all of the objective measures provide significant correlations. The relative value of the coefficients is indicative of the relationship between the design of the objective tests. The Keyboard Objective Measure provided a better correlation ($r_s = .68$, $p < .01$) than the Mouse ($r_s = .57$, $p < .01$) or the Mousepad ($r_s = .52$, $p < .01$). The physical skills required by the Keyboard can also be described as more similar to the Touchscreen (using the hand to touch or press targets) than those of the Mouse and Mousepad (indirect manipulation devices).

The variability of the subjective judgements in comparison to the objective judgements suggests that they are less useful for predicting performance. As they are dependent on the participant having experience of the task they are to complete, this also detracts from their use in predicting performance with new devices. In contrast, the objective measures provided a more reliable measure of performance, despite the differences identified in section 6.2.

Using multiple linear regression, subjective values can be assessed against objective values as described above. The models in table 6.7 described the Touchscreen Objective Measure as: (a) Predicted by Subjective Judgements, (b) Predicted by Objective Measures and (c) A function of both Subjective and Objective Measures. Table 6.8 shows the result of a linear regression run between models (a) and (c).

The β value for the Evaluate Judgement in step 1 reconfirms its usefulness as predictors of the Touchscreen Objective Measure. As the β of the Evaluative

Table 6.7: Mathematical Models Predicting the Touchscreen Objective Measure

$$
\begin{array}{llllll}
\text{(a)} & Y = & \beta_0 & +\beta_1 X_1 + \beta_2 X_2 & & +\varepsilon \\
\text{(b)} & Y = & \beta_0 & & +\beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 & +\varepsilon \\
\text{(c)} & Y = & \beta_0 & +\beta_1 X_1 + \beta_2 X_2 & +\beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 & +\varepsilon
\end{array}
$$

Y = Touchscreen Objective Judgement; X_1 = Touchscreen Predictive Judgement; X_2 = Touchscreen Evaluative Judgement; X_3 = Mouse Objective Judgement; X_4 = Mousepad Objective Judgement; X_5 = Keyboard Objective Judgement;

Table 6.8: Linear Regression Modelling of Objective Mousepad Performance Against Subjective and Objective Measures

	B	Std. Error	β	VIF
Step 1				
Constant	4.950	.307		
Predictive Judgement	.005	.071	.071	1.334
Evaluative Judgement	-.050	.016	-.560**	1.334
Step 2				
Constant	1.598	.872		
Predictive Judgement	-0.001	0.011	-.007	1.431
Evaluative Judgement	-0.042	0.014	-.475**	1.512
Mouse Objective Measure	-0.279	0.247	-.306	4.675
Mousepad Objective Measure	0.124	0.178	-.165	3.522
Keyboard Objective Measure	0.712	0.210	.638**	2.233

$R^2 = .28$, $p < .01$ for Step 1; $\Delta R^2 = .30$, $p < .01$ for Step 2;
** $p < .01$; otherwise $p > .05$.

Judgement reduces in step 2 this confirms its reduction in usefulness once the Objective Measures have been used. The column containing VIF (Variance Inflation Factor) values is included in response to the assumption that there is no multicollinearity. As all of the values are well below 10 (see footnote[3]), the assumption has been satisfied.

Also of interest are the values of R^2 for each of the models. The first model (subjective judgements) provided $R^2 = .28$, $p < .01$, for the second model (both subjective and objective measures) R^2 improved by 0.30 with the same significance. This implies that when a model is constructed from subjective judgements, it can predict 28% of the variability in the outcome. When Objective Measures are included, predictive power doubles and the change is statistically significant. Another multiple regression was subsequently run with Objective Measures ($R^2 = .39$, $p < .01$) included before Subjective Judgements ($\Delta R^2 = .18$, $p < .01$). This time the initial model predicted a greater amount of the variability in the outcome and the addition of the Subjective Judgements provides a smaller contribution. This indicates that although Subjective Judgements are useful in predicting Objective Performance, Objective Measures from other devices provide a greater deal of predictive power.

Figure 6.9 displays two graphs that can be used to check several of the assumptions listed above. Figure 6.9a demonstrates a left hand skew of the residuals from normality. Assessment of the standardised residuals with the Shapiro-Wilk test however proved non significant ($p = .80$) suggesting that they are sufficiently normal to satisfy the assumption of normality[4].

Figure 6.9b addresses the assumptions of linearity and homoscedasticity and independence from external variables. The clustering of scatterplot around (0,0) suggests independence from externally related variables and the lack of funnelling suggests homoscedasticity. The lack of a well defined curve suggests that the relationship is linear and as such, linear regression is an appropriate statistical test.

As all of the assumptions were met this suggests that both Subjective and Objective Measures provide a degree of predictive power and the best model is produced through a combination of both Objective and Subjective Measures. Despite this result, Objective Measures provide greater predictive power than Subjective Judgements for predicting Touchscreen Ability in an ageing population and may be more useful if the choice has to be made.

[3]Values for the Mouse and Mousepad reached 7.7 and 7.1 when the unadjusted (non-logarithmic) Objective Measures were used.

[4]Another regression was run with the Mousepad Objective Measure removed to ensure no collinearity. Results returned almost identical values for both R^2 and statistical significance.

6.3.4.2 Mouse

A significant negative correlation was observed between the Mouse Predictive and Objective Judgements, $r_s = -.37$, $p < .05$ and $r = -.51$, $p < .01$. A visual inspection of the correlation explained the higher Pearson coefficient as a result of slight homoscedasticity and outliers. No significant relationship was however observed between the Evaluative and Objective Judgements indicating that the participants' self-efficacy was not uniformly affected by their performance at the task.

In contrast, all of the objective measures provided statistically significant coefficients (see table 6.5). The Mousepad, $r_s = .76$, and Keyboard, $r_s = .73$, Objective Measures provided better correlations than the Touchscreen, $r_s = .57$, all $p < .01$. The fact that the objective judgements were consistently significant and that they provided better correlations than the subjective measurements suggests that they provide a more reliable indication of performance.

A multiple linear regression was run between models (a) and (c). The initial analysis using subjective judgements resulted in $R^2 = .27$, $p < .01$; when objective measures were included $\Delta R^2 = .49$, $p < .001$. The size of the change in R^2 during step 2 and the significance of the result suggests that the Objective Measures provide a better prediction of the Objective Mouse Measure. This was confirmed by a second regression run between models (b) and (c). The step 1 analysis using Objective Measures resulted in $R^2 = .71$, $p < .001$, when Subjective Judgements were included $\Delta R^2 = .03$, $p > .05$. The size and significance of the step 1 result combined with the small size and insignificance of the step 2 result indicates that Subjective Judgements do not provide any additional predictive power to a model based on Objective Measures. As a result they are less appropriate for predicting the Objective Mouse Measure than Objective Measures from other devices.

The tables showing the analysis can be found in Appendix B. As all of the required assumptions were met this suggests that Objective Measures from alternative devices should be used over Subjective Judgements provided by users when attempting to predict Objective Mouse performance.

6.3.4.3 Mousepad

Neither the Predictive or the Evaluative Judgements had a significant relationship with the Mousepad Objective Measure. The poor predictive ability is understandable given the relative unfamiliarity of the participants with the device. The poor evaluative ability on the other hand may reflect the inability of participants to provide an objective measure of their abilities, resulting in the subjective nature of their evaluations.

Table 6.9: Mathematical Models Predicting Mouse Objective Measure

(a)	$Y =$	β_0	$+\beta_1 X_1 + \beta_2 X_2$		$+\varepsilon$
(b)	$Y =$	β_0		$+\beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5$	$+\varepsilon$
(c)	$Y =$	β_0	$+\beta_1 X_1 + \beta_2 X_2$	$+\beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5$	$+\varepsilon$

Y = Mouse Objective Judgement.
X_1 = Mouse Predictive Judgement; X_2 = Mouse Evaluative Judgement.
$X_4 = X_3$ = Touchscreen Objective Measure; Mousepad Objective Measure; X_5 = Keyboard Objective Measure;

Table 6.10: Mathematical Models Predicting Mousepad Objective Measure

(a)	$Y =($	Subjective Judgements		$)+\varepsilon$
(b)	$Y =($		Objective Measures	$)+\varepsilon$
(c)	$Y =($	Subjective Judgements +	Objective Measures	$)+\varepsilon$

In contrast, all of the objective measures provided statistically significant coefficients. This suggests that Objective Measures provide a more reliable indication of performance than subjective measures. In addition, the highest coefficient was again demonstrated by the task with the greatest similarity (the Mousepad), with the Keyboard also providing a reasonable coefficient.

A multiple linear regression was run between models (a) and (c) as displayed in table 6.10. The initial analysis using Subjective Judgements resulted in $R^2 = .12$, $p > .05$, when Objective Measures were included $\Delta R^2 = .58$, $p < .001$. The small size and lack of statistical significance of the step 1 results, the size of the change in R^2 during step 2 and the significance of the result suggests that the Objective Measures provide a better prediction of the Objective Mouse Measure. This was confirmed by a second regression run between models (b) and (c). The step 1 analysis using Objective Measures resulted in $R^2 = .68$, $p < .001$, when Subjective Judgements were included $\Delta R^2 = .01$, $p > .05$. The size and significance of the step 1 result combined with the small size and insignificance of the step 2 result indicates that Subjective Judgements are not useful for predicting the Objective Mouse Measure.

The tables showing the analysis can be found in Appendix B. As all of the required assumptions were met this suggests that Objective Measures from alternative devices should be used over Subjective Judgements provided by users when attempting to predict Objective Mousepad performance.

6.3.4.4 Keyboard

A significant negative relationship was observed between the Keyboard Predictive and Objective Judgements, $r_s = -.36$, $p < .05$ and $r = -.46$, $p < .01$. No significant relationship was found between the Keyboard Evaluative and Objective Judgements. Participants' prior Keyboard experience is likely to explain the improved predictive ability. The perceived difficulty of the task may then have reduced their self-efficacy. In contrast, all of the objective measures provided statistically significant coefficients with similar correlation coefficients. As with the other devices, this suggests that Objective Measures provide a more reliable indication of performance than subjective measures.

The similarity between the coefficients of all the Objective Measures indicates that the Keyboard task appears to be a more reliable measure of performance across devices than the other devices. This could be explained in relation to the nature of the task itself. The Keyboard task was the only one which got gradually harder as it progressed and as the only task to require text entry, it also contained a cognitive element that was not present in the other tasks. These differences result in the potential to provide a more comprehensive indication of the participants' ability.

A multiple linear regression was run between models (a) and (c) as displayed in table 6.11. The initial analysis using Subjective Judgements resulted in $R^2 = .23$, $p < .05$, when Objective Measures were included $\Delta R^2 = .46$, $p < .001$. The size of the change in R^2 during step 2 and the significance of the result suggests that the Objective Measures provide a better prediction of the Objective Mouse Measure. This was confirmed by a second regression run between models (b) and (c). The step 1 analysis using Objective Measures resulted in $R^2 = .64$, $p < .001$, when Subjective Judgements were included $\Delta R^2 = .05$, $p < .05$. The size and significance of the step 1 result combined with the small size and insignificance of the step 2 result indicates that Subjective Judgements are not useful for predicting the Objective Mouse Measure.

Details of the analysis can be found in Appendix B. All of the required assumptions were met, however, inflated VIF values provided an indication of minor collinearity[5] between Mouse (3.9) and Mousepad (3.6) Objective Measures. The regression was rerun with the the Mousepad Objective Measure excluded and results were similar. The only difference was a reduced significance when Subjective Judgements were added in step two of the regression between models (b) and (c), which supports the original analysis. This suggests that Objective Measures from alternative devices should be used over Subjective Judgements provided by users when attempting to predict Objective Keyboard performance.

[5] The values are well below the critical value of 10.

Table 6.11: Mathematical Models Predicting Keyboard Objective Measure

(a)	$Y =($	Subjective Judgements		$)+\varepsilon$
(b)	$Y =($		Objective Measures	$)+\varepsilon$
(c)	$Y =($	Subjective Judgements +	Objective Measures	$)+\varepsilon$

6.3.4.5 Summary

This subsection provides evidence which can be used to tackle the first three hypothesis defined in section 6.1. $H1_0$ suggested that a significant relationship would be observable between subjective judgements and the objective control mesure. In order to reject the null hypothesis no significant relationships should be observable however significant relationships were observable for correlations between either predictive or evaluative judgements and the objective control measure for three of the devices. This provides evidence which partially supports the null hypothesis meaning that $H1_0$ cannot be rejected. There was however no predictable pattern to whether predictive or subjective judgements would provide useful; in total, only three of a possible eight subjective judgements provided significant relationships. This could therefore be viewed either as very weak evidence to support $H1_1$ with further research being necessary or a repeat of the inconclusive results seen in the literature, as described in Wagner *et al.* (2010).

$H2_0$ suggested that no significant relationship would be observable between the objective control measure and the alternative objective measures. Comparison between the objective measures of different devices however provided consistently significant relationships allowing the null hypothesis $H2_0$ to be rejected. The correlation co-efficients themselves were all reasonably strong (Pearson results varied for the Touchscreen) adding weight to support $H2_1$.

Finally, $H3_0$ suggested that either subjective judgements would provide greater predictive power than, or there would be no significant difference to, alternative objective measures. Multiple linear regressions for all of the devices suggested that alternative objective measures provided greater predictive power than subjective judgements, allowing the null hypothesis $H3_0$ to be rejected. This provides support for $H3_1$, suggesting that objective measures are better than participants' subjective judgements for measuring ability.

There is also evidence to support the atomic approach taken in the previous chapter. The relative strength of the relationships (measured by the correlation coefficient) varied between tasks and the greater the similarity of the tasks the greater the coefficient observed.

6.3.5 How Do Other Factors Affect Prediction Ability?

Having found evidence to support the use of objective measures over subjective judgements, this section will investigate the factors that affect prediction accuracy. Participants' subjective accuracy varies, both by device and between predictive and evaluative judgements. In section 6.3.3, a number of factors were used to assess the validity of the dataset against existing literature. Multiple regression analysis can be used to investigate the effect that each of the factors has on predicting the accuracy of participants' subjective judgement and alternative objective measures.

If the factors improve the coefficient of the regression model between subjective and objective measures, it can be used to indicate the improvement in the accuracy of the subjective judgements. If the accuracy of judgements can be predicted, appropriate weightings can be applied to counter the bias that the factor is producing. The null hypothesis $H4_0$ is defined, suggesting that "inclusion of participant factors will have no effect on the ability of subjective judgements to predict performance", allowing $H4_1$ to be tested. Similarly, if factors improve the coefficient of the regression model between subjective and objective measures $H5_0$ is also defined, suggesting that "inclusion of factors will have an effect on the ability of alternative objective measures to predict performance", allowing $H5_1$ to be tested.

Depending on the level to which factors are able to improve subjective judgements, their appropriateness compared to alternative objective measures may need to be reassessed. As a result $H6_0$ can be defined, suggesting that "alternative measures of performance will have no effect on predictive power in addition to the combination of older people's subjective measures and other factors", allowing $H6_1$ to be tested.

6.3.5.1 The Effect of Factors on Prediction Using Subjective Judgements

Eight (4*2) multiple linear regressions were run based on the models in table 6.12, a summary of the results can be found in table 6.13. They suggest that models based on the factors provide greater predictive ability than models based solely on participants' subjective judgements (based on a comparison of R^2 and significance values between models (a) and (b)). Further, the addition of the factors into a model based on subjective judgements ((a)→(c)) consistently improves prediction power. In contrast, the addition of subjective judgements to a model based on the factors ((b)→(c)) does not appear to increase predictive power with the coefficient of the change being both small and having no statistical significance.

The improvement in predictive power shown by adding the factors to the sub-

Table 6.12: Mathematical Models Evaluating the Effect of Factors on Subjective Measures

(a)	Device Objective Measure=(	Factors		)+ε
(b)	Device Objective Measure=(		Subjective Judgements	)+ε
(c)	Device Objective Measure=(	Factors +	Subjective Judgements	)+ε

jective judgements allows $H4_0$ to be rejected. In addition, the size of the step 2 change for the (b)→(c) models provides evidence to support $H4_1$.

Table 6.13: Summary of Multiple Linear Regressions Investigating the Effect of Factors on Prediction Using Subjective Judgements

	Models		R^2 (Step 1)	ΔR^2 (Step 2)
Touchscreen	(a)	(c)	.337*	.218**
	(b)	(c)	.278**	.278*
Mouse	(a)	(c)	.606***	0.025
	(b)	(c)	.272**	.359**
Mousepad	(a)	(c)	.548***	0.053
	(b)	(c)	0.117	.484***
Keyboard	(a)	(c)	.555***	0.067
	(b)	(c)	.232*	.391**

* $p < .05$; ** $p < .01$; *** $p < .001$; otherwise $p > .05$.
Models are based on those in table 6.12.

6.3.5.2 The Effect of Factors on Prediction Using Objective Measures

Eight (4*2) multiple linear regressions were run based on the models in table 6.14, a summary of the results can be found in table 6.17. They suggest that models based on objective measures consistently outperform models based on the factors (comparing R^2 for models (a) and (b)). Further, the addition of objective measures to a model based on the factors alone ((a)→(c)) consistently improves prediction power. In contrast, the addition of Factors to a model based solely on objective measures ((b)→(c)) did not improve precision power with any statistical significance.

The lack of improvement in predictive power demonstrated by the addition of factors to alternative objective measures allows the null hypothesis $H5_0$ to be rejected. In addition the comparative significance of the (a)→(c) models provides

Table 6.14: Mathematical Models Evaluating the Effect of Factors on Objective Measures

(a)	Device Objective Measure=(	Factors		)+ε
(b)	Device Objective Measure=(		Alternative Objective Measures	)+ε
(c)	Device Objective Measure=(	Factors +	Alternative Objective Measures	)+ε

evidence to support $H5_1$.

Table 6.15: Summary of Multiple Linear Regressions Investigating the Effect of Factors on Prediction Using Objective Measures

	Models		R^2 (Step 1)	ΔR^2 (Step 2)
Touchscreen	(a)$^+$	(c)$^+$	.34*	.17*
	(b)$^+$	(c)$^+$	.39**	.12
Mouse	(a)	(c)	.61**	.17*
	(b)	(c)	.73**	.05
Mousepad	(a)	(c)	.55**	.18*
	(b)	(c)	.68**	.04
Keyboard	(a)$^+$	(c)$^+$	.56**	.15*
	(b)$^+$	(c)$^+$	.63**	.07

* $p < .05$; ** $p < .01$; otherwise $p > .05$.
Models are based on those in table 6.14.
$^+$ Denotes removal of Mousepad Objective Measure to reduce multicollinearity.

6.3.5.3 The Effect of Factors on the Suitability of Using Subjective Judgements and Objective Measures for Prediction

Eight (4*2) multiple linear regressions were run based on the models in table 6.16 ,a summary of the results can be found in table 6.17. They suggest that models based on objective measures consistently outperform models based on subjective judgements with factors included (comparing R^2 for models (a) and (b)). Further, the addition of objective measures to a model based on the factors alone ((a)→(c)) consistently improves prediction power. In contrast, the addition of factors to a model based solely on objective measures ((b)→(c)) did not improve precision power with any statistical significance.

The improvement in predictive power shown by adding alternative objective measures to the subjective judgements and other factors allows $H6_0$ to be rejected. In addition, the significance of the step 2 change for the (a)→(c) models provides evidence to support $H6_1$.

Table 6.16: Mathematical Models Evaluating the Suitability of Subjective and Objective Measures for Prediction with Factors Included

(a)	Objective Measure=(	Subjective & Factors			)+ε
(b)	Objective Measure=(			Alternative	)+ε
(c)	Objective Measure=(	Subjective & Factors	+	Alternative	)+ε

Table 6.17: Summary of Multiple Linear Regressions Investigating the Suitability of Subjective and Objective Measures for Prediction with Factors Included

	Models		R^2 (Step 1)	ΔR^2 (Step 2)
Touchscreen	(a)+	(c)+	.56**	.10*
	(b)+	(c)+	.39**	.27*
Mouse	(a)	(c)	.63**	.17**
	(b)	(c)	.73**	.07
Mousepad	(a)	(c)	.60**	.15*
	(b)	(c)	.68**	.06
Keyboard	(a)+	(c)+	.62**	.14*
	(b)+	(c)+	.63**	.13

* $p < .05$; ** $p < .01$; otherwise $p > .05$.
Models are based on those in table 6.16.
+ Denotes removal of Mousepad Objective Measure to reduce multicollinearity.

6.4 Limitations

The results described in this chapter are based on the data collected for the study. While the data appears to be both usable (as described in section 6.3.2) and representable (as described in section 6.3.3) the applicability of the results may be constrained due to the limited sample size. Although this may be an issue, the adherence of the results to the hypotheses (as developed from existing literature), and the level of significance of the results add credibility to the conclusions drawn.

6.5 Summary

While the relative suitability of the use of an agent is dependent on a number of different factors, the accuracy of the data it provides is a key component of its use. This study is attempting to investigate the suitability of different agents to provide the data required to drive a model based on the framework. A series of hypotheses were therefore developed which describe the suitability of different agents in terms of the accuracy of the data they provide.

H1: No significant correlation is observable between older people's subjective judgements and the control measures of performance.

H2: Older people's alternative objective measures of performance is positively correlated with their control measure of performance.

H3: Older people's alternative objective measures provide a better prediction of performance than their subjective judgements.

H4: Inclusion of participant factors will improve the ability of older people's subjective judgements to predict performance.

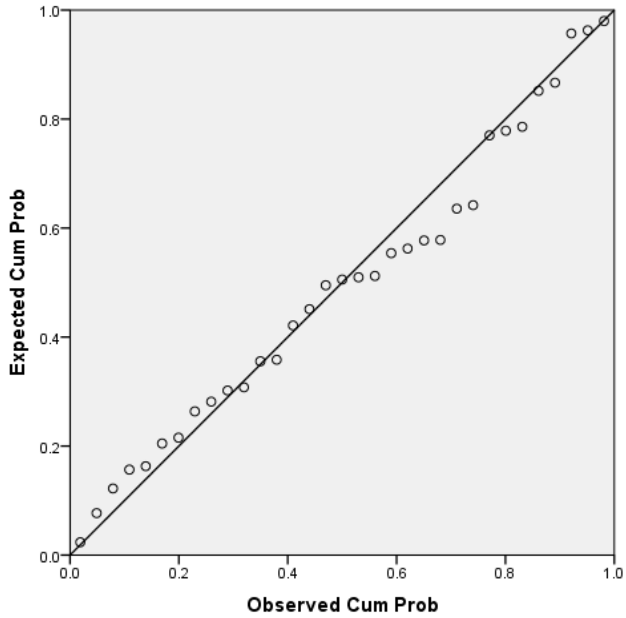
H5: Inclusion of participant factors will not improve the ability of older people's alternative objective measures to predict performance.

H6: Older people's alternative objectives measures will provide a better prediction of the control measure than subjective measures with factors.

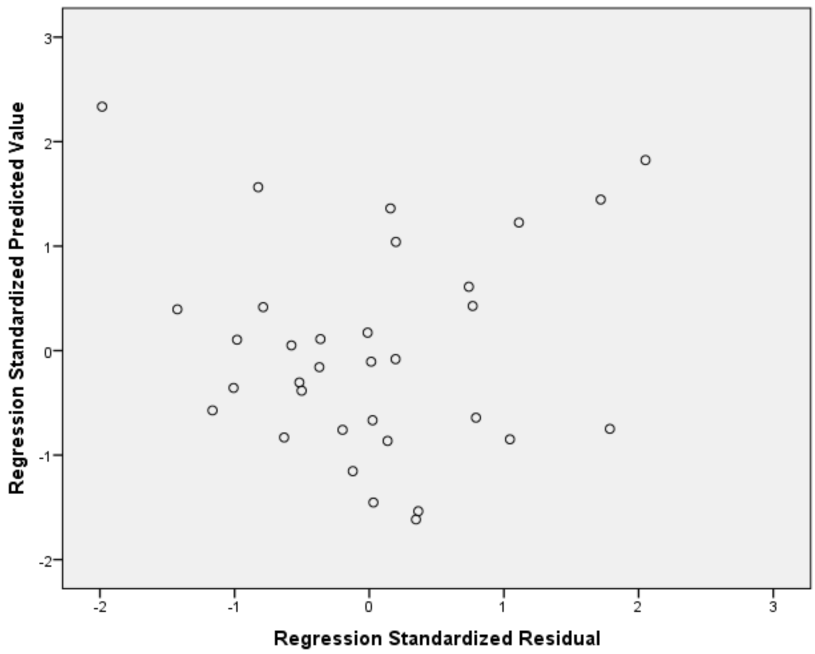
With the exception of $H1_0$ the evidence provided by the results suggested the rejection of all the null hypotheses.

This chapter has therefore validated the inclusion of aim 6 and demonstrated a different facet to its use in the creation of the framework. The implication that automated agents provide a higher quality of data than users themselves results in the need to assess the quality of data (potentially in response to its author) and tailor its use accordingly. In addition, the chapter has also highlighted the need for a reliable method of data collection, owing to the inability of users to reliably provide the data required by a system based on the framework. As costs associated with expert assessment make it an impractical option, automated and semi-automated data collection agents are proposed as the preferred method of data capture. Throughout the course of this study high ethical standards have been adhered to, unfortunately the same cannot be guaranteed from a commercial system.

In the next chapter the ethical considerations surrounding data capture and use are discussed along with other considerations arising from the positioning of the framework within a working system. While chapter 5 provided a means of identifying the similarities between the interaction-based context of different data items, a mechanism is still needed to allow data from different sources to be used with an appreciation of its inherent accuracy. The need to provide users with a level of control over their data is approached and the solution applied to the issues raised in this chapter.



(a) Normal Probability plot.



(b) ZPredicted vs ZResiduals.

Figure 6.9: Scatter-plots displaying Touchscreen Objective measure residuals.

Chapter 7

Ethical Considerations

Having identified the variety in the accuracy of different data collection agents in the previous chapter, there is a need to identify solutions to the issues discussed. The framework is intended to store data about the capabilities of users and the technology, in order to allow systems to be created that are able to predict the adaptations required for the personalisation of interaction. The framework has therefore been designed to be used within a system that is able to provide the required data and provide inferencing to use the data once it has been collected.

Following on from the findings of the previous chapter, this chapter begins in section 7.1 with a discussion of the aims that the will be addressed. Section 7.2 then identifies the stakeholders who may benefit from the development of systems based on the framework and the their individual and combined concerns. In order to move towards an implementable system, some of the 'softer' issues regarding the building of trust and safeguarding user data are then discussed in section 7.3. Following this discussion, the resulting concerns surrounding data storage and use are then raised in section 7.4 and section 7.5 details how they will be addressed within the framework. Finally section 7.6 provides a summary of the contributions of the chapter and its impact on the results of the thesis as a whole.

7.1 Relation to Thesis Aims

In order to move the framework from a theoretical design to an implementable system, the issues raised in chapter 6 surrounding data accuracy must be addressed. The previous chapter provided a better understanding of aim 6 by highlighting the potential differences in data accuracy between collection agents. This chapter now aims to further expand the framework by providing an approach which complements the structure described in chapter 5.

In doing this two of the thesis' aims will be tackled. Firstly, aim 5 deals with the issue of there being a variety of agents both providing and wishing to use the data contained within the framework. While the framework is designed to accept and store all of the data that it is presented with, the theoretical design developed so far does not have a mechanism for restricting access to the data. Secondly, aim 6 deals with the need to appreciate and be able to work with a variety of data of variable quality. While the framework has been designed to cope with variability in the context surrounding a capability at the time it is measured, it does not currently have the facility to restrict access to data based on that information.

As both of the aims are related, a joint approach will be taken to satisfying them. Firstly the ethical approach taken to the collection and use of data will be identified and the ability of the framework to provide the functionality of a permissions system is discussed. Secondly, an approach is developed to store the factors that contribute to the accuracy of a data item and allow filtering based on those factors.

7.1.1 Creating Trust and Issues with Profile Transparency

As a result of working with a large number of older people as part of a participatory research project the importance of building trust has been repeatedly emphasised (Damodaran *et al.*, 2012). In order for the framework to succeed, users have to allow their data to be placed into it and therefore a certain level of trust is required. Trust can be gained through ethical behaviour. In the current climate, people can be hesitant about trusting companies to use their personal data and even more sceptical about allowing them to store it and this has resulted in a desire to understand what ethical conduct actually entails (Li *et al.*, 2010).

The potential need for proprietary and closed systems will be identified, however the solution favoured in this thesis is one of transparency. Transparency is a simple way to build trust by creating open systems and providing accountability. However, in the case of the framework, fully transparent systems may not always be the best idea due to the potentially demoralising effects of highlighting a user's weaknesses (as could be inferred through reference to self-efficacy as described in chapter 6).

The framework is designed to store data about users' abilities, for some users being exposed to this data may not be pleasant. Positivity is an important part of the learning process and motivation is designed to encourage learners to improve. Although learners need to be aware of their performance and have their expectations managed, exposing them to the sometimes harsh reality of their actual abilities (or lack thereof) may be counter productive. Having access to the data collected may change user behaviour, if for example mouse and typing speed are measured, users may understand speed to be an indicator of ability. Although it is, speed is only part of ability indicators, and encouraging users to increase the speed at which they use input devices may be counter productive.

7.2 Stakeholders

Before discussing the ethical considerations of the practical implementation of the framework, the stakeholders that may be interested/affected by its creation will be discussed. Although the theoretical framework itself is the actual output of this thesis, it is the systems that could be driven by the framework that will be most evident to several of the stakeholders. The success of the framework will therefore be based on weighing the advantages it can provide against the issues it raises for each one.

7.2.1 Users

Users are the people that framework has been designed to benefit. As adaptation systems should ideally be embedded within the operating systems of the devices people are using, the framework would become both an integral part of, and indistinguishable from the device as a whole. Many users may be unaware either of the existence of the framework, or its distinction from the devices they are using.

Through the design described in chapter 5, the framework has the potential to assist users by identifying the assistance they require. This knowledge could then be used within services which connect them to appropriate support; either through signposting or automated provision (Atkinson *et al.*, 2012). As the population is both growing and ageing, the production of external support systems could be used to reduce the rate of technology abandonment. The ability to log onto a public terminal and have it personalised based on actual needs would provide users with greater freedom; for example to use self-service checkouts or shopping centre information kiosks.

It is important to note however that automated help systems of any kind should not be used as a total replacement for human support. Many older technology users use technology to facilitate interaction with other people, either electronically or through connections made at computer classes. As a result, they have a preference for face-to-face support and are likely to react poorly to a system that reduces the contact they have with other people (Damodaran *et al.*, 2013). The framework must therefore support and maintain the relationships that older people have rather than disrupting them.

As reliant as users may become on the advantages that the framework can supply, the reliance of the framework on the co-operation of the user must also be appreciated. The framework has been designed as a structure for the storage and comparison of data, and without the permission to collect and store user data, none of the potential benefits can be achieved. Subsequent inferencing, trending and the matching of user abilities and requirements to those of potential adaptations and support is similarly reliant on consent being given. Without gaining acceptance from users, systems will not have the support they need to allow them to reach their full potential. Before they will use adaptive help systems, users may require education about the benefits they can expect to gain. This is very much an issue of marketing, and sits well above the topics discussed in this thesis.

Even if a user can be persuaded to use a system that has been based on the framework's storage abilities, continued use is not guaranteed. The system would rely on the collection of data to allow the provision of support, however if either of these processes are too onerous, the user may choose to ignore them. Firstly in order to access the data required, semi-automated collection agents (similar to those used in the study in chapter 6) may be used. Although data collection tasks can be presented in the form of mini-games (gamification (Eisma *et al.*, 2004)), if they are requested too frequently the data they provide may be biased owing to: (1) learning effects enhancing the users demonstrated abilities and(2) rejection caused by users getting bored or feeling that their privacy is at risk.

As well as considering the way that data is collected, the way that a system deals with the provision of assistance is also important. There is a need to gain consent before changes are made to a user's system. Users may also get frustrated if the assistance is over-active and constantly making suggestions. A fine line must be drawn between support systems being active enough to match the user's needs and making changes so often the user does not have a stable system. This challenge could be faced by storing the user's tolerance for change as a capability.

As users will have different sensitivities it makes sense to include this as a preference during a set-up procedure and the sensitivity could then be changed either periodically or based on feedback/system use. More change may be tolerable, for example, when the user's abilities are improving or they are becoming more familiar with an interface. Alternatively a lower tolerance for change may follow a step-change decline in ability or when the user is observed to be in a generally low mood (Bandura, 1995).

7.2.2 Accessibility System Providers

As mentioned above, the framework provides a method of data storage and comparison allowing cross-system transportation and use of data. This subsection is dedicated to the systems that use the data and the considerations of their providers and/or developers. The aim of the framework is to use information collected about a user in order both to improve their current device as well as reducing the effort needed to personalise other devices. Rather than a specific standalone system, the framework is intended to be embedded in the operating system and will therefore most realistically be incorporated into future technology. It does however have the potential for application to both existing and legacy technologies.

The dynamic, open nature of the framework should allow it to be used both as an open-source resource and to be modified for use as a proprietary standard[1]. It is however understood that new technology is not adopted by commercial developers without financial motivation, and there is therefore a need to protect the investments made on development.

The framework has consciously been designed as an open format with the option to give control of vocabularies to the relevant authorities (World Health Organisation, World Wide Web Consortium (W3C) etc.). This approach is needed in order to provide as much interoperability as possible between a vast range of technologies and user capabilities. Interoperability however is reliant on use of a single vocabulary by all system developers which is clearly unrealistic, given the variety of uses for which systems are developed. There has been a trend for many commercial developers to use closed standards in order to protect their systems both from malicious behaviour and intellectual property theft. By using a proprietary standard developers can encourage or lock in users to avoid them purchasing competing technology.

7.2.3 AT Developers

While the previous subsection focused on systems that are based on the framework, the actual adaptations and assistive technologies that would be suggested as a result of their use have not been discussed explicitly. The framework is designed to improve the discoverability of adaptations and assistive technologies, both those embedded within a piece of technology and those that are retrofitted later. In order to be compatible with the framework, external developers will need to tag their products using recognised vocabularies, and make the data available for discovery. This may be a substantial amount of effort for smaller software houses, but can

[1] Although it should be noted that the intention of the author is the development of an open resource.

be incorporated into either the design or documentation phases of development.

Use of the framework to increase the discovery of support brings both advantages and threats to assistive technology developers. So far, only the potential to increase the consumer base has been discussed. For developers with good quality products that are well tagged and appropriate to the needs of customers, the framework brings potential benefits. The need to advertise could be reduced due to the matching between linking the product to the customer and endorsing it. The improved exposure can then widen the customer base and allow costs to be spread over increased sales.

Weaker products may however suffer. If a product does not meet users' needs—as described by the framework's functional assessment—it is unlikely to be suggested. The creation of a transparent and level playing field allows products to be selected on their functional performance, rather than due to marketing or popularity may currently be the case. Smaller developers may also face a lack of resources that prevent them from developing large producing multi-featured products. Their ability to produce small, well-focused accessibility 'apps' may however prove beneficial allowing them to build up a reputation for quality products that respond to specific user needs.

The development of assistive technology products as micro-ATs was proposed in Vanderheiden (2008). Ideally, all adaptation and assistive technology provision would be delivered out of the box and simply require personalisation, however where it is retrofitted it must be targeted. Just as personalisation is targeted to the individual, by encouraging assistive technology developers to target their solutions, unwanted side effects can be reduced.

7.2.4 Accessibility Community

The term 'accessibility community' is used to represent those people related to the field of accessibility as a whole. There are a substantial number of groups who are interested in developments in the accessibility arena. The users and developers of assistive systems and technology have already been discussed, this section is devoted to community that has built up around the research and provision of different forms of technology support. Stakeholders include: public sector bodies, academia and 'the third sector' all of whom have experience in either providing support to users or providing access to people who need support. Expertise in accessing groups of users is of benefit to both assistive technology researchers and developers.

The amount of conversation between different assistive technology stakeholders is currently minimal, but improving as the challenge is being recognised by academia as a whole. The accessibility field, due to its focus, is reasonably open and there are many links between industry and academia. Knowledge transfer events and partnerships are commonly used as vehicles to connect stakeholders together. They allow academics to present current research in order to understand its potential impact as well as finding out the questions that industry needs answering.

The focus of the community is gradually shifting away from profoundly disabled people and there is recognition that improving accessibility runs parallel to improvements in usability for all technology users [2]. Many of the benefits intended for specific groups have found their way into mainstream use, the tired example of "curb cuts" is being superseded by multiple technological examples. For example as well as improving readability for people with vision impairments, text size can also be manipulated to allow users with better eyesight to see more information in one go. Alternatively, shortcuts created to provide access to functionality for users who are unable to use a mouse can also improve the speed with which 'mouse and keyboard' users can access functionality.

The framework deals with all users in the same way, tracking their abilities and matching them to adaptations that will improve interaction, either to minimise an impairment or to increase performance. This will hopefully help to reduce the divide between disabled and non-disabled people by demonstrating that all users have capabilities that are measurable on a scale. By being able to provide suggestions to all users, systems based on the framework can highlight the potential for all users to benefit from adaptation. As the point at which ability is classified as impaired becomes irrelevant, the accessibility community will be presented with a chance to improve technology as a whole. The knowledge they possess will become relevant to the design of products for all people.

The use of semantic technologies were introduced in chapter 4 with data being exposed using Uniform Resource Identifiers and linked together to make it understandable. The intention is for definitions to be published publicly, this allows common representations to be used in the modelling of people and technology. Personal data would be kept private, however systems could take the data they have, aggregate, anonymise and publish it. This would allow snapshots of stereotypical users and ethnographic information to be made available, reducing the need for data collection exercises and improving the quality of data for statistical organisations.

[2]The W4A 2014 conference was for example themed on "The New Accessibility", describing accessibility as existing on a spectrum http://www.w4a.info/2014/.

7.2.5 External Entities

The final group of stakeholders, while included for completeness, is difficult to define. It is comprised of organisations who would like to use the data provided by the framework, and may therefore include members of the accessibility community. The data produced by the framework could be used to direct commercial accessibility products, research and inform healthcare initiatives. Although the definition of the group is vague, the issue of the use or release of framework data to external entities is an important one.

The responsibility for securely storing the data held in the framework falls to the system providers. Businesses can be built up around the generation and sale of data and a short section is included in this chapter on the potential of financing systems based around the framework. The Data Protection Act (DPA) (Data Protection Act, 1998) provides protection for personal data by ensuring that it is processed and stored appropriately, however if data is anonymised and therefore not attributable to the users it can be used for any purpose or passed on as the system providers wish. Additionally the DPA applies only to data stored in the UK and mandates that data can only be transferred to countries with equally strong legislation (e.g. not the USA). This may have the effect of restricting the transportability of data, by reducing the services that can be offered when a user is abroad or minimising the ability for data collected in the UK to be used to help users in other countries.

7.2.6 Common Issues

Having discussed the potential implications of the framework for each of the stakeholder groups, a number of common factors can be identified that will need to be addressed in order for a practical implementation to be developed. Most academic research requires that appropriate ethical considerations be examined before experiments are performed, in order to safeguard the participants on whom research is performed. While there is no single standard against which ethics can be measured, frameworks have been developed to provide guidance to those working with older people. The Research Ethical Guidance Framework (described in (Damodaran *et al.*, 2012)) is a set of guidelines to promote best practice for participatory research and covers issues such as: recruitment, informed consent, accommodating the needs of older people, use and privacy of data, and the feeding back of results once research is complete.

While academic activities are governed by the ethical codes of their institutions, the production, sale and use of commercial products are governed by pieces of legislation such as the Data Protection Act (Data Protection Act, 1998) and the

Goods and Services Act (HMSO, 1982). Much of the general legislation is however targeted towards consumer products and falls short of the standards proposed in academic frameworks described above. In the UK and USA where technology is classified as a 'medical device', it is subject to a higher degree of scrutiny through legislation designed to ensure a higher duty of care, as befits clinical settings (Waller, 2013). Given the diversity identified within older people and the range of accessibility issues they require, most (if not all) of the accessibility solutions considered within the scope of the framework will therefore not fall into this category.

7.3 Informed Consent

As the framework focuses on the collection and storage of user data, one of the pertinent ethical issues is therefore concerned with obtaining permission from the user to work with their data. While section 7.4 will cover the practical issues of controlling the storage and use of data, this section will focus on collection of informed consent. Originating in the field of medicine, informed consent is "[t]he process of agreeing to take part in a study based on access to all relevant and easily digestible information about what participation means, in particular, in terms of harms and benefits" (Parahoo, 2006). In addition Damodaran *et al.* (2012) identifies the importance of informing the participant what data will be recorded, where it will be stored and what it will be used for. Once the participant has been appropriately informed they are asked for their permission for the experiment to begin, and only with this consent can they be included in the research.

Damodaran *et al.* (2012) also suggests that when working with older people, greater care must be taken to ensure that truly informed consent has been obtained. In the first instance this involves making sure that the person has the capacity to be appropriately informed and secondly there is an onus on ensuring that the process of providing information and obtaining consent is appropriate (without bias).

7.3.1 Ensuring Users are Informed

Before their consent can be sought, a person must be be appropriately informed. They must both understand what they are giving their consent to (data collection/use etc), as well as comprehending the implications of their actions in giving consent (what are the potential outcomes). When working with older people, varying levels of literacy and exposure to technology may pose barriers to both of these requirements (Dickinson *et al.*, 2007).

7.3.1.1 Concerns Relating to Older People

People with low literacy levels may have difficulty perceiving complex ideas, owing to the problems they have with comprehending language. Damodaran *et al.* (2012) suggests that study details and instructions should be provided in both oral and written formats, with vocabulary tailored to the literacy level of the participant. This implies that it is an issue that can be overcome through changing the way that information is presented. As the issue is one of finding an appropriate communication medium, it is in effect an accessibility problem, could therefore be tackled through the application of the framework itself.

People with low levels of technology inclusion provide a more fundamental challenge. As well as presenting a barrier to communication (in terms of a need for 'jargon' to be explained) they are also less able to comprehend the implications that arise through the provision of their consent. For example there are potential issues surrounding the privacy, security and use of personal data that may seem inconsequential, but are actually quite severe. The requirement for accurate and up-to-date information could result in very sensitive capability data being gathered. Its aggregation over time could lead to the ability for pseudo-medical diagnoses to be made which could adversely affect their sense of well-being. Where improper storage leads to this data being made public, this could result in either embarrassment or the need for a decision to be made on whether action should be taken (e.g. the suspension of a driving license) (Sloan *et al.*, 2010).

Interestingly (Sloan *et al.*, 2010) observed that many older people did not raise significant concerns when the issues of privacy or use were described. This could have been for one of two reasons: (1) they did not fully comprehend the severity of the issues described, or (2) they were genuinely not concerned. The issues were subsequently raised during a workshop based on the presentation of Bell *et al.* (2010) and Li *et al.* (2010) and generally dismissed by the medically-biased audience. While it is considered out-of-scope for this thesis, the responses were made with reference to the field of medical ethics. In response to the first concern, a patient is aware that data collection and processing is required in order for a solution to be provided and can be summed up as: "no data/no assistance". The latter concern was covered by the right of patients to self-determination, meaning that a person has the right to decide the value they place on their own personal data.

7.3.1.2 Providing Information

In academia, informed consent is obtained though a strict system that involves producing information sheets, giving the participants time to absorb the information and then asking for their signed consent. Participant information sheets are developed within an ethical framework and focus on ensuring that participants' needs for privacy and safety are met. Commercial software on the other hand tends to favour End User License Agreements (EULAs); legal documents/contracts which are normally presented to the user during the installation process. EULAs are created more for the developer's benefit than the user's and centre around protecting the software from inappropriate use, pirating etc.

As research is normally conducted face-to-face with participants and covered by strict codes of ethics, many of the issues detailed above will not apply. The presence of an investigator matches the preference of most older people for face-to-face support and allows a dynamic approach to be taken to any problems that occur. This also allows the process of informing participants to be personalised based on participant feedback indicating how well they are informed (Shabajee, 2006). Where the participant does not appear to have an adequate understanding, information can be re-presented or further explanation provided.

The automated nature of adaptive systems (like those that the framework is intended to create) therefore create a potential problem. They may be required to personalise information in order to ensure that it is appropriate for informing users, while not having the data necessary to perform the personalisation. This 'catch twenty-two' situation could be resolved through the use of universal design principles, however, the draw-backs of universal design actually formed part of the inspiration for the framework. Given the importance of the rights of participants/users to decide for themselves what they want to consent to, content or service providers should not be deciding what is in their best interests.

7.3.2 Obtaining Consent

Once a user has enough information with which to make a decision, their consent can then be acquired. Like the provision of information, there are a number of methods by which consent can be obtained, each of which can be described in terms of its ethical rigour and the convenience of its use. As the ethical rigour of a method increases its convenience tends to decrease, owing to the additional burden it creates. Methods can be split into two high-level categories (opt-in and opt-out) and further classified by their acceptance of explicit or implicit indications of consent.

7.3.2.1 Opt-In

Opt-in methods take the position that nothing can be done without the prior permission of the user and is the method used by both the academic community and many commercial products. By seeking permission before any actions are carried out the user is kept informed, this adds legitimacy to any actions that are subsequently carried out. Where consent is obtained explicitly, opt-in methods are the most ethically rigorous form of obtaining consent and are used in academia as the gold standard in ethical practice (Damodaran *et al.*, 2012).

As the descriptions suggest, explicit consent is collected via a direct question where an answer is required either providing or denying consent; if the participant does not respond, consent is assumed to have been denied. Although the best option for academia, the need for explicit permission has obvious draw-backs in a low-disruption system. If no action can be taken without express consent being provided, the user will face regular interruptions. This may result in either consent being withheld (owing to frustration), or a reduction in the validity of the consent that is provided (as the user consents without being fully informed). For example Böhme & Köpsell (2010) demonstrates a growing tendency for users to 'click-through' set-up and permission screens and highlights the potential for the rise in ubiquitous EULAs to exacerbate this tendency.

As an alternative to explicit consent, implicit consent is consent implied by the actions of the user. Many websites used implied consent to improve the efficiency of the service they offered their users. Information collected about usage of the site can be used to improve the experience and highlight content that may be useful. Recently, the use of 'cookies' by websites in the UK has been restricted, requiring that users be asked for permission before they are used[3].

Implicit consent, although not as rigorous as explicit, is more appropriate in certain situations, often where the consequences of mis-using data are minimal. The use of user data to either improve services or increase sales does not pose any risk to the user either physically or mentally, and as immoral as up- or cross-selling may appear, the user is under no obligation to make more purchases than they feel comfortable. The only damage that is done is to the reputation or trust that the website has gained.

7.3.2.2 Opt-Out

The alternative to opt-in is opt-out consent, which requires the user to remove there consent rather than provide it. Opt-out methods take the position that consent is assumed and it is the user's responsibility to provide information to the contrary. While it may appear that opt-out is similar to implied consent, this is not the case, as no effort is actually made to obtain consent. Opt-out situations are often found on the internet under the guise of saving the user time,

[3] http://ico.org.uk/for_organisations/privacy_and_electronic_communications

when in reality they are used to attach extra services to an action that the user is sanctioning. The tendency for users to 'click-through' allows opt-out options to be made explicit, while remaining unlikely that consent will be withdrawn. As a result the office of the UK information commissioner suggests that failure of the user to opt-out does not imply consent (Data Protection Act, 1998).

7.3.2.3 Continued Consent

In shorter academic studies, the issue of maintaining consent is often overlooked as the likelihood of the participant revoking their consent is reduced. For longer studies, periodic renewal of informed consent may be required, especially where participants do not have the memory (cognitive capacity) to remember their rights to withdraw (Prentice *et al.*, 2007). In terms of commercial software, after a license agreement has been accepted the user is assumed to agree with the terms and conditions until they uninstall the software. Consent may only be sought again where the terms of the license change, for example in the case of the installation of an upgrade. This places responsibility for revoking consent on the user by requiring them to opt-out.

Both the academic and commercial communities use an opt-in approach to initially gain consent, but different approaches are however then taken to the maintenance of consent. Finding the right balance between the use of explicit and implicit consent is a major consideration within the proposed system. Explicit consent is both ethically superior and more disruptive. Users with reduced attention spans or those requiring more concentration (e.g. if they are completing a difficult task) will not appreciate being disturbed to the same degree as users that are better able to multi-task. Nervous users may appreciate being asked every time the system suggests a change, whereas users who are less concerned about security may be happy for their data to be anonymised and used for a variety of purposes without consultation.

7.3.2.4 Withholding Consent

A user may not provide permission for several reasons, firstly they may not want to because they have a legitimate objection, an appropriate response. Permission may also be withheld when the user does not understand what they are consenting to, they need more information to make a decision. In academia informed consent is required ensuring that the participant is aware of what they are agreeing to, this ensures both their physical and mental well-being. Many users shy away from making decisions when they are intimidated by unfamiliar language or feel pressure from a commercial source. This drop-out rate could be improved by adhering to informed consent standards, users that fully understand the systems and are happy with how it works are more likely to agree to use it.

If the user does not give permission for their data to be stored or used it cannot be and the proposed system will not work. In this case if the user wants to use the framework a compromise must be made, either by the system or the user. A single point of consent reduces burden on the user and makes the system seem less demanding, however it also increases the likelihood that consent will not be given. A gradual consent option will allow users control over what the system can do and potentially ease the informing process for each point.

The potential for the withholding of consent is not the only problem with opt-in systems. It is important that information given to users (just as it is with participants) is appropriate. If users are asked for permission with the implication that it is necessary to be able to use a system, they may give permission without taking the time to understand what they are consenting to.

7.4 Data Handling

In the previous sections the different stakeholders who may have an interest in the data manipulated by the framework were identified, and the use of informed consent as a means of ensuring that users' data is handled ethically was discussed. There are three activities for which informed consent may be required: the collection, storage and use of data. This section will now discuss each of the activities in turn and identify some of the practical issues which need to be considered before a practical implementation of the framework can be developed.

7.4.1 Collection

The success of a system that is based on the framework is dependent on its ability to collect data. Three potential agents were discussed in chapter 2 that may be able to provide the required data: the user themselves, human mediators and (semi-)automated technology-based agents.

Although chapter 6 demonstrated that participants could not be relied on to usefully predict or evaluate their own performance when playing a simple game, data collected from users has several advantages over data from other agents. This can be demonstrated by the preference for user configuration in many currently available technologies. Firstly, data can be collected from users with less effort than other agents. As the user is the one using a device, they are readily available to respond to requests for data and may be able to provide information that cannot easily be collected by other agents. Secondly, assuming they have the option not to provide it, data provided explicitly by the user may also carry an element of implied consent.

Lastly, even if the data cannot be used as a measure of a user's abilities due to its potential inaccuracy, it may still play an important role in the provision of accessibility solutions. The importance of maintaining a healthy/accurate self-image is described in Moores & Chang (2009). Data collected from users can provide an indication of their perception of their own abilities and this can be used to increase the acceptability of any suggestions made resulting from the use of the framework. Where a user is over-confident, support can be chosen to avoid threatening their self-image, while ensuring that they are aware of their deficiencies. This could be achieved, for example, by choosing settings that were near the lower threshold of their ability.

While older people have a preference for human-based support, the use of human-mediators as a method of data collection is mostly impractical due to the frequency with which data may need to be collected. Human-mediators do however have some potential advantages which may make them suitable for infrequent sources of data. Although the accuracy of data provided by human-mediators was not investigated in the study, (owing to their role in providing assistance and the external perspective from which they can view user's interactions), the data they provide may be inferred to be of a higher accuracy. In addition, depending on their relationship with the user, they may either be able to assume consent (where the user has requested their assistance) or provide a level of legitimacy which increases the likelihood of the user providing consent themselves.

Given the results from chapter 6, semi-autonomous agents are likely to provide the most reliable data, owing to the objectivity of their assessments. Although intrusive, by forcing the user to carry out set actions, specific data can be collected which may occur infrequently or otherwise be difficult to objectively measure, their overuse may however result in a distraction which will ultimately annoy the user. Their intrusiveness also has an impact regarding the ethical nature of their use. If the user is appropriately informed of their purpose and is given the ability to opt-out of completing them, a user's willingness to use semi-autonomous agents implicitly implies consent.

Autonomous data collection agents are also available to monitor user activity and their autonomous nature provides both potential benefits and drawbacks. Although they were not investigated, data collected without the user's knowledge (but with their consent) should be less biased as the user will not be under the pressure of assessment. This should make them more objective than other measures, but also presents a challenge regarding the collection of informed consent. Autonomously collecting data removes the need for the user to attempt contrived ability measurement tasks, but may be more likely to concern them if they are not sure when and what data is being collected.

If the user chooses not to consent to their data being collected and declines to provide any personal data of their own, then it will be difficult for any form of personalised service to be offered. If this is the case there are a number of options available to ensure that data is available for the framework to work with. As described in chapter 2, default or stereotypical profiles could be used to provide a rough approximation of the user's abilities, without requiring any personal data to be stored. Alternatively, less intrusive (and accurate) data collection methods could be used which extrapolate from any data that is available.

7.4.2 Storage

Although data collection and storage are discrete activities, they are sufficiently similar that many users may require educating as to their distinction. Whereas collection is the process of observing a user to identify useful data, storage is the process of recording data in a persistent format so that it can be retrieved at a later date, or transported for use in other situations. A user may consent to the collection of their data in order to allow assistance to be provided, but not to its storage. There are two main considerations which are likely to influence the provision of consent with regards to data storage:

Where the data will stored – influencing the level of control the user perceived themselves to have in determining the privacy of their personal data.

How long the data will be stored for – influencing the impact that individual events may have on a profile, through the longevity of their inclusion in a profile.

7.4.2.1 Location

Principle 8 of the Data Protection Act restricts the transfer of data, in order to ensure that it is not stored in a location that is likely to threaten its security (Data Protection Act, 1998). As described in chapter 2 a number of options are available for data storage based on either physical or cloud based media. The choice of location is a pertinent issues and while not necessarily manifested in the current older population (as seen in section 7.3.1.1), increased media attention may lead to increased interest in data security in the general population. As people become more aware of the need to protect their personal data, their likelihood of providing consent to its collection may be based, in part, on the location of its storage.

A number of trade-off can be observed between physical and cloud based storage methods, relating to both the efficiency of their use and users' perceptions of their security. Users are often worried about the unknown and as far as their data is concerned they are likely to withhold their consent to data storage if they are unsure where or how secure the storage will be.

One solution to this issue is to store data exclusively on the device on which it is captured and will be used. This may ease the uncertainty felt by some users if they perceive the data to be 'safe' on their own device, allowing them direct control over its access. This strategy requires that enough storage capacity is available to cope with the extra load that the data places on the device and also requires that processing is done by the device as well. As the framework has been designed to allow data collected on one device to be applied to the use of another, restricting storage to a single device is not an option.

As a slight improvement, the next step would be to allow data to be stored on a physical portable storage medium, able to interface with multiple devices. Many people have USB storage devices for documents, so storing the data in this form has the benefit of familiarity. USB is an established standard[4], most traditional computer hardware devices are USB compatible, some smaller device like phones or tablets do not however have USB capability. Requiring a user to carry an extra mobile storage device may not however be an ideal solution as it may be lost or broken; destroying the data. With this in mind, storage could be incorporated into an existing piece of technology such as a mobile phone, with Near Field Communications, Bluetooth or wireless capabilities used for connectivity.

The use of any form of physical local storage medium places the responsibility for the security and backing up of data onto the user and requires a standard to be in place which can be used to transfer data to/from all of the devices the user wishes to use. While it provides the control many users may be looking for by moving the profile online many of these issues could be handled for the user, at the expense of requiring a higher level of trust. Although the idea of "cloud computing" may currently be an overused buzz-phrase that is often employed without understanding the limitations of the need for an always-on internet connection, the use of cloud-based profiles is increasing. Access to the internet in the UK is improving, with high-speed broadband available in the majority of populated areas [5], 90% 3G coverage[6] and the recent role-out of 4G services.

The majority of larger/more adaptable devices (computers, terminals and televisions etc.) are often able to guarantee connection to the internet given the loc-

[4]Its use in existing projects has been described in Liffick & Zoppetti (2007) and Vanderheiden & Treviranus (2011b)

[5]https://www.gov.uk/broadband-delivery-uk

[6]http://media.ofcom.org.uk/2013/11/07/ensuring-3g-coverage-compliance/

ation they are used in. Many smaller, mobile devices are also becoming more connected.

7.4.2.2 Time

Principle 5 of the data protection act dictates that personal data may not be stored (in the UK) “for longer than is necessary” to achieve the purpose it was collected for (Data Protection Act, 1998). An argument could be made for retaining data for as long as a user uses a system, as the ability to tracking changes provides useful information. Historical data can be used both as a point of comparison for newly collected data, and as a source of data for trends to be identified. As every user is unique, the more historical data that is available, the more accurately trending can be performed.

While the persistent storage of data can be used to provide a reference facility, it is important to highlight the temporal nature of peoples’ abilities. Not only do capabilities change over time, but different capabilities will change at different rates. The General User Modelling Ontology (GUMO) for example defines a number of “expiry-classifications” which describe the expected length of time for which different types of collected data may remain valid (Heckmann *et al.*, 2005). Although some data is not likely to ever expire (gender or date of birth), the “time to live” for most data will range from seconds (for a heartbeat) through to years (for a personality type).

Users who are hesitant about providing consent to their data to be stored may find it more acceptable knowing they can set an expiry date, after which, the data will be removed. This time limit could be as short as the length of the session (effectively not storing it at all) or indefinitely if they wish to leave it available to assist other users. As a graduated component the length of time data is stored for can be varied, allowing the user to initially set the time short, and increase it as they feel comfortable.

7.4.3 Use

Assuming that a user has given consent for the collection (and potentially storage) of their personal data, it seems illogical that they would object to the data being used to identify appropriate support. However as identified in section 7.2 the information collected by the framework has the potential to be of interest to a range of stakeholders. As different stakeholders may use the data for purposes other than identifying accessibility problems or providing appropriate support, the user’s explicit consent should be requested. The process provides transparency to the user by describing how their data could be used and highlighting the ways that a lack of consent may reduce potential functionality.

Unlike the storage of data, if a user has consented to their data being collected, it can probably be assumed that they have given a basic level of consent to its use. This basic level of consent is necessary to provide the functionality described in chapter 5. While the comparison of personal data against different pieces of technology is part of the basic functionality, the use of data for inferencing has the potential to prove problematic. As described in section 7.3.1.1, inferencing could be used both to apply information about the user's performance in one task to their performance in another, as well as to provide pseudo-medical diagnoses.

Cunningham *et al.* (2012) describes a computer-based assessment that has been developed to assist in the identification of hand movement difficulties in people with Parkinson's disease. The task chosen and resulting data produced share many similarities with the tests developed in chapter 6. While the paper does not currently claim to be able to diagnose Parkinson's disease, it does suggest that it could assist in diagnosis by confirming the presence of Bradykinesia or rigidity. In addition it suggests that data could be analysed against other related conditions such as slowness of movement due to age and longitudinal monitoring be used to detect changes due to medication or disease deterioration.

Inferring information about individuals is only possible if a set of rules or corpus of data already exist from which rules can be developed. While rules can be developed through consultation with domain experts, the approach is similar to the use of human-mediation during collection activities, and as a result is prone to similar problems. An alternative approach is for rules to be created through machine learning (Sosnovsky & Dicheva, 2010); providing the potential for automation and the generation of new knowledge with regards to trends relating to age-related capability decline.

This form of inferencing relies on the ability to access and use user data. The use of information within a single device (or on multiple devices for the benefit of a single person) is an easy concept for a user to grasp; the collection of personal data for the purposes described in the previous paragraph may result in privacy concerns similar to those described in section 7.4.2.1. If information is not of a personal nature or is not attributable to an individual, it is not covered by the Data Protection Act. An anonymisation process could be used to allow data to be extracted without the user's permission, however many ethical frameworks would still insist informed consent be collected.

So far the use of data has been discussed in relation to the benefits it can provide for an individual and an inferencing system. A system base on the framework will also be able to generate information about individuals which may be used to identify trends within its user-base. These trends may be beneficial to other communities such as:

Government/Medical: Just as data could be used in the diagnosis of individual medical conditions, aggregated trends may be able to provide a picture of the declining capabilities in the ageing population as a whole. This information could be used to inform policy decisions.

Academic: There is currently a lack of detailed and long-term information about the needs of both disabled and non-disabled users (Atkinson *et al.*, 2008). As well as providing information to direct future research and foci for academic funding, the provision of an anonymised corpus of data would be beneficial as an artefact for use in the research itself.

Commercial: The commercial community includes entities that seek to make money from exploiting the information contained in the data produced by the system for their own financial gain. The trending data used to inform policy can be used as market research data which enables the targeting of products and services and the data from the anonymised corpus could similarly be sold.

The three communities listed are not distinct and elements of each can be found within others; for simplicity they will be discussed separately as stereotyped communities.

The different uses of the data identified so far can be used to highlight a trade-off between their potential benefits and the issues they raise in terms of user-privacy. As a final example the use of a framework-based system in a workplace situation will be discussed. The ability to use data, gathered from the use of personal devices, in providing assistance when using devices in the workplace is likely to speed up the process of making those devices accessible. The user may however not wish some of this information to be divulged to their employer due to fear of discrimination. If a system that was designed to provide assistance suggested new adaptations were needed, this could form evidence that a particular ability had degraded. Conversely if no change was detected, evidence would be available to suggest that the user was able to carry on with their work. The monitoring of employees is a contentious issue and could be seen as an invasion of privacy, however it could also be beneficial where evidence is required to support requests for time off related to medical conditions.

7.4.4 Profile Management

In order to provide control over data that is stored, it can be grouped together within profiles (as described in chapter 2 (Peissner *et al.*, 2012b)). The process of building up and managing profiles has its own issues. In chapter 5 a user profile was built up logically using a series of actors, which had a series of capabilities, measured through a series of data items. The focal point of a profile was its top-level actor and profiles could then be navigated using the defined relationships, with all data being attributable to a profile.

In this section, the issues surrounding the attribution of profiles to actual users and the advantages provided by the structure of the framework in terms providing control over data will be described.

7.4.4.1 Personal Identity

In the technological world the concept of identity is becoming increasingly important. Many pieces of technology, websites, products and service require the creation of a profile with which their use is associated. The use of profiles has several advantages, many phone operators allow the contents of mobile phones to be backed up to the cloud and certain services to be accessed online. Apple uses profiles to assist with the attribution of devices allowing both the delivery of purchased content and tracking of lost devices.

Many people have multiple email addresses, allowing them to separate communications relating to different parts of their life. Behaviour exhibited via profiles attached to social networking sites for example is often different to that displayed on job-finding services. They may also create multiple different personas for a single service, allowing the same separation to be achieved in order to personalise the service to their varying needs Gross (2009). One of the major divides for example is between social and work-related presences. The question arises therefore, over how to manage profiles controlling the assistive system suggested.

Just as people have different personas to allow for their different behaviours, there may also be differences in the way technology is acquired and used in different situation, for example at home, at work or in public setting. In a work environment, as technology use is mandated as part of a job role, the onus is on employers to provide safe suitable working environments (including any appropriate assistive technology) for their employees. In turn the employee is obligated to interact with the technology in order to do their job and technical support is likely to be provided in order to 'keep the business moving'. At home, while technology may be restricted based on a user's financial personal circumstances, they are less constrained by any workplace related restrictions. Users are therefore free to choose technology to suit their own preferences (rather than just needs), both in terms accessibility and any desires to follow technology trends.

The number of profiles that a user has does not matter, so long as they are

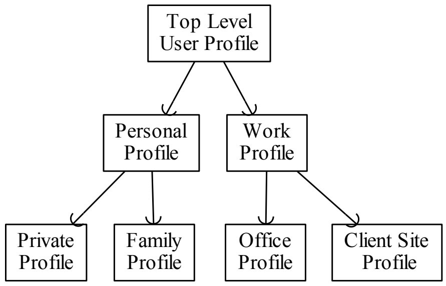


Figure 7.1: An example of the hierarchical structuring of profiles using actors and the '*has*' relationship.

able to log into the one that is appropriate for any particular device or situation. As profiles are based on the attribution of data to a top-level actor, the creation of a profile will involve creating an actor on which the profile can be based. A user could then have different profiles, based on different actors, representing their capabilities in different contexts. This allows the use of data to be controlled by physically partitioning data into different profiles.

Another approach is to limit each user to a single profile, with all data relating to a single user being stored together in a single location. Putting all available data into one profile has the advantage that trending and inferencing will be more accurate, owing to a more complete dataset being available. Restricting (or mandating) a single profile for every user, may not however work in the real world given the desire of some users for multiple personas as described above.

As any actor can be used as a sub-actor for another actor, different profiles can be linked through the relationships defined between their top-level actors. As shown in figure 7.1 a master profile could be created to act as a focal-point against which all of the user's individual profiles were attributed. Through the use of Unix file-system style permissions, access could then be granted to the data in different profiles dependent on the situation that a user was in, creating a permission system. For example when at work, read and write access would be granted only for the work profile to the user's work persona to be maintained. When at home, read and write access could then be granted for the user's personal profile, with read permission also given for the work profile. This would allow any issues identified at work (such as fatigue after an intensive typing session) to be used to inform decisions made about support needed at home.

7.4.4.2 Collaborative Working

The ability for profiles to be combined through the techniques described above also provides the potential for collaborative working to be supported. Where multiple people use the same device it can be personalised for each individual's personal whenever they are logged on. Where a device is used by more than one individual it will be important to know whether users are using it in isolation or at the same time. Collaborative working (where two or more users wish to access the same device, interface or data) presents another challenge that, although out of scope, is worth considering in order to ensure the framework is able to cope with the potential demands of future research.

Atkinson (2012) describes an approach that identifies the suitability of assistance for both individuals and collaborative working. By describing the abilities of individuals in terms of their human capabilities, their shared capabilities can be identified and interfaces customised or support provided as appropriate. One consequence of collaborative working is that compromises must be made between the requirements of the individuals involved with interaction being tailored to the needs of the group as a collective, rather than each of the individuals individually.

The ability of the framework to attribute data in multiple lower-level profiles to a higher-level profile complements the approach taken in Atkinson (2012) and could therefore be used to support it. Firstly, an actor could be created to act as the focal point of the collaborative profile, with the profiles of the individuals involved being attributed through the relationship already described. Any data related to individuals can be retrieved from and stored against their individual profiles. However if suggestions are accepted or rejected by individual users based on their own abilities the collaborative profile could become unstable, as it is "pulled" between the abilities of individuals rather than becoming an aggregation of all the users it represents. Where there is ambiguity with regards to the individual that is responsible for specific data items being collected, they can be attributed to the higher level collaborative profile instead.

7.4.4.3 Recognising Users

As alluded to in the previous subsections, the collection, storage and use of data in the form of profiles is dependent on appropriate profiles being identified and used. This section will briefly discuss the suitability of different methods enabling users to be recognised by systems that are based on the framework, easing the login process during which profiles are identified for use. The Global Public Inclusive Infrastructure acknowledges the need for "user initiation" and will seek to explore a range of technologies including (Vanderheiden & Treviranus, 2011b):

- URL activation and USB key based URL activation.
- Special data formats (USB and Smart Cards).
- Facial recognition and visual codes (e.g. 2D barcodes).
- Near Field Communication (NFC) phone and ring activation.

Approaches to deal with user initiation can be split into two broad categories: (1) those that require the user to remember login details and (2) those that are tied to a physical item or device. Approaches based on the former exclusively provide an identifier, however, approaches based on the latter have the potential to store both a profile identifier and the profile data itself. Ideally, a number of different approaches will be available, allowing a the user to choose the one that is most appropriate for their given situation.

One of the key issues that creates a challenge to ICT use by many older people is that of degrading memory. Damodaran *et al.* (2013) specifically identifies the remembering of passwords and recalling of sequential instructions as problems and this would make reliance on an approach which required the user to login a potential barrier. By linking profiles to existing email/password combinations, the extra burden of additional login details could be reduced, and login processes can be designed to guide users efficiently through. Requiring users to log in does pose a potential problem; accessibility settings will not be available until after the login process has been successfully completed, however users may require assistance in order to log in, creating a catch twenty-two situation.

Physical devices can alleviate this problem by allowing configuration simply due to their presence. RFID key rings are used by bar staff to log onto tills and at busy periods two or more staff can toggle between their separate orders, multiple times, in a short space of time. Being able to identify individuals quickly avoids the need to compromise between the needs of users who use a device separately, but within a short space of time of each other. Memory issues can affect users' abilities to find and use additional devices (for example keypads used to access online banking facilities), but there is the potential to use existing devices, such as NFC embedded mobile phones or any contact-less card that the user owns[7].

7.5 Data Management

So far the chapter has discussed approaches which have have been used to organise data into profiles and provide users with control over access to their data. A similarity can be observed between:

[7]As trialled by the SNAPI project http://www.snapi.org.uk/.

- The use of informed consent (from section 7.3) to mediate between the interests of the different stakeholders (identified in section 7.2), and ...

- The need for an approach to deal with the issue highlighted in chapter 6 regarding mediation between data provided by different sources, given their different levels of accuracy.

In both cases, an approach is required to mediate between different data items; dictating whether they are suitable for use by applying a threshold or gate-based test. Section 7.4 discussed the use of profiles both as a means of attributing data to an individual as well as providing a mechanism through which a user can control access to their data. Checking that consent has been collected provides a simple permission-based threshold that has already been widely accepted in both academic and commercial communities.

Chapter 6 identified how data collected/produced by different agents may vary in accuracy. As a result, either the *creator* themselves, or a measure of their apparent accuracy could be used as the basis for another test of data suitability; through their comparison against appropriate thresholds. Additionally, different uses of data will have different required accuracy thresholds. There is a higher requirement for data to be reliable if it is used by professionals to assist in making medical diagnoses than if it is used to assess the accessibility of a public terminal that is incapable of adaptation. As different systems are developed to provide different services, the framework will benefit from having the capacity to allow appropriate data to be selected for use.

In order to determine the suitability of data for different uses, any number of different threshold-based tests could be developed. The type of data required for these tests is data which provides information about the accuracy of a data item. Metadata is data about data, and can be used to ensure that data is used appropriately. The contextual data described in chapter 5 is in fact metadata about the data item to which it is related as it allows the suitability of a data item to a given situation to be described. Three different types of contextual data were listed, all relating to the context surrounding the capability described by the data item at the time it was observed.

Rather than describing capabilities, this chapter has focused on metadata that describes the data item itself, and as such is known as 'Data Context'. Unlike the other types of context, data context is not stored as an actor, but rather as a series of discrete data items, each of which has its own data type and methods by which it is processed. By noting their similarity both issues can be addressed with the same approaches, which will be developed through the rest of this section.

The Dublin Core Metadata Initiative[8] is an organisation which "supports shared innovation in metadata design and best practices". Its primary output is the Dublin Core Metadata Terms (DCMI, 2012), a recommendation which includes a set of metadata terms which have been published within various standards. The terms can be used to describe both physical and online resources and are the de facto standard for the provision of metadata in semantic document systems (Sosnovsky & Dicheva, 2010). As the Metadata Terms deal with resources they define a number of terms which may be of use in the description of the data context of a data item.

AccessRights: "Information about who can access the resource or an indication of its security status."

Creator: "An entity primarily responsible for making the resource."

Created and Valid: "Date of creation of the resource" and "date (often a range) of validity of a resource."

Each of the metadata terms described above can be the focus of a threshold-based test to determine the suitability of a particular data item for use. Section 7.4.4 described how profiles could be managed to produce through a permissions system, which could be controlled by metadata, similar to the "*accessRights*" term from the Dublin Core. When data is requested from a data store by a system that is based on the framework the access rights of each data item can be checked against those of the system requesting the data. In this way, the use of sensitive data items can be restricted to a small set of trusted systems which need access to it.

Whereas permissions are used to protect data, other metadata terms can be used to provide an indication of its accuracy. As suggested in chapter 6, people are unable to provide appropriate data for the assessment of their own capabilities. Where higher accuracy data is required, data could be passed through a filter which removed data items based on the *creator* term from the Dublin Core. Section also described a time-based measure which could be used to assess the accuracy of a data item based on its age and predicted expiry time. Another filter could therefore be created using the *created* and *valid* terms, which was able to filter data items by checking that $[(created + valid) > currentDate]$.

[8] http://dublincore.org/

7.6 Summary

Many user modelling related research projects deal solely with their theoretical content and ignore the needs of users, regarding their ethical concerns. In order to move the theoretical designs of the previous chapter forward towards an implementable system, this chapter has been used to discuss both the softer issues surrounding the use of the framework and the need for a method of mediating between data provided by data sources.

Informed consent forms a fundamental aspect of the process of building trust, which will be required in order for a successful implementation of the framework to be developed. From an academic perspective, research cannot be carried out without the consent of participants to have their data stored and manipulated in order for adaptations to be suggested. From a practical perspective, legislation such as the Data Protection Act prevents the storage of personal data in commercial systems without prior consent from the user.

7.6.1 Contribution

Based on the above discussions, a permissions system has been proposed that allows control to be maintained over the storage and use of personal data and in doing so, an approach has been developed that also allows mediation between data from different sources to be performed. Through the use of thresholds, data can be chosen for use based on its suitability, measured in terms of:

- The permissions settings of the data and the permissions granted to an agent wishing to use it.
- Alternative pieces of metadata that can be used to imply quality and the quality level required for an intended use.
- Other forms of context and the required level of similarity to the context under investigation.

7.6.2 Impact

In formulating the mechanism described above, this chapter provides the theory necessary to address both of the intended aims and allow the design to be practically implemented. Aim 5 identified the importance of the ability for a variety of agents to be able to produce and use the data within the framework. While chapter 5 provided a design with which interoperable models could be created, the approach taken within this chapter has resulted in the use of a range of metadata from existing standard, which can be used to describe data *produced* by a variety of sources. In addition, metadata describing the permissions associated with particular pieces of data can then also be used to restrict the *use* of data to recognised agents.

Aim 6 then identified the subsequent variability of data quality that arises as a result of the variety of agents displayed in aim 5. Through a similar threshold style approach to that created to achieve the control of data in aim 5, the other forms of metadata (data context) identified above could also be compared against thresholds to allow the quality of data to be controlled. Thresholds could then be applied to the other forms of context (identified in the previous chapter 5) to control the wider quality of data in relation to its context of use.

With the design of the framework now able to support all of the goals identified in chapter 2 the thesis is now in the position to start performing an evaluation. As the framework is currently just a theoretical design, the first stage of the evaluation process will be the production of a prototypical system to assess the feasibility of its practical implementation. The next chapter will therefore discuss the suitability of a number of technologies for use in the implementation process as well as the construction of the actual prototypical system itself.

Chapter 8

Implementation

With the theoretical basis of the framework developed in the previous chapters, this chapter now describes the practical implementation of a system for modelling the accessibility of interaction between different actors. Although the theoretical comparison of actors has been discussed, the framework has thus far been solely described in terms of its structure – as dictated by the particles and their relationship that have been defined. In order to allow (and as part of) the evaluation of the framework, its ability to form the basis of implementable systems needs to be understood.

This chapter will begin in section 8.1 by discussing the design of the framework in order to highlight the advantages it affords to implementations. Section 8.2 then identifies a number of technologies and techniques that can be used to provide practical implementations and describes how they could be used to allow the development of a prototypical system. The modular design of the system is introduced in section 8.3 with more detailed descriptions of individual modules and how they have been implemented provided in the following sections (8.4, 8.5 and 8.6). This will be followed by a short worked example in section 8.7 that will be used to bind the implementation work together. Finally, section 8.8 provides a summary of the impact of the chapter by describing the contribution it makes to the thesis as a whole and how the prototype supports the evaluation of the framework in the following chapter by demonstrating the feasibility of its practical implementation.

8.1 Modularity

The framework has been designed using three layers, which allow differentiation between the structural, data and contextual elements, that may be considered during an accessibility assessment.

Structural layer components identify potential channels of communication through

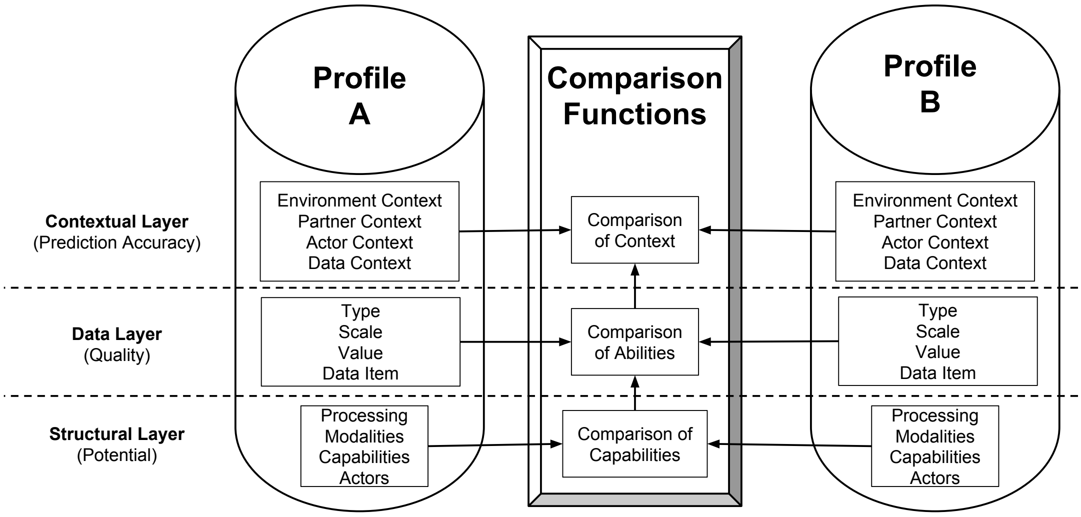


Figure 8.1: Using modular construction to provide separation between data and processes, based on layers of abstraction.

the creation of actor profiles and comparison of their attributed capabilities.

Data layer components then quantify the quality of communication within each channel through the comparison of ability data.

Contextual layer components predict the accuracy of the preceding predictions through the comparison of contextual information.

Each of the layers has a different focus, allowing comparisons to be made with increasing amounts of detail. This additional detail requires the provision of additional particles which can be used to provide additional functionality, as seen in figure 8.1. As the layers build on top of each other, higher layers are dependent on the output of lower layers to provide a point of reference on which their output can be focused. However, the output of higher layers can also be used to filter the data used at lower layers.

The use of a standard set of particles, linked together by a standard set of relationships, allows profiles to be built using a series of functional layers. Each additional functional layer provides additional data, thereby increasing the information available and the informative content of the comparison. Although linked, the data provided at each layer is effectively separate, allowing the separation of concerns between layers.

The separation of concerns philosophy (as described in chapter 4) advocates the creation of modular programs by separating the storage and processing of data. In addition, a system based on the framework should also be able to separate the storage and processing of data between different layers of abstraction. The use of a modular design has a number of advantages relating to both the development process and the resulting system:

- Different modules can be designed, coded, tested etc in parallel – reducing development time.
- Responsibility for each module can be given to 'experts' with appropriate experience – providing improved quality.
- Different versions of the same module can be produced – providing different biases for different purposes.
- Implementations can then be customised through their inclusion/exclusion of different modules.

8.2 Technologies and Techniques

The description of the framework that has been provided so far has been sufficient to produce a theoretical modular design. The use of standard particles connected by flexible relationships, rather than a rigid data structure, has resulted in a design that is not dependent on any specific technology. The design that has been suggested does however present its own requirements and biases.

The graph-based structure is reliant on a language that can represent dynamic networks of data, rather than a static flat file structure. In order for comparisons to be made, inferencing technologies are required that are able to provide pattern-based representation and matching. Finally, as comparisons are made at gradually higher layers, techniques are required to take account of the potential uncertainty inherent in the use of data that has been collected from various sources. This section will now discuss the desirable characteristics of the techniques and technologies that can be used to provide an implementation of the framework with specific attention to those that have been chosen for prototyping purposes.

8.2.1 Representing and Storing Data

The data structure described in chapter 5 is a directed graph, made up from a collection of standard particles connected by a series of standard relationships. Atoms (of different particle types) can be created to provide generic components, which can be subsequently instantiated as profiles for describing a specific user, technology or assistance. New data can then be appended to a profile, extending its graph and increasing the potential knowledge-base available with which comparisons can be made. As a result, a format is required that is able to represent collected data in the proposed graph-based structure and facilitate its storage, transportation and manipulation by numerous systems.

The Semantic Web is a series of standards that have been developed by the World Wide Web Consortium (W3C) to allow the encoding, storage and manipulation of semantic information over the internet. They have been designed to provide a platform independent format that can be used within large distributed systems. Produced for a growing community of developers, a range of resources are available including libraries for most major programming languages, development environments and data-stores[1].

As described in chapter 4, the semantic web stack is built up in a hierarchical fashion, similar to the framework itself. It moves from low-layer addressing (URI) and encoding (XML, RDF) standards, to higher layer data organisation (OWL) and querying standards (SPARQL). The standards can be used to deliver the functionality required by the framework including: encoding, storage and comparison of semantic data.

The Resource Description Framework (RDF) is itself a series of specifications for the storage and representation of semantic data[2]. RDF represents data as a series of triples, consisting of three elements in the form [*Subject, Predicate, Object*], where the *subject* denotes a resource which is related via the *predicate* to the *object*. Each element in a triple can take the form of a URI (often presented as a URL) which provides an address representing an object or concept. The subject and object can also take the form of a 'blank node' (used to aid the construction of graphs by allowing a number of triples to be linked together through an anonymous resource). Finally, the object can be represented as a 'literal', allowing data in various formats (including strings and integers) to be stored. Triples can then be connected together via their common subject and object URIs in order to create directed graphs, with the subject and object representing nodes and the predicates representing edges.

As described in chapter 4, ontologies provide a structured method of representing knowledge. The Web Ontology Language (OWL) is a specification providing a set of semantic relationships and syntax for building and storing ontologies using RDF[3]. A number of ontologies are available already to provide information that can be used for many of the particles within the framework (inc. people, capabilities, assistive technologies and context). Where suitable descriptors are not available in existing ontologies, the web-based nature of OWL allows new ontologies to be created and shared over the internet.

[1] http://www.w3.org/2001/sw/wiki/Category:Tool
[2] http://www.w3.org/2001/sw/wiki/RDF
[3] http://www.w3.org/2001/sw/wiki/OWL

This use of relationships to store data in terms of its semantics structure is identical to the graph-based structure of the framework. A number of serialisations/notations are available, allowing the same patterns to be displayed in different formats. Owing to its machine readable XML-based structure, 'RDF/XML'[4] is often used for sharing semantic data on-line. However for ease of comprehension 'Turtle' (Terse RDF Triple Language)[5] will be used within this chapter.

True to its name, Turtle has minimal syntax and allows the use of prefixes to aid readability of URIs. Prefixes allow URIs to be abbreviated through the use of short prefix names to represent long prefixes in repeated URIs (e.g. '*ex:concept* to represent '*http://example.com#concept*') Triples are then declared by listing their elements in order—Subject, Predicate, Object—with a full stop acting as an end of line character (e.g. '*ex:Subject ex:Predicate ex:Object.*').

Listing 8.1 provides an example of a short RDF file using Turtle notation (hashed lines are comments). The first block defines prefixes representing: [line#2] the RDF vocabulary, [line#3] the Foundational Modal of Anatomy, [line#4] the World Health Organisation's International Classification of Functioning, Disability and Health and [lines#5-7] example domains representing the Adaptive Capability Profiling Framework (ACPF) vocabulary and example capabilities and modalities. The rest of the listing ([lines#9–25]) describes the graph shown in figure 8.2, which represents snippets taken from example user and screen profiles. The relationships are represented within the figure through the use of shapes and colours, in order to reduce clutter and remain consistent with the representations in chapter 5.

In the example a user (user@example.com) is represented by their email address, which is identified as an actor through the '*rdf:type*' relationship (and shown in figure 8.2 by its rectangular shape). The actor representing the user is then assigned a sub-actor (fma:12513) via the '*acpf:has*' relationship. The URI 'fma:12513' is a reference to an ontology of anatomical structures that has successfully been implemented using OWL (Golbreich *et al.*, 2006) and represents an eyeball. An eyeball has a number of capabilities, however the listing specifically identifies visual acuity (icf:b21001) which inputs via mod:visualSize (an example reference to a modality that is focused on the size of a visual object). Similarly a monitor (monitor@example.com) is an actor with a capability to output via the same modality.

Listing 8.1: Turtle Notation for Figure 8.2

```
1 # Prefixes
2 @prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>.
```

[4] http://www.w3.org/TR/rdf-syntax-grammar/

[5] http://www.w3.org/TR/turtle/

```
3 @prefix fma: <http://www.bioontology.org/projects/
    ↪ ontologies/fma#>.
4 @prefix icf: <http://who.int/icf#>.
5 @prefix acpf: <http://acpf.ex#>.
6 @prefix cap: <http://capabilities.ex#>.
7 @prefix mod: <http://modality.ex#>.
8
9 # User Profile Snippet
10 user@example.com rdf:type acfp:actor.
11 fma:12513 rdf:type acfp:actor.
12 icf:b2100 rdf:type acpf:capability.
13 mod:visualSize rdf:type acpf:modality.
14 #--------
15 user@example.com acpf:has fma:12513.
16 fma:12513 acpf:can icf:b2100.
17 icf:b2100 acpf:inputsVia mod:visualSize.
18
19 #Screen Profile Snippet
20 monitor@example.com rdf:type acpf:actor.
21 cap:imageSize rdf:type acpf:capability.
22 mod:visualSize rdf:type acpf:modality.
23 #--------
24 monitor@example.com acpf:can ex:imageSize.
25 cap:imageSize acpf:outputsVia mod:visualLength.
```

While the listing provides an example file, in reality different users' and technologies' data would be stored in separate places and a reasoner used to identify the nature of relationships between them. Data can be stored and transported in flat files (e.g. the listing) or in 'triple stores' (databases customised for the storage of triples). Different files can be created, or triple stores partitioned, to allow the separation of data via abstract layer. Each of the snippets represents data that would be stored in structural layer sections of the user and screen profiles respectively.

- The upper portions of the the snippets ([lines#10–13 and #20–22]) provide type definitions to describe what kind of particles each of their predicates (representing atoms) represents.
- The lower portions of the snippets ([lines#15–17 and #24–25]) then describe the relationships between those atoms, using the relationships defined in chapter 5.

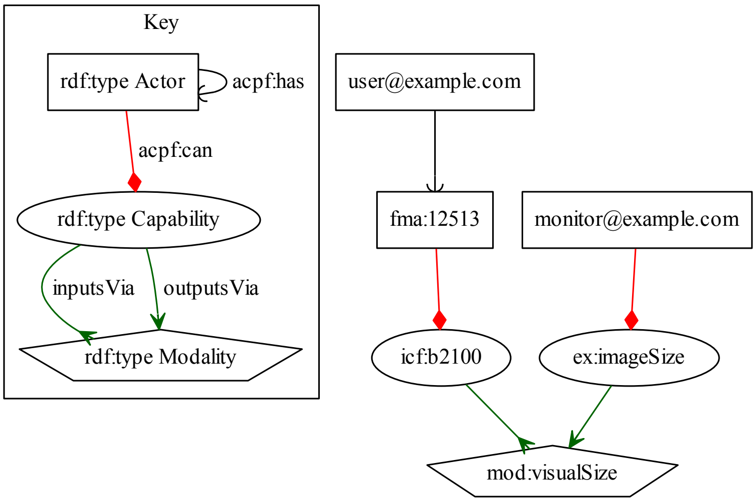


Figure 8.2: A graphical representation of the example structural layer graph.

8.2.2 Providing Reasoning Capabilities

The framework is designed to perform comparisons in order to assess accessibility between two profiles. Due to the graph-based structure of each of the profiles created using the framework, the ability for comparisons to be made between them is dependent on the identification of the presence or absence of similar modality nodes in each of their graphs. If this is the case the graphs of the profiles will overlap. A basic comparison 'between' two profiles can therefore be performed through the adherence of the profiles to a series of rules. However, the use of assistive technologies, and the ability to nest actors, results in a need to trace the information as it flows through multiple profiles.

SPARQL **P**rotocol **a**nd **R**DF **Q**uery **L**anguage (SPARQL) is a standard which sits above RDF in the semantic stack and provides the ability to specify SQL style queries on RDF graphs[6]. It can be used for exposing information from multiple interlinked graphs, based on their adherence to a specified pattern. Rule Interchange Format (RIF) is a standard which occupies a similar position to SPARQL in the semantic stack and has been designed for exchanging rules between rule systems, in particular Web rule engines[7]. While SPARQL and RIF provide the ability to specify queries and rules, they rely on the use of reasoning engines for their implementation. The semantic web stack has been chosen for its focus on interoperability and future research could involve the definition of rules and processes using semantic web standards. The production of a proof-of-concept

[6] http://www.w3.org/TR/sparql11-query/
[7] http://www.w3.org/TR/rif-overview/

solution will however be more easily be provided through the use of an integrated inferencing technology, owing to the variation in support that current reasoners have for the above standards.

In terms of the design developed in chapter 5, reasoning capabilities have the potential to both facilitate some of the most powerful features of the graph-based framework and simultaneously simplify the use of the numerous relationships that are required. Firstly, pattern matching is the method by which two profiles will be compared to assess their accessibility, as will be described in section 8.5.1. As a querying language, SPARQL could be employed to provide the pattern matching required to identify matches, based on rules stored in RIF format.

Secondly, the meronomic nature of capabilities is mirrored in the modalities through which they are able to transmit or receive information. Through the use of the '*infers*' and/or '*inferredBy*' relationships (sections 5.2.2.1 and 5.2.3.2) capabilities and modalities can be inferred based on the existence of their parent or constituent (higher- or lower-level) capabilities and modalities. When an actor acquires additional sub-actors, the lower-level capabilities they provide may result in the actor acquiring additional higher-level capabilities. This acquisition may be facilitated via inference, using rules encoded in RIF, which references the inferential relationships between modalities and provides a mirrored set of capabilities in the actor.

SWI-Prolog is a multi-platform implementation of Prolog containing libraries that facilitate both semantic web and graphical user interface (GUI) programming. Prolog is based on first-order predicate logic and programs are written through the definition of a series of rules which are then tested against a knowledge-base that is made up of 'facts'. The profiles provide the knowledge-base and the rules can be used to for pattern-matching.

As Prolog programs are written as a set of rules, they can be developed in a modular way. Functionality can be altered through the inclusion and exclusion of individual rules (or rule libraries), and higher layer rules can be built using a series of lower layer ones. Libraries can be produced, which provide different rules to suit different situations and this mirrors the use of abstraction within the framework.

8.2.3 Dealing with Uncertainty

Given that the framework has been designed to facilitate the modelling of interactions between real people and pieces of technology in different situations it is likely to require a significant amount of data. As well as the variability in actors' abilities that is caused by a number of contextual factors, chapter 6 identified the potential for the data itself to be inaccurate depending on the objectivity of the agent providing it. The contextual layer of the framework attempts to address the need for a description of the suitability of individual pieces of data for use in different situations. It does this by providing information about the context in which a piece of data was collected.

Although a structure has been defined to allow the provision of contextual information about different data items, additional techniques are needed to represent the resulting level of uncertainty that arises from their actual use. As an example, a piece of data collected could be collected in a given context (χ_0) and then reused to represent an actor's abilities in a number of different contexts ($\chi_1, \chi_2 \ldots \chi_i$). The smaller the difference between χ_0 and χ_i the greater the accuracy of predictions made using the data. As the difference between χ_0 and χ_i increases, the data becomes more unreliable and this should be reflected in its use within the data layer.

When measuring a capability against many lower level objective scales, an actor is likely to be able to display a range of abilities (e.g. many humans can hear sounds in the range 20–20,000Hz). However, this range will vary depending on either contextual factors (as described in section 5.4) or the accuracy of the data item itself (as described in section 7.5). The maximum and minimum values of an ability range could therefore be described using error bars rather than fixed points. The error bars would represent the 'confidence interval' of the values, with their size being dependent on the accuracy (or level of confidence) of the contextual comparison (Field, 2009).

Confidence intervals could also be used in a similar way to describe the comparison of data items, as seen in figure 8.3. In the figure, two capabilities are being compared to identify whether or not they overlap (and are able to support a channel of communication). Each actors' capability is described as a range, within which their actual ability lies. As there is a lower probability that the actors have abilities that lie at the outer edges of the range, they are given a lower probability than those nearer the middle of the range. This could be reflected by error bars that are calibrated to represent a given confidence level and will result in the amount of overlap between the capabilities being dependent on the confidence level at which they are described.

Confidence intervals are based on the use of probabilities to describe the likelihood that a given result will occur. Probabilistic modelling is a field in its own right and although there are numerous techniques available to allow for quantifying uncertainty (Ghahramani, 2013), rather than being incorporated within this thesis, they are currently out of scope and provide interesting avenues for future research.

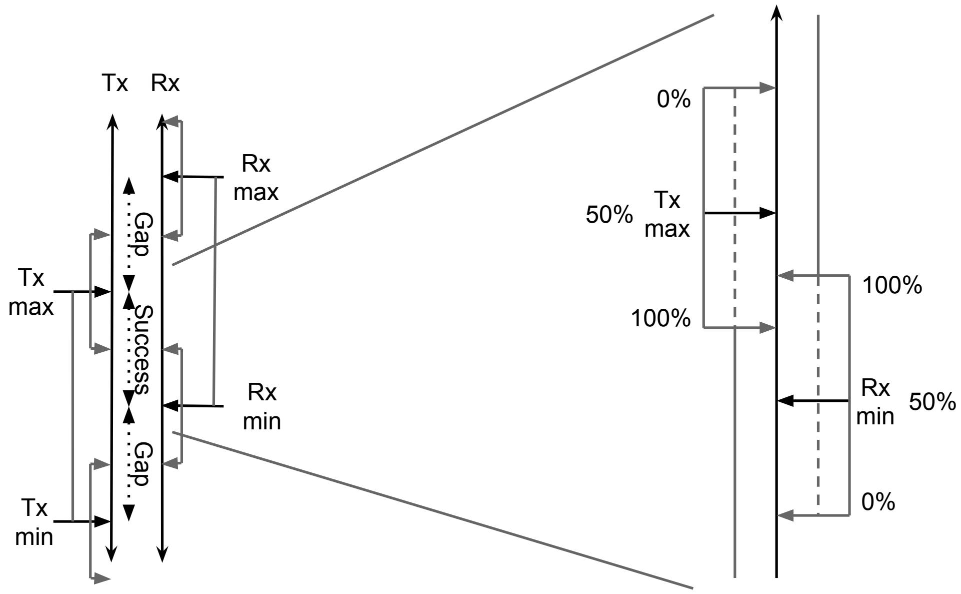


Figure 8.3: Using confidence intervals to provide a measure of the accuracy of a prediction.

8.3 Functional Modules Required for Implementation

The next three sections (8.4, 8.5 and 8.6) will discuss the implementation of a system that is based on the framework. They will include descriptions taken from both an actual prototypical implementation that has been developed for evaluation in the next chapter, as well as illustrative examples of extensions or alternatives that could be used in further developments. As seen in figure 8.4, a system could be implemented in a modular fashion. The figure will be gradually filled in throughout the chapter and the completed version can be found in figure 8.10 (on page 227). The separation of concerns described in section 8.1 allows the separation of: (1) data storage, (2) reasoning and (3) the provision/use of the outputs that are generated.

The framework that has been described in previous chapters forms the basis of the reasoning concern that takes data about actors from a data-store and provides information about their potential accessibility as an output. As a result it can then be divided again based on the functional layers of the framework (structural, data and context). The implementation of each concern will be discussed in turn with emphasis on their functionality and the approach that has been implemented for evaluation purposes.

Where appropriate a running example will be used to illustrate the functionality of the framework. Each section will present a short discussion regarding the generic considerations of implementing the framework which will then be followed

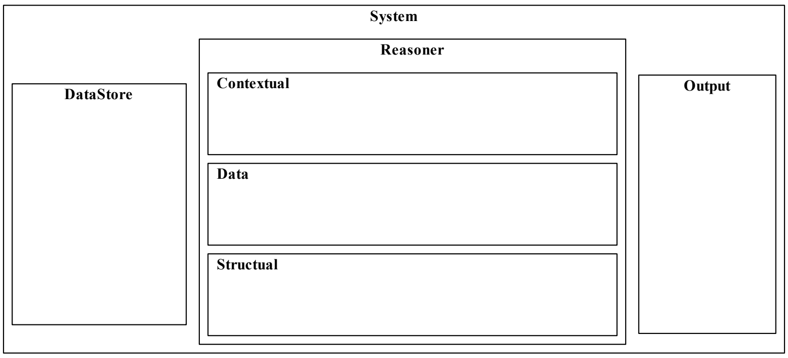


Figure 8.4: Implementation – The basic modules.

by examples taken from a prototype that has been developed. The prototype is a stand-alone system that uses the elements of the framework to identify the accessibility of interactions between a user and piece of technology.

8.4 Data Storage

There are two main options for the persistent storage of triples; flat RDF files and relational database based triple stores. In addition to the options available for storing RDF graphs, the RDF graphs themselves can be partitioned in different ways to provide additional benefits.

As described in chapter 7, profiles may be stored either locally or remotely and the storage approach taken will be the one that best addresses the concerns that each raises. The principle concerns of local storage formats are likely to be linked to their use of local system resources (e.g. storage, and processing). The use of remote storage instead raises speed and security based concerns due to the method by which data will be accessed.

8.4.1 RDF Files and Triple Stores

Flat RDF files have a lower requirement for storage space, but must be read (in their entirety) into memory before their data can be used. As they are simple text files, they have a lower barrier to entry than triple stores, which are reliant on the instantiation of a database. However, although they are both easily accessible and transportable (RDF files can be hosted 'as is' online and have the potential to be human readable), they have no inherent security mechanisms.

As the alternative option, triple stores use relational databases for the storage of triples. While they have a higher requirement in terms of hard disk space, they provide a more efficient method of manipulating higher numbers of triples (large datasets can require millions of triples), allowing only those needed to be returned rather than all-or-nothing approach of flat files. In addition the use of SPARQL "endpoints"[8] provide an external interface for web-accessible triple stores. As single static points of entry into what could potentially be a large dynamic data-store, endpoints can be used to both to improve accessibility and control access through the use of authentication protocols.

8.4.2 Partitioning Graphs

As well as the options above for storing graphs, the graphs themselves can be stored in different ways (through different forms of partitioning) to achieve different benefits. The use of functional layers within the framework has already been discussed with regards the benefits that the separation of concerns provide. The partitioning of graphs into separate data-stores, based on the particles they contain, mirrors this approach and can be used to speed up data retrieval given that smaller graphs will be quicker to search.

In addition to providing a smaller search space, the partitioning of graphs also provides the potential for data retrieval functions to be optimised based on the type of data that they have been developed to query. For example, structural layer stores may be indexed either by actor (for the retrieval of profile data) or modality (for finding capabilities that match channel-based queries). Another method that is specifically related to the use of triple stores is that of "vertical partitioning" of data – a technique that separates data (triples) into tables based on their predicate (Abadi *et al.*, 2007). On data sets of 50 million triples (which could be generated by the framework if a remote mass storage approach were taken for the storage of many different user's data) query times were reduced from minutes down to seconds.

Partitioning data based on the actor (or capability) to which it is attributed may also be used to provide security related benefits. As described in chapter 7, if different actors are used to contain data with different access rights, by storing the data attributed to each actor separately, access can be controlled by only providing data related to appropriate actors. In this way flat RDF files can be used, with actors (and their related data) stored in separate files and files provided based on the access rights of a given process.

[8] http://www.w3.org/wiki/SparqlEndpoints

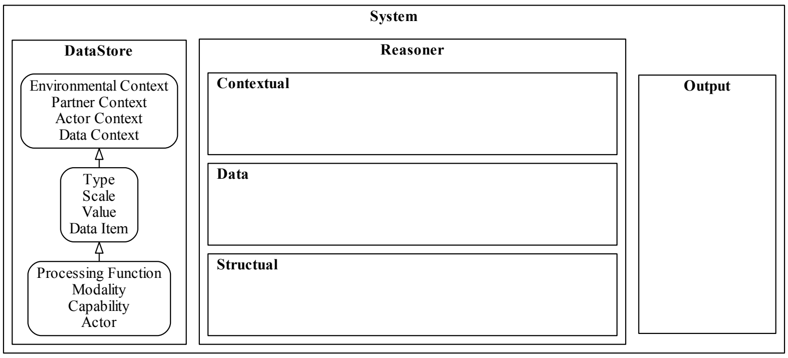


Figure 8.5: Implementation – The data storage module.

8.4.3 Data Storage Within the Prototype

For ease of implementation the prototype uses a series of text-based RDF-XML data-files. The ontology editor Protege[9] was used for the creation and editing of data-files, however a series of proof-of-concept mini-games were also created that provided their output in RDF-XML. Profile data was partitioned into a number of actor-based data-files with each file storing data relating to a different actor. This allowed actors to be created and edited independently.

Profiles relating to different pieces of technology were stored as a series of separate actors, with lower level actors stored in separate files that were used by a higher level actor (again stored in its own file). A different approach was taken for representing users; while both the actor and all of its sub-actors representing a single user were stored in the same file, the data representing their data items (created through the use of mini-games) were stored separately. This demonstrated the potential for data collection to be performed by a number of different semi-autonomous agents. Through the inclusion of their data files the outputs of different agents was then inserted into existing user profiles, allowing the data they generated to be used within the framework.

SWI-Prolog provided the facility to read-in all of the data-files, storing them in memory in order to allow querying and inferencing to be performed. The variety of potential methods that are available for the storage of data demonstrates an ability for independent data-stores to be created to facilitate storage and subsequently retrieval in real systems. While only one method of data storage was used in the prototype, 'data loading' functionality was provided in a separate library to the reasoning processes within the framework. This provides the potential for different methods of storage to be developed and used for different situations

[9]http://protege.stanford.edu/

without affecting the data inferencing processes.

8.5 Reasoning

The reasoner is the module within a system that allows two actors to be compared and a measure of their compatibility to be produced. *Comparison* is a process that involves identifying the similarities or dissimilarities between two (or more) items; it is different to *matching* which involves the identification of items which (when compared) are similar. The reasoner has the potential to be included within (or expanded to provide) a matching service[10] and will use matching as a means of achieving its functionality. As such its ability to provide both comparisons and matching will be discussed.

Given that the framework (on which the reasoner is an implementation) is separated into a series of functional layers, the same separation can be provided within the reasoner. Each of its layers can being provided by a separate (sub-)module resulting in similar benefits to the separation of concerns described in section 8.1. Although they provide different types of comparison, each layer of the reasoner can be broadly described in terms of the same three basic functions that where identified in section 2.4:

Retrieval of data from the available dataset (efference).

Comparison proving layer-specific functionality (inference).

Exposure of the result of the comparison (afference).

Once again, each of the functions can be provided by its own (sub-sub-)module with a series of plug-ins being used to provide specific parts of the functionality (as will be described in the following subsections).

Each layer of the reasoner should be able to work autonomously, carrying out the comparisons that it has been designed for, as with the rest of the framework. The two upper layers are however each dependent on the layers below them to provide them with a focus for which their comparisons are carried out. Contextual layer comparisons provide information about the applicability of data items (from the data layer) to a situation, and the comparison of data items provides a measure of the quality of interaction (bandwidth of a channel) between two capabilities (as identified in the structural layer). The two lower layers may also benefit from the additional knowledge provided by the layers above them; with higher layers filtering out unsuitable data before lower layer comparisons are run.

[10]such as that proposed in the architecture being developed by the Global Public Inclusive Infrastructure

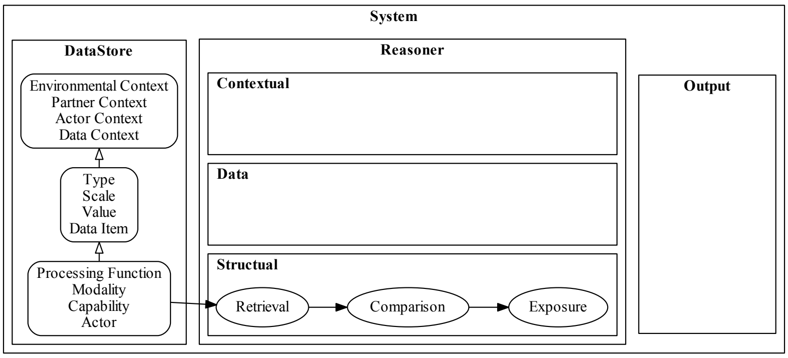


Figure 8.6: Implementation – Including the structural layer reasoning module.

The implementation of each of the functional layers will now be discussed in turn with emphasis on: (1) the input and information it requires, (2) the comparison functionality provided and (3) the uses of its functionality in terms of outputs provided. Where appropriate, the plug-ins required to support the functionality of each layer will also be discussed.

8.5.1 Structural Layer Comparisons

The structural layer of the reasoner provides comparisons between actors, based on the properties (modality and direction) of their capabilities. This results in the identification of channels of communication that indicate the ways in which actors may be able to interact with each other and the resulting routes that can be used for the transfer of information. Actors are represented in a graph-based structure, possessing capabilities and potentially being made up from a series of sub-actors with capabilities of their own. The structural layer is therefore composed of modules that perform pattern matching, based on the graph-based structure described in chapter 5. As represented in figure 8.6, modules are required to retrieve the appropriate data, provide comparisons and then expose the channels that have been identified.

As it is the lowest layer of the reasoner, the structural layer is the point of entry into the framework and the required capabilities can be identified in response to a request from outside of the system. For example, when using the prototype a user may request information about the accessibility of their interactions with a particular device or device component. If this is the case, the input would be two actors ($\mathbb{A}$ and $\mathbb{B}$) for which a comparison was required. Actor $\mathbb{B}$ (representing the device) will be considered the 'subject' from whose point of view the assessment will be carried out. Actor $\mathbb{A}$ (representing the user) will then be considered the 'partner' whose capabilities will be used to satisfy those of Actor $\mathbb{B}$.

While this may appear to be the wrong way round, it makes sense when considering the assessment as a comparison between their two sets of capabilities. In order for the user to be able to fully access the device, all of the device's capabilities must be satisfied by compatible capabilities belonging to the user. If the 'device' was a word processor within a personal computer (PC), its lack of any audio capabilities would be of little consequence; unless the user was unable to satisfy its (primary) visual and kinaesthetic requirements. This would make the set of capabilities displayed by the device ($\mathbb{B}$) a subset of the set of capabilities displayed by the user ($\mathbb{A}$).

Alternatively, a sequence of actors may be provided ($\mathbb{A}$, $\mathbb{I}$, $\mathbb{B}$) where $\mathbb{I}$ is an actor that has been placed *in-between* $\mathbb{A}$ and $\mathbb{B}$ in order to improve their accessibility. As an example a screen reader could be used to augment the capabilities of the word processor in the previous paragraph where the user was unable to provide the required visual capabilities. The sequence can be broken down into two comparisons ($\mathbb{I}$-$\mathbb{B}$ and then $\mathbb{A}$-$\mathbb{I}$) that can be tackled individually. If this is the case then processing functions are used to identify the capabilities exposed by $\mathbb{I}$ in the second comparison as dictated by the result of the first comparison. Using this approach any number of actors could theoretically be added to the chain.

8.5.1.1 Input/Retrieval

The structural layer allows the comparison of actors by matching between the properties of their capabilities and therefore requires series of capabilities to be identified, on which comparisons can be performed. As suggested above, given that it is the lowest layer, the focus for the structural layer will originate outside of the reasoner. If the request is for a comparison to be performed, two (or more) actors or capabilities will be specified and the data required for their comparison will be retrieved from the data-store. If the request is for a match to be performed, the data relating to the two (or more) sets of actors or capabilities will be required in order to allow multiple comparisons to be made. Two actors will be chosen, one from each set, that are most compatible.

The data retrieval module is responsible for providing the data required to allow comparisons to be made. While the graph-based nature of the framework should make it possible for queries to be composed and performed directly on the data-store the use of a specific data retrieval module provides the potential for pre-processing. Pre-processing involves the preparation of data in order to improve speed/efficiency when comparisons are performed.

Three uses of pre-processing will be briefly described, based on a request for the comparison of two actors. Firstly, although the framework has been designed with a semantic graph-based structure, the different comparison modules may use different formats (e.g. linked lists) to represent the data. Pre-processing can be used to pre-navigate the graph-based structure of the the data-store in order to present the data in the most efficient format for comparison. Secondly, as actors inherit the capabilities of their sub-actors, pre-processing can be used to perform inference which attributes the capabilities to them directly. Lastly, actors are compared at the structural layer through the modality and direction of their capabilities. As they may have multiple capabilities with the same modality and direction pre-processing can be used to select the most appropriate one by referencing higher layer information. As an example, a person with two hands is likely to present the same kinaesthetic capabilities in each, however many people have a dominant hand. By comparing their abilities at the data layer the dominant hand can be identified and presented within the structural layer comparison.

The Prototype is able to accept queries that are presented in the form of either two ($\mathbb{A}$–$\mathbb{B}$) or three actors ($\mathbb{A}$–$\mathbb{I}$–$\mathbb{B}$) for which a comparison is required. As described in the previous section profiles are stored in RDF-XML files and data is read into memory on initiation of the prototype system by existing parsers provided within SWI-Prolog. Once read in, the data is held in memory and a data retrieval module is used that uses pattern-matching to query the data and return its results into nested lists; the most prominently used data structure in prolog. A series of pattern-matching rules have been created, the first of which returns all of the capabilities of an actor (but not its sub-actors) along with their modalities and directions. A second rule has then been defined that lists the sub-actors of an actor and (recursively) reapplies the first rule to identify the capabilities of the sub-actors and attribute them to the top level actor. A similar rule is also available to recursively identify the constituent sub-capabilities of a higher-level capability. Finally, where comparisons are requested for more than two actors a fourth rule is available to identify data relating to processing functions. When a comparison is requested, the modules are run for each actor in the request as appropriate and the resulting data is fed into the structural layer comparison module.

8.5.1.2 Comparison

The comparison module of the structural layer of the reasoner is responsible for identifying the extent to which two (or more) actors' capabilities match. This is measured in terms of the modality and direction of the capabilities being compared and could result in either a binary [0,1] or a more descriptive [0–1]) result being provided. In the first instance, where two actors are being compared, the comparison will be performed from the point of view of a subject interacting with a partner, as described in section 8.5.1. Matching is performed to create channels of communication for each of the subject's capabilities by identifying capabilities from the partner with the same modality and opposing direction.

If all of the subject's capabilities are satisfied by the partner, their interaction is likely to be accessible and this is equivalent to the binary result of 'true' or '1'. Owing to the recursive nature of the framework, where a match is unavailable, inference can be used to perform the comparison in terms of lower level (constituent) capabilities. Where all of the sub-capabilities of a capability can be found, its existence can be inferred and the match confirmed. However if a match cannot be found for just one capability (or sub-capability), the corresponding binary result of 'false' or '0' would indicate a complete lack of accessibility. As this is clearly not the case a more descriptive result could prove more informative.

As an example, in a situation where some of a capability's sub-capabilities are available and some are missing, there is the potential for the lower level comparison to return a more descriptive decision (e.g. the percentage of successful sub-channels). This resulting value could then either be compared against a threshold value–creating a filter that allows channels to be inferred with a (quantifiable) degree of uncertainty—or returned directly.

Where a comparison is requested for a chain of more than two actors, processing functions can be used to identify the linkages that allow actors to translate and transform information between their different capabilities. For a three actor chain the comparison process is simply an extension of that used for comparing two actors (again described in section 8.5.1). Once channels have been identified between the first pair of actors, the processing functions of the partner are used to identify additional capabilities that can be added to those of the original subject for use in the second comparison.

The Prototype has a comparison module that consists of a series of pattern matching rules which accept a combination of actors and capabilities and assess the potential for channels of communication to be formed between them. The module is based on a pattern matching rule that takes two capabilities and checks whether they have the same modality and opposing directions. From this rules were developed to allow the systematic comparison of two sets of capabilities (with one being designated as the subject) in order to identify all of the potential channels of communications that exist between them. Where the subject had a capability that its partner could not satisfy, recursive querying of sub-capabilities was provided to identify if a match was subsequently possible. Next, by allowing capabilities to be identified based on their attribution to actors, rules identifying the ability of an actor to satisfy a particular capability, or the capabilities of another 'subject' actor were created. Finally, a separate rule was created to allow the matching of capabilities (and actors) through the use of processing functions that could themselves be attributed to one or more actors.

The comparison functions are all focused on the subject of their comparisons, producing a list of the capabilities that have been successfully satisfied by the subject. Although they are (at present) based on a binary comparison plug-in, this could be replaced by a more descriptive version in response to further research regarding the appropriate combination of descriptive values.

8.5.1.3 Output

The output module is responsible for taking the results of the comparison module and exposing them for use by the output module. In terms of functionality, its main role is one of ensuring compatibility and this can be achieved in two different ways. Firstly, the range of potential clients demonstrates a need to ensure that the content that is outputted is appropriate for the client it is serving. As an example, a visualiser may require information about the success of individual channels, whereas a single value indicating an 'accessibility score' may be sufficient to inform an accessibility provision agent whether its proposed solution is appropriate. Secondly, similar to the data retrieval module, the format that the content is provided in will also be important to ensure that the output itself is actually accessible to its clients.

The Prototype provides a visualisation output module, as will be described in section 8.6. In order to provide it with appropriate data, an output rule was created that took the list of capabilities from the comparison module and presented them in a format that could be used by the graphical interface programming functionality to highlight which were matched and which unmatched.

8.5.2 Data Layer Comparisons

The data layer of the reasoner provides comparisons between capabilities, based on the properties (units, data type, value) of the data items used to provide a measure of an actor's ability. This results in a measure being produced that describes the quality of interaction that can be sustained via individual channels. In order to quantify the ability of a channel, data is required that records an actor's abilities with regards to the capabilities between which the channel is formed. The data layer is therefore composed of modules that deal with ability measurements, in response to requests from the structural layer. As represented in figure 8.7

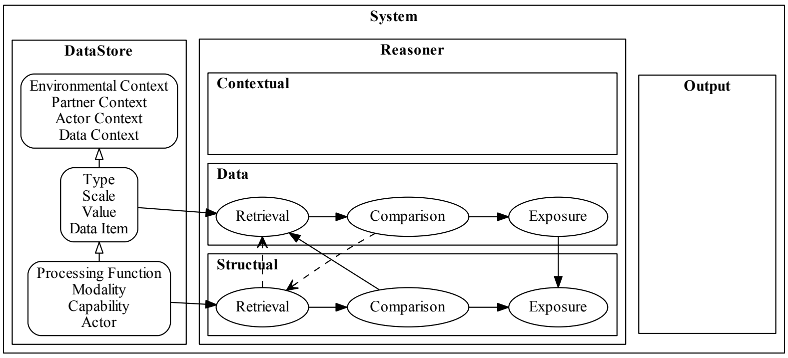


Figure 8.7: Implementation – Including the data layer reasoning module.

functionality is provided through the use of similar modules that: (1) retrieve ability data from the data-store, (2) provide comparison functions and (3) expose resulting channel bandwidths once they have been identified.

Two types of edges are identified in figure 8.7. Solid lines denote the flow of information relating to the construction of channels of communication. As described in the previous layer, channels are created through the comparison of data from separate actors, between whom interaction is to be modelled. Dashed lines denote the flow of information relating to the selection of data that will be used in the above comparisons. Again, as described in the previous section, multiple capabilities may be available to facilitate the creation of the same type of channel. Where this is the case, the data layer can be used to choose between the available capabilities by selecting the one with the greatest potential to result in an accessible interaction.

8.5.2.1 Input/Retrieval

The data layer provides comparisons between capabilities, by matching between the properties of their data items. It therefore requires pairs of data items to be identified, on which comparisons can then be made. Data items describe actors' abilities with regards to specific capabilities and as such, the identification of data items is dependent on the specification of capabilities (by a structural layer module) against which they can be focused.

Where a query is generated by a structural layer comparison module, this will be in the form of a pair of capabilities from different actors, for which a comparison is required. Where a query is generated by a structural layer retrieval module, this will be in the form of one or more sets of capabilities from which selection of the most appropriate one is required (for pre-processing purposes). In either case the data layer retrieval module will then be required to identify any data items that provide information about each of capabilities and pass them to the comparison module.

As capabilities may fluctuate over time, a single capability may have more than one data item attributed to it. If this is the case, the data items used for comparison can be selected as a result of pre-processing provided by the contextual layer – identifying the most appropriate data items for the situation being modelled.

The Prototype retrieves data related to the the data layer using a series of pattern matching rules in a similar way to its equivalent structural layer module. For each capability that is identified, a query is performed to retrieve any attributed data items and store them in the form of nested lists within the capabilities data structure for use by the comparison module. In order to demonstrate its potential to use different types of data, one overall data retrieval rule has been defined that is able to call on multiple data type specific rules.

As described in section 5.3.1.1, the construction of data items will be dependent on their data type. In the same way that different modules are required to interface with different data-stores, different 'plug-ins' will be required to retrieve and present data with different data types. At present rules exist to enable the use of binary and range data[11]. When a data item is first identified, its data type is identified and the appropriate rules for its manipulation are subsequently used.

At present binary and range data can both be retrieved and presented for comparison. As the overall data retrieval rule references the pool of data specific rules, the retrieval of new data types can be supported through the simple creation of an appropriate data specific rule.

8.5.2.2 Comparison

The comparison module of the data part of the reasoner is responsible for describing the bandwidth of a channel. This is measured in terms of the extent of the 'overlap' between the capabilities between which the channel is positioned. The diversity between different data types leads to the need for a variety of methods to be used for performing comparisons between data items. For example, range data can be compared in terms of the existence of an overlap between the ability ranges of the transmitting and receiving actors and binary data can be compared using a binary 'AND' function.

At their most basic, data layer comparisons are attempting to confirm whether the potential channels that were identified at the structural layer are actually able to support interaction. As described in section 5.3 it may be possible for

[11]Binary data is a value from the set [0,1] and range data comprises of two values; a maximum and a minimum.

two capabilities to have a common modality and opposing directions, but still be incompatible, due their lack of a shared ability range (e.g. a speaker producing sound that is too quiet for a user to hear). Similar to the output of the structural layer comparison module, the data layer module can produce results at different levels of measurement. However unlike the structural layer module, as well as being based on the combination of the results of multiple lower level channels, the data layer module can also provide results at different levels for a single channel. Although binary measurements can be used to acknowledge the existence or non-existence of an overlap in absolute terms, more descriptive measurements provide a greater level of detail about the magnitude of the overlap; which is in effect a measure of the bandwidth of the channel.

The potential range of different data types leads to a greater variety in the range of levels of measurements that can be used to provide a descriptive measure of their overlap. For example, if two ordinal values were compared the result of their comparison could describe whether they were equivalent or if one was greater than the other. However, if two interval or ratio level values were compared the result of their comparison could be much more descriptive (Field, 2009, p10).

The other use of the data layer comparison module is pre-processing, which involves selecting capabilities for use in structural layer comparisons through the differentiation that data layer data can provide. Rather than comparing data items in order to provide a measure of the accessibility of a channel of communication, data items are compared to identify the most suitable capability for use in such a comparison. As with functions used to measure the accessibility of channels, a number of techniques are available for pre-processing selection and although individual techniques will not be discussed there are two categories that the techniques may fall into. Capabilities could be selected based on either their conformance to a pre-determined criteria that describes the most appropriate ability, or through comparison based on a particular feature (e.g. highest/lowest value or widest range).

In longer chains of actors ($\mathbb{A}$–$\mathbb{I}_1$–. . .–$\mathbb{I}_x$–$\mathbb{B}$) picking the best combination of $\mathbb{A}$–$\mathbb{I}_1$ and then moving onto the best $\mathbb{I}_1$–$\mathbb{I}_2$ and so on may not give the best overall chain. However any other algorithm would be significantly more complicated.

The Prototype provides an overall comparison function that (like the retrieval module) uses a series of data specific plug-ins to allow the comparison of different data types to be performed. Comparison plug-ins have been created to match those identified in the previous section. At present the comparison module provides a binary output that identifies whether or not a match exists between two capabilities.

8.5.2.3 Output

Similar to the structural layer output module, the data layer output module takes the results of the comparison module, identifies the appropriate content and exposes it in a format that is compatible with the systems or processes that have requested the data. As the data layer provides information relating to the quality of the interaction that a channel can sustain, it is reliant on the structural layer output for the identification of channels against which its output can be focused. Alternatively, as seen in 8.7, the module can also be used to inform the output of the structural layer. In chapter 7 thresholds were used as a method of decision making and they can also be applied to data layer output in the same way. As an example, a minimum bandwidth limit could be set, which if not exceeded would result in a channel being declared as inaccessible. By turning the descriptive data layer channel measurement into a binary one, the data layer output is effectively reduced to a structural layer one that identifies the existence of channels.

The Prototype extends the structural layer output module by providing the additional data required to confirm or deny the existence of a channel. Similar to the structural layer output module, the data layer output module took the results of the comparison module and exposes them in a format that was compatible with the systems or processes that have requested the data.

8.5.3 Contextual Layer Comparisons

The contextual layer of the reasoner provides comparisons between data items, based on the contextual information that is stored about them. This results in the production of a measure of their contextual similarity, which can be used to describe the appropriateness of a data item for use in modelling a particular situation or comparison. The layer is therefore composed of modules that deal with the comparison of contextual data, which is (mostly) stored in the same graph-based structure as actors, capabilities and modalities. For contextual information that can be stored in the same graph-based structure as capability information, the functionality of contextual layer modules will be very similar to that developed for the previous two layers. As represented in figure 8.8 functionality is provided through the use of modules that retrieve contextual data from the data-store, provide comparison functions and then expose the resulting measure of applicability once it has been identified.

As described in chapters 5 and 7 there are four types of context and similar to the data layer, different plug-ins are required to deal with each.

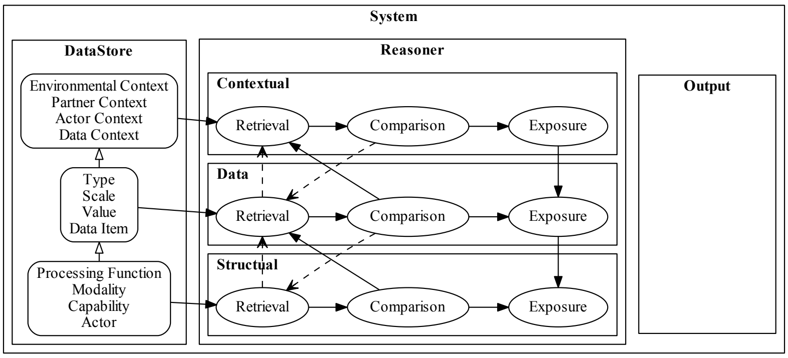


Figure 8.8: Implementation – Including the contextual layer reasoning module.

Data context provides information about how the data should be used (e.g. author, date and permissions).

Actor Context provides more holistic information about the state of an actor when the data was recorded.

Partner Context providing more holistic information about the state of the actor with whom interaction was taking place when the ability data was recorded.

Environmental Context provides information about the environment in which the data was recorded.

As described in chapter 7, data context requires the creation of custom functions. The reuse of lower layer structures by the other three types of context however results in the potential to reuse a number of the functions defined in the lower layers.

8.5.3.1 Input/Retrieval

The contextual layer provides comparisons between data items, by matching between the contextual information that is stored about them. It is the highest layer of the reasoner and the identification of contextual layer data is reliant on the specification of data items by the data layer. As seen in figure 8.8 the contextual layer retrieval module may receive requests from both the data layer retrieval and comparison modules, mirroring a similar relationship between the data and structural layers.

Where a query is received from the data layer comparison module, it will be presented in the form of two data items for which a contextual comparison is required. Where the query is generated by the data layer retrieval module, it will contain a list of data items for which contextual pre-processing is required. For either of these queries, the contextual retrieval module will be required to identify the contextual information that is attributed to each data item and present it to the contextual comparison module.

As described above, given that they share the same structure of actors, capabilities and data items, functionality relating to the actor, partner and environmental context can be reused from lower layer retrieval modules. However, as it uses a different structure, data context requires the creation of new functions, these can be created in the same way as the data layer retrieval module, which is able to work with different data types using a series of plug-ins.

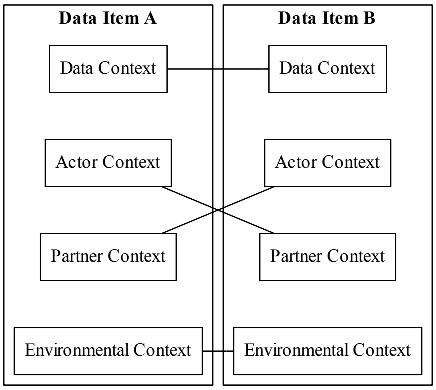
The Prototype provides functionality for retrieving both data and environmental context. For environmental context, lower layer data retrieval rules are re-used by a new pattern matching rule, allowing recursive retrieval of data relating to the contextual actor that has been attributed to the specified data item. In addition to retrieving contextual information relating to data items, the environmental plug-in is also able to retrieve data relating to the current context. A new ‘overall’ rule has then been written to retrieve data context that is able to use a number of contextual item specific rules in the same way that the data item rule was created in section 8.5.2.1. So far one contextual item specific rule, which retrieves date and time-to-live data, has been created.

Both of the rules retrieve data from the the dataset and append it in the form of a nested list inside the capability to which the data is attributed.

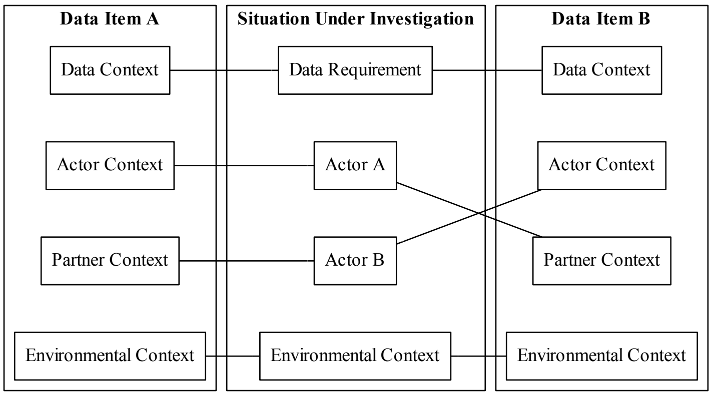
8.5.3.2 Comparison

The comparison module of the contextual part of the reasoner is responsible for identifying the extent to which data items are appropriate for use in modelling a given situation. This is measured in terms of their similarity, as described by the previous two layers. However, rather than comparing similar types of context between data items, each type of contextual information should be compared against the context it represents. As shown in figure 8.9a this means that the actor context of a contextual item $\mathbb{A}$ would be equivalent to the partner context of a contextual item $\mathbb{B}$.

The direct comparison of contextual information provides a measure of their similarity, which can be used to indication the appropriateness of the comparison of their related data items against each other. It does not however identify the appropriateness of the data items for modelling a particular situation. As a result, a comparison module could be developed to provide this form of comparison, as demonstrated by 8.9b.



(a) Directly comparing the context of two data items.



(b) Comparing two data items against a specific interaction.

Figure 8.9: Visual representations of the comparisons required to match between the contextual layer information of two data items.

The Prototype has plug-ins that are able to deal with both data and environmental contexts. The plug-in that provides data context comparisons uses the date on which a data item was created and its time-to-live to identify whether the data item is still valid. As described in chapter 7 the plug-in acts as a threshold, which if not passed, restricts the use of the data.

Two plug-ins are available to provide environmental context comparisons, representing each of the situations described in 8.9. Unlike the lower layer comparison modules, a descriptive value is provided in the range [0–1] that represents the probability the data is appropriate.

8.5.3.3 Outputs

The output module of the contextual layer of the reasoner exposes the results of the contextual comparison. It is similar in nature to its data layer counterpart in that it can either output its results directly, or feed them back to the layer below it. Where the output is exposed directly, this will be in the form of a confidence measure that is related to a specific data layer comparison. Where information is fed back to the layer below, this will be through the use of threshold functions that are able to further restrict the reported channel bandwidth to reflect the particular confidence that has been reported.

The Prototype extends the data layer output module by providing a confidence measure for the lower layer comparisons that have been performed. This is provided as a probability and stored inside the data layer output.

8.6 Output

The output module of a system takes the results provided by the reasoner, translates them, and exposes them to clients who can make use of them. As the last module of the reasoner to be described, it is effectively an inverse of the data storage module (providing rather than collecting data). As seen in figure 8.10 the output module may take output from any of the layers and the module may provide data to a range of output clients including:

Visualisers: Programs that aim to inform a user or developer by allowing them to explore the accessibility of interactions through the presentation of data.

Data-Store: Where the existence of a capability has been inferred from its constituent sub-capabilities this fact can be stored for future reference, reducing the need for recursive searching.

Accessibility Provision Agents: Potentially the widest range of outputs including agents that: identify that a assistance is required, suggest assistive technologies or provide/deliver assistance.

The Prototype has an output module that visualises the results of the reasoner. It is intended to represent a tool that could be used to inform users or developers about the accessibility of a particular device for a particular user. Two actors are described in terms of their sub-actors and the capabilities that they have. The potential for one to access the other is then identified through the identification of

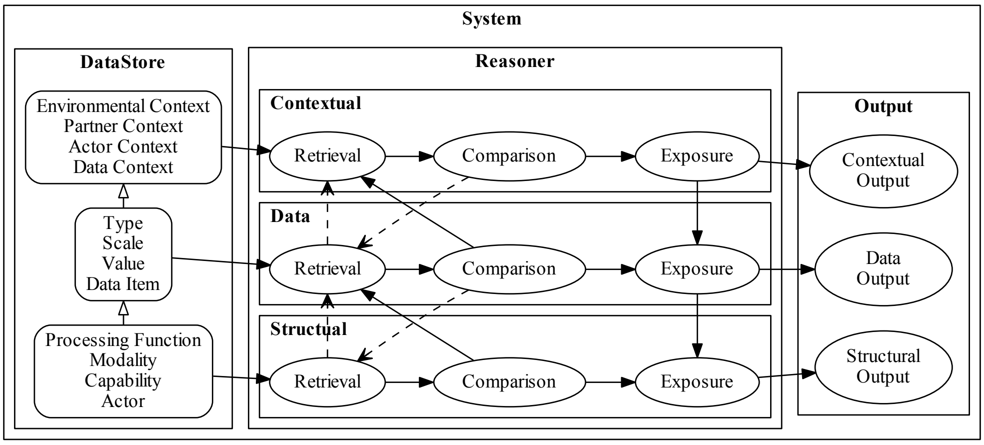


Figure 8.10: Implementation – Including the outputs module.

channels of communication between them. Accessibility is demonstrated through the use of colour to highlight the capabilities of each actor that the other is able to access.

8.7 Overview of the Prototype Using a Worked Example

While an evaluation of the utility and appropriateness of the implementation developed in this chapter will be provided in the next chapter, this section will provide a short worked example as a means of binding together the work presented in this chapter. The prototype was developed as a means of evaluating the ability of the framework to be used for manipulating data. The worked example will therefore show how the prototype provides a Graphical User Interface that allows the comparison of a person and a computer.

When loaded, the prototype provides a simple interface, shown in Figure 8.11. Working from Left to Right, Actor A ($\mathbb{A}$) is attempting to interact with Actor B ($\mathbb{B}$) and comparisons are therefore performed to investigate how many of $\mathbb{A}$'s capabilities $\mathbb{B}$ is able to sustain. In this example $\mathbb{A}$ is a person (User 1) with a series of sub-actors (in the upper list) and a series of capabilities[12] (in the lower list). The actors and capabilities are displayed based on the file that they are stored in and their name (e.g. [file]:[actor/capability]). The top level actors (user1 and computer) do not have any capabilities attributed to them directly; instead,

[12] The convention used for naming capabilities with a prefix ('i_' or 'o_') identifies whether they are used for inputting or outputting.

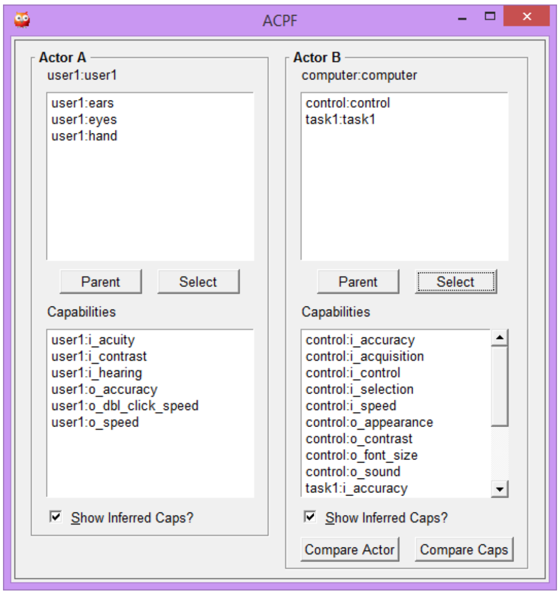


Figure 8.11: Worked Example – The graphical user interface of the prototype.

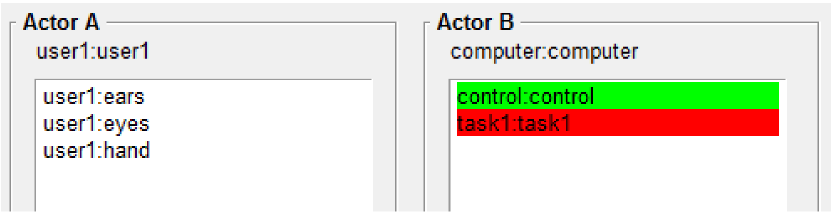


Figure 8.12: Worked Example – Using the prototype to compare Actor $\mathbb{A}$ against the sub-actors of Actor $\mathbb{B}$ to predict their accessibility.

their capabilities are inferred from those attributed to their sub-actors (using the checkbox at the bottom of the window).

Data relating to each of the actors is stored in RDF files and through the data structures used by the prototype the data can be used within comparisons between the actors to assess their ability to interact. In Figure 8.12 the button marked "Compare Actor" has been pressed and as a result, $\mathbb{A}$ has been compared against each of the sub actors of $\mathbb{B}$. The green highlighting over "control:control" implies that $\mathbb{A}$ is able to interact with[13] all of the physical devices that make up the computer. The problem, as denoted by the red highlighting over "task1:task1", is with the software-based task that the computer is being used to complete. By looking at the sub-actors within the task (using the "Select" button) it is apparent that the problem is with the formatting of the labels.

As well as comparing $\mathbb{A}$ against the actor $\mathbb{B}$, $\mathbb{A}$ can be compared against $\mathbb{B}$'s capabilities (using the "Compare Caps" button). Figure 8.13 shows the result of this comparison, highlighting three Capabilities that the user is unable to perceive. While "contrast" is the specific capability that is inaccessible to the user this

[13] A simple binary comparison, looking for overlapping ranges.

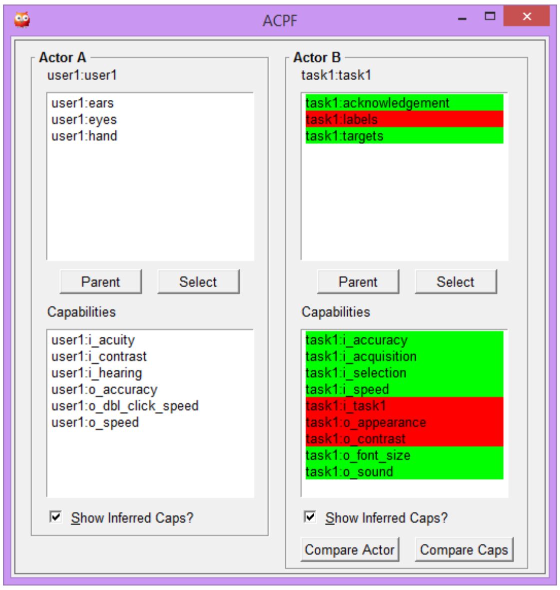


Figure 8.13: Worked Example – Using the prototype to compare Actor 𝔸's ability to interact with each of Actor 𝔹's capabilities.

results in both the appearance of the labels and the overall task being inaccessible. The information that the prototype provides can be used to identify the types of adaptations that are required for two actors to be able to interact.

8.8 Summary

This section will now summarise the achievements of this chapter and describes how the implementation it provides allows the for further evaluation of the framework to be performed in the next chapter. This chapter has described how the theoretical structure of the framework provided in chapters 5 and 7 can be implemented to provide a system that can be used for the assessment of accessibility between people and pieces of technology.

8.8.1 Contents

A number of suitable technologies and techniques were identified that could be used to produce a system based on the structure and particles defined within the framework. The semantic web was identified as a means of practically representing and transporting data via existing web-based standards. Programming languages that provide logical inference were identified as a means of achieving the graph-based pattern matching required for comparing data from different profiles. Finally, probabilistic modelling techniques were identified as providing a means of both representing and reasoning with the uncertainty that is inherent in systems that use data collected from a variety of sources.

The design and practical implementation of a prototypical system was then described. By using a modular design, both the overall system and the reasoner itself have been provided with the ability to be developed in stages, with additional functionality built on top of existing data and modules. While a basic system may be produced that includes modules representing only the lower layers of the framework, higher layers can then be added as required through the inclusion of appropriate higher layer modules and plug-ins. As well as building modules up in layers, different alternatives can also be produced to provide functionality that is appropriate to a given use. As an example, both data item (retrieval and comparison) and processing function (retrieval and translation) plug-ins were written as separate rules which were used on demand, as dictated by the data that their higher level rules encountered. Alternatively, in future developments, reasoning could use different levels of measurement to produce either binary or more descriptive matches.

8.8.2 Impact

This chapter provides two important outputs to the thesis as a whole, improving the quality of the contribution it makes to the wider research domain. Firstly, the chapter demonstrates the ability for a practical implementation of the framework to be produced in the form of a prototypical system that allows the accessibility of interactions between different actors to be performed. The system uses both the same structure as the framework itself (as developed in previous chapters) and the approach of separating concerns. Secondly, the system itself can be used as a reference implementation to aid the evaluation of the usefulness of the approach developed in chapter 4 in providing a means of performing accessibility assessments.

The approach itself will be further evaluated in the following chapter through an evaluation of the prototype.

Chapter 9

Evaluation

With details of the motivation, design and implementation of the framework provided in previous chapters, this chapter now addresses the need for an evaluation to be performed. In chapter 2 a series of aims were identified that have been used to guide the research detailed within this thesis as a whole. It was also described how these aims would result in the thesis making a contribution to the wider field of accessibility modelling.

In order to provide a thorough evaluation, this chapter evaluates both the ability of the practical implementation to deliver the contributions, and the ability of the framework to deliver against the research aims. Section 9.1 recaps the aims and functional contributions described in chapter 2 and describes the dual approach that has therefore been taken. As a result of this approach, the rest of the chapter's content is then organised into three parts:

Section 9.2 describes the use of a data collection study to provide real user data and the storage of that data within an ontology. In section 9.3 the data is used to drive a practical assessment of the framework and demonstrate its ability to deliver against a series of functional contributions. Section 9.4 reviews each of the aims in turn by assessing the validity of the theoretical approach taken to deliver them and the practical evidence provided by the data collection study. Finally section 9.5 provides a discussion of the achievements and limitations of the chapter by summarising its contents and describing its contribution to the impact of the research provided within the thesis as a whole.

9.1 Approach

This section will describe the approach that has been taken within this chapter, in order to provide an evaluation of the research produced within this thesis as a whole. The literature review in chapter 2 resulted in the production of a number

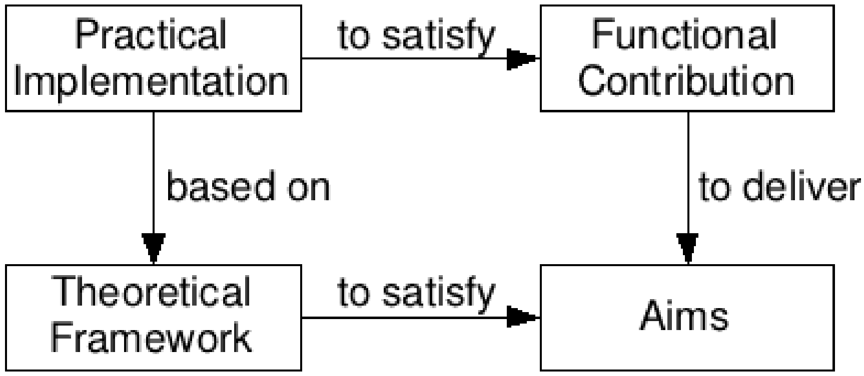


Figure 9.1: The use of aims and functional contributions in evaluating the approach.

of aims that describe the complex nature of accessibility modelling:

Aim 1: Functional Assessment of Accessibility

Aim 2: Variability Between and Within Individuals

Aim 3: Variety Within and Interaction Between Accessibility Barriers

Aim 4: Variety Within and Interaction Between Accessibility Solutions

Aim 5: Variety of Agents Producing and Using Data

Aim 6: Variability of Data Quality

Although all of the aims had previously been addressed in different pieces of existing research, no single solution was found that was able to address them all simultaneously (as described in chapter 4). While identifying an approach for providing a unified solution, it became clear that a static modelling format would not provide the flexibility that was required to incorporate all of the aims simultaneously. This resulted in a theoretical framework being designed in chapter 5 to satisfy the aims. The framework provides a series of particles, connected together by a series of standard relationships, which can be used to build models of people and technology that facilitate accessibility assessments. Chapter 6 then highlighted the difficulties faced in collecting data to drive a system based on the framework. As data may be collected with varying degrees of validity, a mechanism was required that was able to filter data based on it characteristics and chapter 7 discussed how the framework can integrate this requirement as well.

As seen in figure 9.1, the aims may be delivered through the creation of a system that provides a number of functional contributions:

- Representation and storage of collected data in a form that encourages compatibility and transportability between different systems.

- Support for comparison of users and technology (with reference to existing support) at varying levels of fidelity, in order to identify potential interaction in a given context.
- Mediation between data from different sources and provision of a confidence measure to describe the accuracy of any predictions made.
- Identification of mismatches in terms of the elements responsible and the nature of the problem; resulting in the provision of a description of the assistance required.
- Enabling speculative augmentation to depict different technology configurations or the use of different forms of assistance.

With a theoretical design for the framework provided in the preceding chapters, chapter 8 then provided details of a prototypical system that was developed based on the framework. The system was designed to demonstrate both the feasibility of implementing a system based on the framework and the ability of a system based on the framework to provide the functional contributions described above.

The evaluation performed in this chapter pulls together all of the previous threads through the use of a data collection study that has been performed in order to provide real data for use within a practical implementation. In section 9.2 the design of the study is described and the data is used to provide a practical evaluation of the ability of the framework to deliver the functional contributions. Each of the contributions is discussed in turn and the study is used to provide data to support them. Following this practical assessment, section 9.4 will evaluate the research conducted throughout the thesis as a whole, in terms of the aims identified in chapter 2. Each of the aims will be described in terms of: (1) how the framework addresses it, (2) the theoretical validity of the approach and (3) supporting evidence from the evaluation study.

9.2 The Data Collection Study

The practical evaluation was designed around a scenario in which a person uses a computer to complete a task (based on the one developed in chapter 6) and a high-level depiction of the scenario can be seen in figure 9.2[1]. The use of this scenario provides a realistic (yet constrained) example of a situation in which the framework would be beneficial and allows the demonstration of the functional contributions.

[1]The figure uses the same notation as developed in chapter 5. A computer is interacting with a user. The computer contains a task, which has some adaptations that alter its presentation.

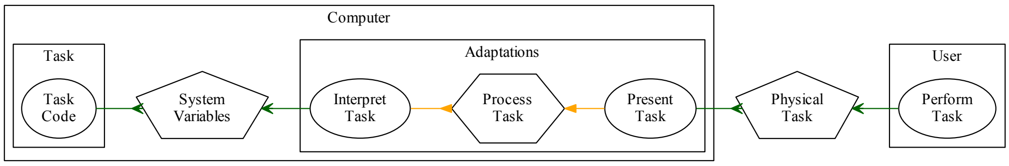


Figure 9.2: A high-level representation of the scenario (using the notation developed in chapter 5).

In order to assess the ability of the practical implementation to predict a person's ability to complete the evaluation task, data was collected via a number of 'mini-games'. The mini-games were lower-level (constituent) tasks that produced data, which was combined and used to predict participants' abilities to complete a higher-level evaluation task. In section 9.2.1 the data collection study will be introduced and the suitability of the framework for use in the collection of data will be described. In section 9.3 the use of the framework to perform comparisons with user data will be discussed and the use of the framework as a means of estimating ability will be evaluated.

9.2.1 Data Collection Methodology

The evaluation task scenario was used as the basis for a data collection study, further details of which are are provided in appendix C. The study collected data from real people, in order to provide additional evidence demonstrating the ability of a system based on the framework to provide the functional contributions. Nine participants were recruited for the study, which provided the data necessary to demonstrate the ability of the prototypical system to accurately determine the separate accessibility needs of a number of different people, in relation to a single piece of adaptable technology.

The study consisted of a series of mini-games that were used to collect data relating to four different capabilities: 1) sight, 2) hearing, 3) hand dexterity, and 4) finger dexterity. Participants then attempted an evaluation task (that had been subject to a number of adaptations), in order to assess the potential for predictions to be made based on the data provided by the mini-games.

There is a long history of relatively small numbers of participants being used in usability trials (Macefield, 2009). While the data is not being collected as part of a usability trial, it is providing a representative sample of users who may be involved in one. The ability of the implementation to accurately predict the participants' abilities to interact with the evaluation task will demonstrate a level of accuracy that is acceptable for common industry testing. For example, the number of participants recruited was greater than in a study testing a user simulation tool designed to evaluate interface layouts for impaired users Biswas & Robinson (2013).

9.2.2 Data Gathering Mini-Games

In order to facilitate the collection of user data, three mini-games were developed. They were based on existing standardised assessments, which had themselves been designed and validated to measure different capabilities (further details provided in appendix C). Three types of mini-games were developed (audio, visual and motor) a number of mini-games were developed to assess visual capabilities and the single motor mini-game assessed two separate capabilities. As the study used a within-subjects design there was a potential for a learning effect to be observed, however, a standard order was used to avoid the effects of motor fatigue as the study progressed. The potential for learning effects to bias the results was then minimised through the provision of participants with training time before each mini-game.

The mini-games all exposed different abilities, as seen in figure 9.3a and similarly, the participants exposed a series of abilities through their sub-actors, as seen in figure 9.3b. When 'playing' the games a participant's abilities could be identified, based on the mini-games that they were able to interact with. Each of the mini-games measured a discrete capability:

9.2.2.1 The Audio Mini-Game

The Audio mini-game measured a participant's ability to receive information over an audio channel, through the use of sounds that were transmitted out of the computer's speakers. The audio mini-game was constituted from volume and frequency elements and took the form of a number of tones with different frequencies that were played at gradually decreasing volumes until the participant could no longer hear them. Tones were played sequentially at random intervals and participants had to press the enter key within one second of the tone being played. Participants were instructed to press the key as soon as they could *comfortably* hear a tone. The use of a relatively short time period and random intervals reduced the potential for participants to either guess when the tone would play or retrospectively decide that they had heard a tone. This decision increased the likelihood that an accepted key-press indicated that the participant had actually heard the tone (rather than learned or guessed when to press the button), which therefore increased the validity of the assessment.

The mini-game resulted in a low-resolution audio-gram being produced, which provided a representative picture of a participant's hearing, similar to that produced by an audiologist conducting an audiometric hearing test.

9.2.2.2 The Visual Mini-Game

The Visual mini-game measured a participant's ability to receive information over a visual channel, through the use of light transmitted out of the computer's monitor. Three different games were used based on: (1) the Snellen visual acuity test (BS 4274-1:2003, 2003) (2) a preference based measure used to assess visual acuity in Atkinson (2012) and (3) a visual search task based on the Landolt C Test (EN ISO 8596:2009, 2009).

The Snellen-based visual acuity test used sequences of six letters which the participant had to identify; the letters got gradually smaller and/or lighter and the participant was given a score according to the smallest and lowest contrast letter they could identify. The preference-based game was similar, but the participant was asked to identify the smallest/lowest contrast that they could "comfortably" read. Finally, the visual search task randomly presented either a capital C or its reverse (as displayed in figure C.1) on the screen; the participant had to acquire and accurately recognise the shapes by pressing the arrow button corresponding to the side on which the hole appeared.

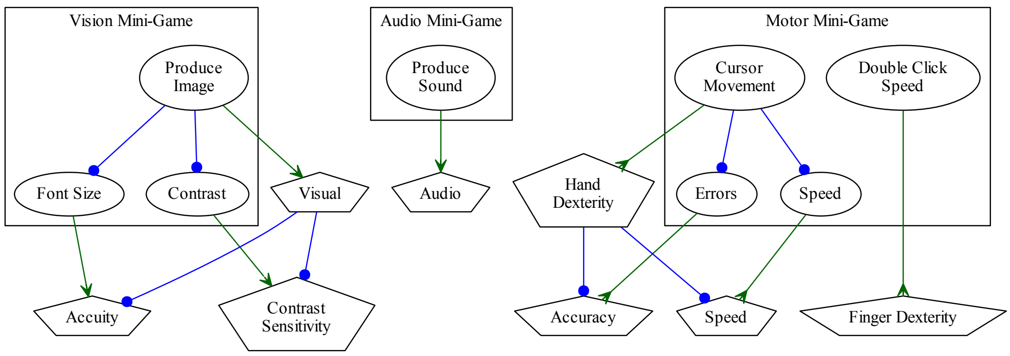
The games provided three different scores which all describe a participant's visual acuity[2], a single representative score, similar to that provided by an optometrist conducting an ophthalmic examination.

9.2.2.3 The Motor Mini-Game

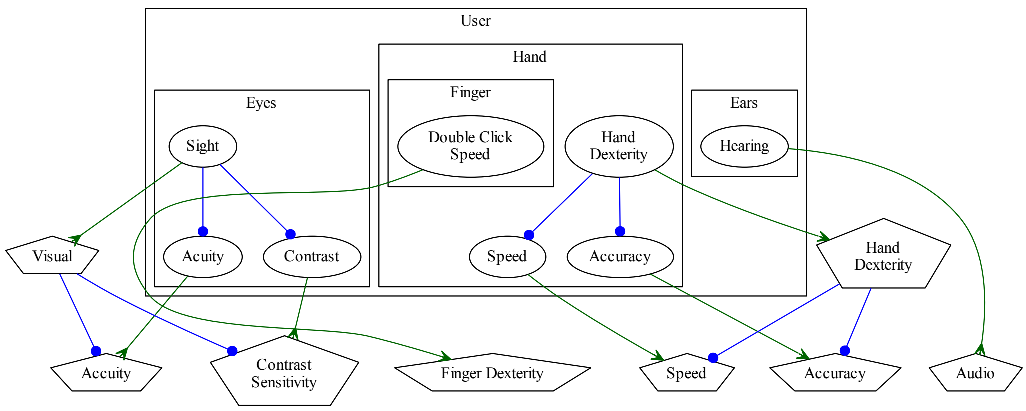
The Motor mini-game measured a participant's ability to transmit information via a motor channel, through the use of the physical force on the computer's mouse. A joint movement and double-click test was created based on the multi-directional point-select task defined in ISO 9241-9 (2000) as used in MacKenzie *et al.* (2001). Six circular targets were arranged in a circular layout; both the diameter of the targets and the diameter of the layout circle within which the targets were arranged were adaptable. Participants had to move the mouse in a set pattern and double click on the targets in the correct order.

By assessing the time taken to move between the targets and double click on them, measures of a participant's arm/wrist and finger dexterity were provided, similar to those collected during the assessment of Parkinson's disease (Cunningham *et al.*, 2012).

[2] the 'clearness' of their vision



(a) The mini-games.



(b) A user.

Figure 9.3: A graphical representation of the profiles created for the the evaluation study in terms of actors, capabilities and modalities.

9.2.3 Higher-Level Evaluation Tasks

In order to evaluate the ability of the framework to perform comparisons, a multi-modal computer-based task was then created (again further details are provided in appendix C). It required the combined use of the capabilities that had been previously assessed by the mini-games and could be adapted in the same ways. An interaction model is provided in figures 9.4 and 9.5 showing how a user would require all of their abilities to interact with the evaluation task.

The task involved the use of a mouse to double-click on 12 labelled targets, in the right order, with an audio acknowledgement providing feedback in order to inform the participant that they could move onto the next target. For each target the participant was required to (1) perform a visual search to identify the appropriate target, (2) acquire the target using a mouse to move an on-screen cursor, (3) double-click on it to select it and (4) hear the audio feedback to acknowledge successful completion and indicate they could move on to the next test.

Four versions of the task were created, with different configurations of adaptations used to 'vary their abilities' and cause different accessibility barriers. Table 9.1 provides a summary of the tasks, including their focus and capabilities (at both the higher task level and underlying capability level). Although not explicitly shown, the tasks' capabilities could be used to identify their requirement on the participants.

The first configuration was used for both training and to provide a control measurement. It was therefore designed to be fully and easily accessible. The subsequent three tasks were then adapted to reduce their accessibility in either one or more modalities (which were not disclosed to the participants), in order to either make them harder to complete or induce failure. After the control task, the order of subsequent tasks was randomised in order to reduce the potential for a learning bias to occur. As identified in chapter 6, repeated negative experiences can have a detrimental effect on a person's self-efficacy. Randomising the order in which the tasks were presented also allowed the intensity of the barriers to be varied between tasks in order to avoid demoralising the user.

9.2.4 Assuring the Ontology's Integrity and Representativeness

In preparation for the data collection study, the integrity and representativeness of the ontology had to be validated. An ontology's representative validity describes its ability to provide an accurate portrayal of the subject it is attempting to model. Representativeness is therefore a subjective property and requires a qualitative assessment to support its assurance. An ontology's integrity relates to

Table 9.1: Summary of Evaluation Tasks

			Task Abilities			
Task #	Focus	Overall	Visual	Audio	Hand Dexterity	Finger Dexterity
0	Control	↑	↑	↑	↑	↑
1	Audio	↓	↑	↓	–	↑
2	DblClick	–	↑	↑	↑	–
3	Combined	–	–	↑	–	↑

↑ High task ability (Requires low participant ability).
– Median task ability (Requires median participant ability).
↓ Low task ability (Requires high participant ability).

its internal consistency, ensuring that it obeys the rules specified by its schema (the framework) and contains accurate values. Owing to the structural nature of the framework and the construction of the prototypical system, the integrity assurance process is objective and supports automation.

9.2.4.1 Representativeness

In total six different ontologies, representing five top-level actors, were created to support the evaluation of the framework. Three ontologies were created for the three mini-games, these were used to identify the data that would be stored in the fourth ontology, representing the user. The user ontology was then compared against a fifth ontology, representing the evaluation task, within the prototypical system. While they represented different actors and contained different instances of similar capabilities, all of the ontologies referred to the same set of modalities and therefore shared references to a sixth ontology, holding their details.

The scenario upon which the evaluation was based was developed in consultation with two academics that had experience in the field of accessibility modelling (who were consulted separately). To assure their representativeness the ontologies were created independently (by the author) and subsequently returned to the academics for validation. Additionally, the ontologies were all developed following the identification of the scenario described above, which allowed them to be designed to capture only data that would be relevant to the intended comparisons. The controlled nature of the study meant that it was possible to collect only that data from the mini-games that would be required in predicting ability at the evaluation task. In reality however, ontologies would be created firstly to represent actors and their tasks, meaning that some extraneous data would be likely be collected.

9.2.4.2 Integrity

With an assurance process used to ensure the representativeness of the ontologies, they were judged as appropriate to be practically implemented. However, before they could be used in the prototype another assurance process was required, which focused on the integrity of their use within the evaluation. Rather than a subjective assessment, an objective process was used that mirrored those seen in a commercial system. Firstly, the structure of the framework allowed the ontologies to be constructed in a systematic way, reducing the number of errors that could be caused through human involvement in their crafting. Secondly, the automated generation of parts of the ontologies by the mini-games removed the human element entirely. Finally, the prototype itself provided an independent means of semi-automatically assessing structural integrity.

The hierarchical structure of the framework lends itself to the systematic construction of ontologies that are based upon it. Each of the ontologies was created with the same "top-down" strategy. The highest level actor was created along with its required relationships (e.g. [Actor] is a Actor). Each of the actor's capabilities were then created in the same way: with their required relationships defined first, followed by a relationship to the actor and then the creation of their data item. Data items (and their related contextual information) were also created in the same way again. Once the top-level actor had been created its sub-actors were created following the same format. Through the breadth-first creation of actors the ontology was populated a layer at a time and integrity could be assured before moving onto the next level.

While a single ontology could be created for each of the mini-games and the evaluation task, each of the users required their own version of the user ontology. For this reason a generic user ontology was created that could be populated with data from each of the users separately; the use of a template in this way reduced the risk of human error. In a similar way the visual mini-game was designed to output its data in RDF, which was suitable for direct integration with the rest of the user's ontology, minimising the potential for human error.

In addition to the approach used to ensure that ontologies were well formed before they were used within the prototype, the prototype itself provided a level of assurance as to the integrity of the ontologies loaded into it. Actors and their capabilities were only displayed in the prototype if they were well-formed and followed the design developed in chapters 5 and 7. This provided the ability to validate each ontology as it was used, by visually checking that each actor contained its expected capabilities.

9.2.5 Summary

This section has demonstrated the use of a data collection study to provide data from potential users of a system that has been based on the framework.

9.3 Evaluation Against the Functional Contributions

While the previous chapter demonstrated the feasibility of building a system based on the framework, it provided no evaluation of the effectiveness of that system's ability to assess accessibility. This section therefore presents an evaluation that has been designed to assess the ability of the practical implementation to deliver the functional contributions described in section 9.1 and thereby meet the objectives developed in chapter 2. This evaluation is performed using the data collected from real users during the data collection study introduced in the previous section.

The data collection study provided real data that would be collected and processed by a system that was based on the framework. The study also provided a number of observations that highlight areas where a wider system, based on the framework, would be beneficial. This section draws on both the data and observations in order to provide an understanding of how the prototypical system would deliver the functional contributions identified in chapter 2 and the associated wider benefits this could provide.

9.3.1 Representation and Storage

The data collection study was designed around a scenario that involved a person interacting with a computer-based task. The subsequent evaluation activities (e.g mini-games for predicting accessibility and an evaluation task) revolved around a number of associated capabilities. The accessibility of the task was dependent on the match between its capabilities and those of the individual attempting to use it. Figure 9.4 provides a simplified box and graph-based depiction of an interaction model constructed to allow the prototypical system to assess the accessibility of the scenario. In providing reasoning based on this interaction model, the prototypical system demonstrates its ability to represent both of the actors involved and the accessibility of their interactions.

Based on the approach taken by the framework, the mini-games were described as having a series of capabilities – depicted in figure 9.3a. Each of the mini-games transmitted or received information via different modalities and at varying levels of ability, controlled through the use of in-built adaptations. Participants were

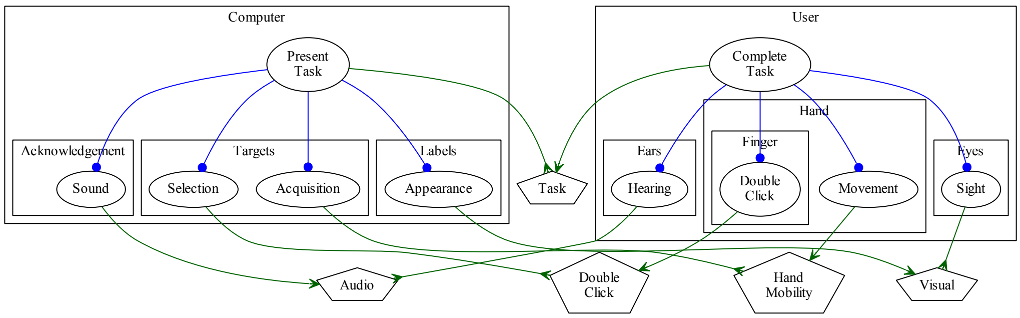


Figure 9.4: A 'box and graph'-based interaction model depicting a user and the evaluation task.

then described in terms of their own (separate) capabilities, which received and transmitted information via the same modalities as those of the mini-games (figure 9.3b). By observing a user's interactions with the mini-games, data was produced that described the user's own capabilities. Successful interactions were then used to infer participants' abilities, allowing data to be appended to their profile.

The data collected by the mini-games was stored in such a way as to represent its provision, both from different systems and at different hierarchical levels of abstraction. Although only one of the mini-games presented its output directly in the RDF format required by the prototypical system, all of the results could be coded into the RDF-XML format required by the reasoner. Additionally, by storing each of the outputs in a separate file, the provision of data by different systems can be represented.

9.3.1.1 Differentiating Between Capabilities

As well as being able to store data in a distributed nature, the data itself was also collected at different hierarchical levels of abstraction between mini-games. Figure 9.5 shows a more complete interaction model. As an example, the motor mini-game's ability to receive information via the hand-dexterity modality (used to identify an ability to move a mouse around) was measured in terms of two underlying capabilities: speed and accuracy. This allowed participants to be described both in terms of an 'index of performance' (or throughput) describing the nature of their higher-level speed-accuracy trade-off (MacKenzie & Isokoski, 2008), as well as the underlying capabilities themselves. The same mini-game also had the ability to receive finger dexterity information, however this was collected in terms of a single capability: double-click speed.

In a similar way, the visual capability 'sight' was measured in terms of its underlying capabilities: 'acuity' and 'contrast sensitivity'. By isolating the two

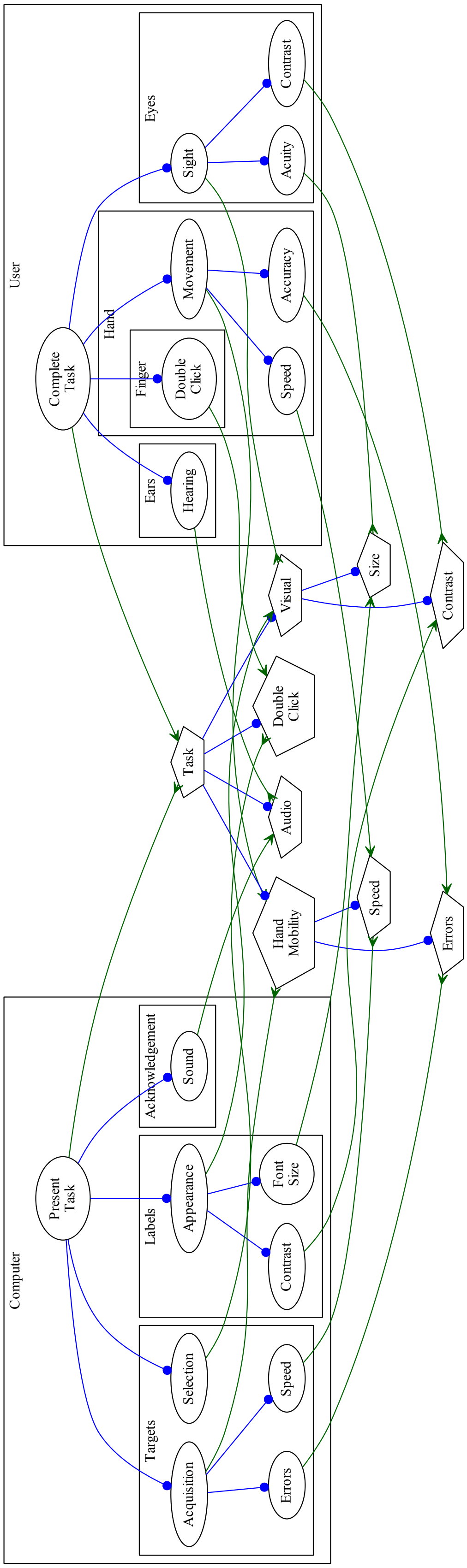


Figure 9.5: A model of the interaction between a user and the evaluation task including an additional hierarchical layer.

lower-level capabilities it was possible for them to be used separately, as well as being combined to indicate participants' higher-level capability.

Acuity is an ability to distinguish between objects based on their detail and shape.

Contrast Sensitivity is an ability to distinguish between objects based on their difference in colour.

Both (sight) acuity and contrast sensitivity can then be combined to provide a higher-level measure.

As described in section 9.2.2, recognised visual tests were adapted for use in measuring each of the above capabilities. While the two lower-level capabilities could be scored using their objective measurements (e.g. font size or hex colour code) the higher-level ability required a more abstract 'ability-based' scale to be created. The data collection study acuity was measured using a font size in pixels and contrast was measured using the hex colour code that the text was displayed in. The higher-level sight ability scores were calculated by using the contrast and acuity values in table 9.2. The two lower-level capabilities could also be scored against a similar scale and the ability scores used in comparisons performed in the following two sections were worked out using the values in the table.

Table 9.2: Definition of Ability-Based Scores

Score	1	2	3	4	5	6	7
Acuity	8px	12px	16px	24px	36px	48px	72px
Contrast	EFEFEF	EEEEEE	CCCCCC	AAAAAA	888888	666666	000000

It could be argued that the visual capabilities appear to be so closely related that there would be little benefit in measuring them separately. If this were the case, lower-level capabilities would be so closely related to their higher-level abilities that ability scores would be indistinguishable. Although this would support the hierarchical nature of capabilities, it would reduce the requirement for inference as two higher-level capabilities sharing a lower-level capability would also be highly correlated. Insufficient data was collected to allow a correlation- or regression-based analysis to be performed to confirm the relationships between the capabilities; their similarity, however, results in the relationship being implicitly evident. The ability-based scores that were collected can however be used to perform a comparison between the different capabilities, showing their independence.

Paired t-tests (displayed in table 9.3) suggest that participants ability scores for the three capabilities had statistically different means. The difference between the higher-level sight capability ($\bar{x} = 3.9$, $s.d. = 1.3$) and its two underlying capabilities supports the idea that multiple minor barriers can combine to create a significantly greater barrier (in terms of the higher-level capability they infer). The difference between the acuity ($\bar{x} = 6.4$, $s.d. = 0.9$) and contrast sensitivity ($\bar{x} = 5.4$, $s.d. = 3.9$) scores further demonstrates their independence and the validity of treating them as two underlying, but separate capabilities.

Table 9.3: Paired T-Test Results Between Visual Capabilities

Capabilities		t	df	Sig. (2-tailed)
Acuity	Contrast	-5.8	26	.0
Contrast	Both	10.7	26	.0
Acuity	Both	6.62	26	.0

9.3.1.2 Measurement of Capabilities at Different Hierarchical Levels

Given that all aspects of the study took place on the same hardware, each of the mini-games was able to record participants' abilities in terms of measures that were tied to the testing equipment. However the technology-focused preferences collected by the mini-games were converted to more generic ability measures as a means of providing better transportability between devices. The data collection study provided two practical examples of methods that could be used for converting lower-level data for use at higher hierarchical levels.

The audio mini-game measured hearing by gradually reducing the volume of a tone being played and recording the level at which participants were unable to hear it. This resulted in a low-resolution ability-based audiogram being produced as shown in figure 9.6. Through reducing volume by a factor of three and recording ability in terms of the number of reductions that were made, ability levels followed a similar logarithmic scale to that used for the measurement of sound in dB[3]. The measure resulted in a reversal of the traditional volume scale by attributing higher levels (representing better hearing) to lower values (the ability to hear lower volumes). This is however appropriate given that way that the reduced abilities of one actor within an interaction can be offset through the increased abilities of another. By comparing a participants' individual values against a series of standardised ability curves, the most appropriate match could then be used to provide a single ability value, characterising a user's hearing across the

[3] http://www.britannica.com/EBchecked/topic/155074/decibel

full spectrum of frequencies (similar to the use of Fletcher-Munson curves or other equal-loudness curves e.g. ISO 226:2003 (2003)).

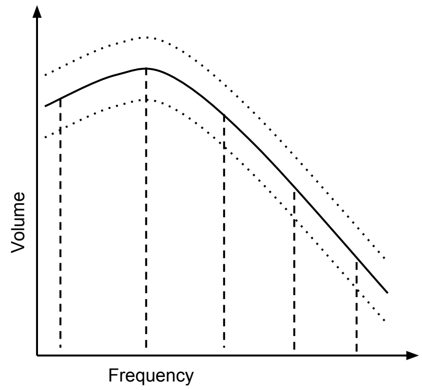


Figure 9.6: An ability-based audiogram.

While a similar approach could have been used to identify a user's visual abilities, the more consistent negative relationship that can be observed between their sub-capabilities (shown in figure 9.7) allows a different approach to be taken. Rather than measuring capability in the same way across the whole spectrum of abilities and then fitting the curve, three 'assessments axes' were used. The first and second are represented by the horizontal and vertical dashed lines and equate in the data collection study to the measurements used to identify the underlying sub-capabilities (acuity and contrast). The third 'axis' is then represented by the diagonal dashed line that varies both sub-capabilities at the same time and is used to provide a measure of the general higher-level capability, by providing an indication of the location of the curve.

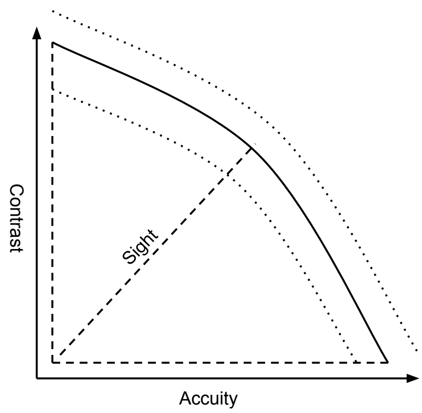


Figure 9.7: Relationship Between Contrast and Acuity Capabilities

9.3.2 Supporting Comparison

The study collected data through the use of mini-games and higher-level tasks that both relied on the same capabilities. Using the experience gained from the data collection study, this section will now demonstrate the ability of the framework to support comparisons between users and pieces of technology (at increasingly lower levels of detail, as appropriate), in order to predict accessibility.

While the ability of a user to complete higher-level tasks was collected, the regurgitation of task data as a means of predicting a user's ability to perform a task is not so much a prediction as an evaluation. It does however serve as a starting point to describe the comparison process. Before comparisons can be attempted, the availability of suitable data must be confirmed. This is achieved by querying the user's profile for capabilities that match those from profile of the technology that they are being compared against. An initial structural layer query allows the capabilities of the user to be identified and if one cannot be found that provides a suitable match, the technology's profile can be queried to identify a series of sub-capabilities that can be used instead. These sub-capabilities can then be queried against the user, with successively lower hierarchical levels of data being queried until a suitable match is found.

The study collected data at three hierarchical levels; the highest being that of the task itself; followed by a series of underlying capabilities; some of which were also measured in terms of their underlying capabilities. As data from all of the participants was collected through the use of the same tasks and mini-games, all of the participants exhibited some form of ability with regards to all of the capabilities (and sub-capabilities) that were measured. The removal of the higher-level task-related capabilities therefore resulted in the ability to perform successful structural layer matches through querying of recursively lower-level capabilities.

If a structural layer match is identified, the data layer is then be queried to identify if any data items exist, in order to allow a data layer comparison to be made. Once again, given that all participants followed the same procedure, the existence of appropriate data items could be guaranteed. While section 9.3.3 will describe how the system is able to handle a situation where inappropriate data items are available, this section will now discuss the ability of the system to perform matches following the removal of task related data.

As described in section 9.2.3, four tasks were created, based on the same capabilities that were measured by the mini-games. While users attempted the mini-games separately (requiring the use of the capabilities in isolation), the tasks required the use of all of the capabilities together. As such, whereas it was possible for a user's performance at each of the mini-games to vary without affecting the other mini-games, a deficiency in any single one of the capabilities required by the task would affect the user's ability to complete it as a whole. Each task was described both in terms of a single ability range describing its overall difficulty (based on the highest requirement of its sub-capabilities), as well as sets of ability ranges describing its underlying capabilities individually.

In the same way that the mini-games provided a measure of ability through recording a participant's performance, it was also possible to make a prediction of a participant's abilities to complete a task, based on their performance at other tasks. For example, as the control task required the lowest levels of ability (at the task related modality), completion of any other task provided the data necessary to 'predict' completion of the control task. Completion of the control task, could not however be used to predict performance at other tasks, given that they required a higher level of ability than was displayed through the use of the control task alone.

Where no suitable task level data was available to suggest that a participant was able to complete a particular task, comparison was then performed in terms of their underlying capabilities. If it was also not possible to identify a match in terms of the underlying capabilities the ability to recurse down to lower levels of detail provided a more informative prediction of performance.

9.3.3 Mediating Between Data from Different Sources

Although the interaction model only includes a single set of visual abilities, three different mini-games were created (details of which are provided in section 9.2.2 and appendix C) that all provided the same combination of sight, acuity and contrast related ability measures. The mini-games were designed to both use the range of potential data collection methods identified in chapter 2 and demonstrate the need for assessments that accurately reflect the capability under investigation.

- The Snellen-based mini-game provided an objective measure of visual acuity (human mediated).
- The word selection variant provided a subjective, Preference-based measure of visual acuity (self-reporting).
- The Landolt-based mini-game provided an objective measure of visual search ability (semi-automated).

By comparing between the results of each of these mini-games the importance of choosing (and validating) a representative assessment is highlighted. Table 9.4 contains the results of Paired T-Tests, run between each of the three mini-games described above and shows how they provided different scores for individuals' visual capabilities. Over the three capabilities that were measured (acuity, contrast and both combined) the Snellen-based mini-game provided on average a higher ability rating ($\bar{x} = 6.4$, $s.d. = 0.9$) than the preference-based mini-game ($\bar{x} = 5.4$, $s.d. = 1.3$) which was higher again than the Ladolt-based ($\bar{x} = 3.9$, $s.d. = 1.3$) mini-game.

Table 9.4: Paired T-Test Results Between Different Visual Mini-Games

Mini-Games		t	df	Sig. (2-tailed)
Snellen	Preference	4.1	26	.0
Preference	Landolt	6.6	26	.0
Snellen	Landolt	10.8	26	.0

The difference between the Snellen and preference-based mini-games was expected owing to their difference in subjectivity and is in line with the findings of chapter 6. However, the result is still interesting given that chapter 6 actually investigated self-efficacy, where-as the difference exhibited here is more attributable to participants' selection of a preference.

The difference between the Landolt and the other two mini-games was also expected as they actually measured different capabilities. By measuring participants' visual searching abilities, the Landolt-based mini-game should be positioned hierarchically above the other two tests; being comprised of visual acuity (seeing the targets), cognitive (reacting) and motor (hitting the appropriate key) elements.

By using the Dublin-Core "Creator" metadata tag, the data from each of the visual tests was coded to identify it as being provided by different sources. This both allowed the Landolt-based data to be selected for use over the other two scores and a reduced confidence rating to be provided when non-Landolt-based data was used in a comparison and demonstrated the ability of the framework to handle inappropriate data.

9.3.4 Identification of Mismatches

Following the demonstration of the ability of the framework to perform comparisons and identify matches between profiles at multiple hierarchical levels, the functional nature of those comparisons will now be demonstrated. The preferable result of an accessibility assessment is one indicating the potential for successful interactions to take place between two actors. Where this is not the case, a functional assessment should describe the mismatch that has occurred in terms of the gap that exists between the two actors (or the abilities that are available to facilitate communication via a different route).

Where a mismatch was observed between a participant's abilities and those required by a particular higher-level task [4] the mismatch that had occurred could be described at increasingly lower levels, in order to allow the nature of the problem to be identified. As an example, five of the nine participants were unable to complete task #2, due to the double-click setting of the task being inappropriate for their capabilities. Of those who failed the task, four participants tried to progress through the targets even though no sound had been played. Three of the four stopped after the second or third target, commenting about their inability to hear the acknowledgement sound and one user continued all the way to the end, ignoring the lack of feedback. Further questioning confirmed that all four were under the impression that their failure to complete the task was due to their hearing being insufficient for the requirements of the task. This was incorrect as the real mismatch existed between their finger dexterity and the double-click time required by the task (an inability to hear the acknowledgement sound). After being informed of the real nature of the problem, one participant was able to successfully complete the task—with increased effort and a significant number of errors—by increasing their double-click time.

Of the four participants who were able to complete the task on their first attempt, three had average double-click times that were lower than that required by the task and successfully clicked the first target on their first attempt. The other one had an average time that was 3ms higher than the requirement, however they successfully clicked the first target on their second attempt and were able to consistently increase their double-click time for the duration of the task.

Not only does this re-affirm the validity of the functional requirement, but also demonstrates the potential benefits provided by the frameworks ability to address it. Barriers to accessibility may occur for a number of different reasons and it was only through the comparison of the underlying capabilities of the task against the user's abilities that the actual barrier that existed was identified.

9.3.5 Enabling Speculative Augmentation

Where a mismatch is found to occur between two actors, a need arises for some form of assistance that is able to bridge the gap. The hierarchical nature of the framework facilitates the ability for successive actors to be speculatively placed inside of one another. This allows their capabilities (and processing functions) to be automatically attributed to the higher-level actor that they have been placed within, augmenting its capabilities and potentially improving its accessibility.

Task #1 was designed to be inaccessible to all of the participants, owing to the reduction in hearing ability (specifically in the higher frequencies) that is charac-

[4] And it was also not possible to find a match by recursing down to lower-level capabilities

teristic of the ageing process (Moller, 2006). By adapting the acknowledgement tone to output at a frequency of 12,000Hz, the sound produced was expected to be imperceivable no matter what volume it was played at. As such even though the volume of the tone was kept the same as in other tasks, none of the participants was able to complete the task.

The ability for speculative augmentation to be used to propose an adaptation to this problem was evaluated through the use of an adaptation that took the audio capabilities produced by the acknowledgement tone and transformed them to output via the visual modality. The adaptation was represented within the framework as an actor with two capabilities (one input and one output) and a processing function that was used to link them together. By augmenting the task with an additional visual output, although the audio mismatch was not resolved, the use of a different route ensured the accessibility of the acknowledgement component of the task. This then resulted in the accessibility of the overall task being reinstated.

9.3.6 Summary

Following the building of a prototypical system to assess the feasibility of implementing the framework, a practical evaluation has been performed to identify the ability of the system to deliver the function contributions. The result of this evaluation suggests that the system does deliver the contributions and allows this evaluation to now move on to assessing the validity of the underlying approach on which the system was developed.

9.4 Evaluation Against Thesis Aims

With a practical evaluation of the feasibility of delivering the thesis objectives performed in the previous section, this section now moves on to assessing the underlying aims described in chapter 2. The aims are used to provide an evaluation of the framework in terms of the research produced throughout the thesis as a whole. Each of the aims will be reviewed individually through the identification of its theoretical validity and supporting evidence provided by the data collection study. Validity is described in terms of six inter-related aspects (described further in appendix C.2). The aspects will be used to identify the individual contribution that each element of the thesis makes to the to the overall validity of the framework.

9.4.1 Functional Assessment of Accessibility

The functional assessment of accessibility is based on the functional model of disability and takes a neutral approach to the identification (and solution) of accessibility issues. Rather than the impairments of a user or the limitations of their technology, the functional approach focuses on the interaction itself; viewing it in terms of a series channels of communication that allow information to be passed from transmitters to receivers (in either direction). Where those interactions can be completed, the two actors are able to interact with each other and if one actor ($\mathbb{A}$) is able to fully satisfy the needs of the other ($\mathbb{B}$), $\mathbb{B}$ can be said to be fully accessible to $\mathbb{A}$. An accessibility problem (or barrier) can then be identified as existing when there is a mismatch (or gap) between the capabilities of the two actors, such that they are not fully compatible.

Where an accessibility barrier exists, rather than apportioning blame, the functional model describes it in terms of the gap that exists between the capabilities of each of the actors involved. By viewing interaction as the transmission of information between a series of actors, interactions are broken down into measurable units. Through the use of interaction-based vocabularies, the capabilities of both users and technologies can be captured and compared by the framework providing an ability to describe not only of the problem, but also the assistance that is required to bridge the gap.

9.4.1.1 Theoretical Validity

The framework carries out accessibility assessments through a combination of the use of the hierarchical description of actors and the provision of semantic and syntactical standardisation. Individual actors are able to expose their capabilities through the use of a hierarchical, interaction-based vocabulary and the accessibility of specific interactions can then be assessed through the direct comparison of each actor's relevant capabilities. The results of the comparison describe the matches and gaps between the two actors' capabilities and provide a functional assessment of their accessibility.

Through their shared theoretical basis in the Shannon-Weaver model of communication, substantive validity is provided to this claim, given the similarity of the framework in this aspect to the Universal Access Reference Model (UARM). As an extension, the UARM could also then be incorporated, with the work of Sala *et al.* (2011) and Iglesias-Perez (2010) used to provide the basis of a hierarchical layer of capabilities.

9.4.1.2 Evidence From the Practical Evaluation

Chapter 8 described the construction of a reasoner that provided a pattern-matching comparison function to traverse the relationships provided by the framework. The demonstration of the utility of this functionality in section 9.3.2 provides evidence to support the ability of the framework to deliver this aim.

9.4.2 Variability Between and Within Individuals

The appreciation of the variability between and within individuals relates to the requirement for the representation of diversity within profiles. As highlighted in chapter 2, every user is different and their profile should ideally reflect their personal characteristics, rather than fitting them to a stereotype. As well as variety between individuals, the characteristics of a single individual may change over time and their profile should also reflect this variability as well.

In a similar way, while much technology is mass produced, there are a large number of potential device, AT, software and adaptation combinations. As technology ages there may also be variability in its characteristics, either through wear-and-tear or 'upgrades' that change its functionality. This results in the potential for technology to display the same variety of abilities as is seen in users.

While the acknowledgement of diversity is important, profiles must still be comparable and there is therefore a need for standardisation in order to facilitate comparisons. This standardisation arises naturally from the variety seen in users and technology as they both vary in the same ways. The framework provides a mechanism by which this variability can be captured, by allowing a number of data items to be stored against any single one of an actors capabilities.

9.4.2.1 Theoretical Validity

The framework has been designed from the start to represent the diversity of individuals in a comparable way. The introduction of particles, atoms and instances in chapter 5 bears a resemblance to the object-oriented programming philosophy, which breaks the world down into classes, objects and instances. This provides a further element of substantive validity, due to the similarity of the framework to the existing accept approach. In terms of Messick's validity criterion, the behaviour of the framework is also substantively similar to the use of overlay models. Sosnovsky & Dicheva (2010) describes the 'historical' use of ontologies for modelling the structure of a domain and how their elements can then be employed in the form of an overlay to represent atomic user characteristics. The framework takes this approach and applies it to the modelling of both users and technology.

Additionally the use of a series of standard particles that are connected by pre-defined relationships provides a dynamic yet predictable semantic structure from which profiles can be created. Atoms then provide a means of expressing the ways that profiles vary by representing examples of actors, capabilities and modalities from which instances can be created. Finally, instances allow profiles to be individualised, whilst remaining comparable, due to their basis on atoms. In terms of Messick's validity criterion as the structure used for data storage is representative of the real life attribution of capabilities to users (demonstrated by the Universal Access Reference Model (Carter & Fourney, 2004b)) this also provides an element of structural validity.

9.4.2.2 Evidence From the Practical Evaluation

Through the implementation of the framework, its ability to represent the variety between and within individuals has been demonstrated. The use of RDF to implement the syntax (in section 8.4.1) provided the standardised, machine-readable format needed for comparisons to take place, while remaining dynamic enough to represent the diversity that is present between individuals.

Within the evaluation study, the use of mini-games based on pre-existing standardised assessments provides a demonstration of convergent validity within the approach. Convergent validity describes the extent to which a measure can be substantiated through its relationship to existing, pre-validated metrics and as the framework is able to accept data from existing metrics, its outputs will be convergent with the metrics that are used to provide data.

9.4.3 Variety Within and Interaction Between Accessibility Barriers

The variability and interaction between accessibility barriers relates to the requirement to represent diversity, both in the ways that an individual may be unable to access a particular piece of technology, and the underlying reasons behind that inaccessibility. Chapter 2 highlighted a number of different accessibility barriers, both in terms of their causes and effects. It was possible for both a single barrier to occur as the result of multiple different types of interaction as well as different barriers being caused as a result of the same interaction in different situations. In addition, it may be possible for multiple barriers to be present between two individuals and where multiple barriers exist, the interaction between the barriers can result in barriers of greater complexity and significance.

9.4.3.1 Theoretical Validity

The framework has been designed to addresses both of the concerns surrounding barriers to accessibility. Firstly, the diversity in accessibility barriers arises naturally from the individual nature of the profiles that are used for comparisons. Given the diversity of individuals' highlighted in the previous subsection, it follows that interaction between individuals may be impaired for a variety of reasons – based on the compatibility of their individual abilities. This results in diversity between the levels of success (or failure) of individual interactions and results in variety between the accessibility barriers observed.

Secondly, the hierarchical nature of the framework allows the representation of higher-level capabilities in terms of their lower-level constituents. Where a higher-level capability can be inferred from multiple lower-level ones, the existence of lower-level accessibility barriers can be used to to predict potential higher-level issues. In addition, where an individual has multiple minor (lower-level) impairments, it is often the case that their combination can lead to higher severity higher-level impairments. Within the framework, inferencing functions are used to describe the relationship between a higher-level capability and each of its constituents. Inferencing functions therefore provide an appreciation of the interaction between different accessibility barriers.

In terms of Messick's criterion of validity, the similarity between the representation of accessibility solutions within the framework and their real life construction provides an aspect of structural validity. In addition, the use of the same structure to represent users, technology and assistive technology is an approach that has already been proven within the Universal Access Reference Model (Carter & Fourney, 2004b) adding an element of substantive validity.

9.4.3.2 Evidence From the Practical Evaluation

The evaluation study highlighted the way that the accessibility of an interaction is dependent on the underlying capabilities via which that interaction is facilitated. As a general observation, each task required the combined use of four separate capabilities and a reduction in any one of them resulted in a reduced accessibility of the overall task.

More specifically, task #3 used a reduction in both target and label size as a means of demonstrating the effect of the combination of multiple minor impairments in creating a more fundamental accessibility barrier. Although the targets and their labels could be adapted separately, the reduction of both together was used to represent the effect caused through the resizing of interface widgets. A reduction in widget size not only impacts the physical size of a widget (resulting in a change to its motor related capabilities), but also reduces the space available for its label (resulting in a change to its visual capabilities).

Similar to the effect observed in section 9.3.1, the minor-to-moderate impairments observed as a result of each of the capabilities separately resulted in the failure of the task as a whole. This can be explained with relation to the Model Human Processor (providing an element of content and convergent validity), in terms of the effect that the two capability changes have on the perception and motor processes required for the completion of the task. The reduced visual bandwidth resulted in an increase in the time taken to identify each target, and the reduced target size then had a separate (but similar) effect on the target acquisition.

9.4.4 Variety Within and Interaction Between Accessibility Solutions

The variety and interaction between accessibility solutions describes the need for the framework to be able to represent the diverse range of ways that assistance can be provided. The accessibility solution that is appropriate for an individual in a given situation may be provided through variation in the choice of device, assistive technology (AT) and adaptations used (Sloan *et al.*, 2010). Further, the accessibility solution may also be provided through variations within the user's abilities, as well as those of the technology they are using. Chapter 2 described both the ability of each solution to individually influence the accessibility of an interaction and the cumulative effect that arises from the use of several solutions in combination.

Section 9.4.2 discussed the need to represent all *actors* as individuals. This aim describes the need to view all *accessibility solutions* as individuals and recognises that their components can be applied to multiple situations, based on their ability to span the functional gaps of different accessibility barrier. In addition, solutions may have unintended consequences which, while increasing the bandwidth in one channel, decreases the bandwidth of another. A solution to a high-level barrier may therefore require multiple components, all of which are targeted at different low-level barriers and may interact with each other.

9.4.4.1 Theoretical Validity

The framework tackles this aim through the combination of the approaches taken in the previous sections. Firstly, all actors are represented using the same structure, meaning that as far as the framework is concerned, there is no difference between an assistive technology and any other piece of technology. The variability that exists within different accessibility solutions is therefore addressed in the same way that variability can be represented between different individuals.

By treating accessibility solutions as actors, their use within interactions can be understood through the same comparison functions used to assess any other interaction. If changes are made to either of the individuals in an interaction, or an assistive technology is placed between them, re-comparison will reflect the change that has been made and describe the resulting accessibility.

As described in chapter 5, all of the different types of accessibility solution can be represented using the framework. Their use in combination can also be modelled through the inclusion and removal of different sub-actors within a profile. In this way, speculative augmentation can then be used to make changes to the actors within a comparison, in order to assess their effect on its accessibility. In terms of Messick's validity criterion, the representation of accessibility solutions in a form that facilitates their direct assessment against user and technology capabilities provides an element of structural validity.

9.4.4.2 Evidence From the Practical Evaluation

The use of the visual adaptation in task #1 demonstrated its ability to transform audio-based acknowledgement information into the visual medium in order to overcome the mismatch that existed. This form of adaptation actually replicated the use of a traditional assistive technology, however, another example of an accessibility solution can be identified in relation to the data collection study.

Although the study took place on a single piece of hardware, the screen that was used to display the mini-games/tasks provided a translation between the font size (px) that had been specified for the label and the actual size (mm) with which the text was displayed on the screen. By changing the resolution of the screen, it would have been possible to alter the size at which a single version of the visual mini-game/task was displayed. While the resolution of the screen was kept static[5] the resultant behaviour of different screens can be modelled through the use of actors with appropriately calibrated processing functions.

9.4.5 Variety of Agents Producing and Using Data

The data required to drive assessments using the framework may be provided and/or used by a number of different agents including: the user themselves, human mediators and technology-based automated (or semi-automated) agents. This range of agents lead to the development of a common representation for the storage and manipulation of data. The basic design of the framework was provided in

[5] Due to the effect that the change would have had on controlled elements of the task, e.g. the cursor speed.

chapter 5 and consisted of a standard series of particles from which models could be created.

The actual technologies chosen to provide an implementation of the framework were based on their ability to represent semantic data and perform pattern matching. The use of technologies taken from the Semantic Web stack, was however also intended to demonstrate the potential interoperability and transportability of the data stored by the framework.

9.4.5.1 Theoretical Validity

The semantic standardisation of the framework and syntactic standardisation within the implementation provided the potential for data to be collected from multiple agents. In the same way, by standardising the format in which data is stored, the framework has also been designed to facilitate the storage of capability data for multiple temporal and contextual situations. 'Data item' particles are self-contained, providing both a measurement value and the information necessary to interpret it. Values are stored complete with their scale (indicating measurement level) and the units in which they were measured. This results in an element of content validity, through the ability for a description of the measurement type to be used as a means of providing the information necessary for the selection of an appropriate comparison function.

9.4.5.2 Evidence From the Practical Evaluation

While the evaluation study was conducted on a single device, a number of steps were taken to replicate the provision of data by multiple agents. Firstly, a number of different mini-games were produced to act as semi-autonomous data collection agents. While they measured a variety of different modalities, their outputs could all be stored in the RDF-XML format identified as appropriate to facilitate the storage of data in a format that would be widely transportable. The outputs from different mini-games were then stored separately in order to replicate the potentially distributed nature of data storage in a real system.

Data was provided by several of the mini-games at multiple hierarchical levels with data being provided to describe both a higher-level capability and its underlying sub-capabilities. In addition, participants' vision was also measured using three different mini-games, which all produced a different values in response to measuring a similar capabilities. Through the attribution of appropriate metadata the 'data context' of the capabilities could be identified and data used appropriately.

9.4.6 Variability of Data Quality

With data being provided by a range of agents, there will be variability both in terms of its accuracy, and applicability to different situations. Given the temporal variability of individuals profiles, multiple data items may be stored in order to represent the variety in individuals' fluctuating capabilities and the agents measuring them. Additionally, the capabilities of an individual are constantly changing (even if only slowly). A decline may be observed in a given set of capabilities owing to the effects of ageing process, while at the same time improvements may be observed in others owing to learning and skills acquisition. In addition to the natural changes driven by the longitudinal processes described in the previous sentence, variability may also be caused in the short term by changes to the context in which interactions are taking place.

The framework has therefore been designed to facilitate the storage of both capability data (through the use of its structural and data layer components) as well as contextual information. In this way profiles have the facility both to be easily updated, as well as being able to hold multiple measurements about a single capability, based on their performance in different contexts.

While the storage of multiple data items for each capability allows the most suitable to be chosen, there may still be discrepancies between the chosen data item and the situation to be modelled. Section 9.4.5 described the use of data items as a means of capturing and representing the context in which data is collected. Contextual data can then be used to determine the suitability of a piece of data for use in modelling a specific situation.

9.4.6.1 Theoretical Validity

Each data item included the contextual information necessary to describe its suitability for use in modelling a given situation and by storing multiple data items against a single capability, the most appropriate one could then be chosen. In terms of Messick's validity criterion, this behaviour of the framework provides an element of generalisability validity, by ensuring that assessments are conducted using data that provides an accurate representation of the abilities of the actors involved. The validity of data is tied to its context of measurement and the ability to describe similar situations to which the results can be applied, provides another example of the ability of the framework to quantify the generalisability of an assessment.

9.4.6.2 Evidence From the Practical Evaluation

The contextual element of the data collection study was confined to the provision of data context regarding the author of the data. However, much of the rest of the contextual layer (the actor, partner and environmental contexts) relies on the same structural components as the lower two layers. The evidence provided to demonstrate the ability of the lower layers to meet their functional contributions, can be used to infer similar abilities for the contextual layer. Given the reliance of the contextual layer on the collection of appropriate contextual data (and the collection agents this then required) the validity of this inference as a measure of demonstrating this aim is limited.

9.5 Summary

The evaluation contained within this chapter has suggested the potential utility of the framework as a means of providing accessibility assessments between people and pieces of technology.

This chapter has provided an evaluation of the framework by discussing its ability to satisfy the aims and contributions developed in chapter 2. The evaluation was performed as both a discussion of the theoretical underpinnings of the framework and the practical concerns surrounding its use. The use of a data collection study provided real data against which the contributions of the practical implementation were assessed. The theoretical validity of the aims was then identified in terms of the inter-related validity criterion identified in Messick (1995), with the discussion focused on the potential benefits that the framework provides through the use of its previously described functionality.

9.5.1 Limitations

While the research has been built on top of existing proven theories, their use has resulted in two limitations being present in the approach taken.

9.5.1.1 "Brute Force" Approach to Data Storage and Comparison Process

The framework provides a comprehensive method of storing and reasoning data. Data is stored semantically using a series of connected elements that result in profiles being built up gradually using a number of different layers. This approach has been taken in order to provide a separation between the different functional components of the framework and allow their associated reasoning processes to be built and upgraded independently.

As a result, it is intended that two profiles will initially be compared at a superficial level of fidelity, through the use of simple pattern matching algorithms. Greater levels of detail will then be added, by increasing:

- The number of capabilities used to describe a single interaction – describing it using multiple lower levels to provide additional detail.
- The number of data item that are attributed to a single capability – describing its ability in multiple contexts (either temporally or environmentally).
- The amount of data that is attributed to a single data item – describing it in terms of different types of context (e.g. data, actor, partner and environment).

In this way, while the approach has been designed to allow more efficient comparison and searching algorithms to be developed, an increase in functionality is likely to result in an increase in the amount of data that is required. As the complexity of comparison processes increases, their requirement for more extensive data collection and storage facilities will also increase accordingly. The approach taken within this thesis could therefore currently be considered to have been provided through "brute force" data collection, storage and comparison processes. As described above however, there is a great deal of scope for increasing the efficiency of matching algorithms. This could potentially be achieved through the use of expert knowledge to augment the matching process, with higher layers employed that are aware of the relationship between different modalities (as identified in Atkinson (2012)). Data storage could then also be rationalised through the regular use of data aggregation and maintenance processes, with data retrieval functions similarly written to maximise their efficiency as well.

9.5.2 Dependencies

As it draws together a number of threads from different areas of research, the practical implementation of the approach is therefore dependent on a number of existing lines of inquiry:

Semantic web technologies are not yet fully realised, meaning there are still issues to be overcome before they can be widely adopted. A lack of accepted standards at the upper levels of the technology stack has resulted in the need for additional technologies to be used in order to achieve the functionality described at present. There are also outstanding questions regarding the scalability of graph-based data storage.

As they are researched further, many of these issues will disappear, assuming the further development and widespread use of semantic technologies. In addition, the identification of current deficiencies can be used to inform new research and development efforts. In terms of scalability, the modular and multi-layered design of the framework can be used to take advantage of pre-processing techniques that only present the data that is needed to evaluate a particular situation.

Data acquisition is a particular challenge with the problem being twofold: (1) initial population may require a bootstrapping procedure, and (2) information would need to be kept up-to-date. ISO 24756 for example, relies on the user (or other human agent) to create and update CAPs. However for mass adoption, the user cannot be relied upon either because (1) they may be unwilling, or (2) the information they provide may be unreliable (Godoy & Amandi Godoy & Amandi (2005)).

The framework is therefore reliant on the availability of agents that are able to populate user profiles. Given the nature of their task, current agents are highly specialised and tend to be developed as part of standalone systems (e.g. Hurst (2009); Trewin *et al.* (2006); Cheng *et al.* (2013)). The intention of this framework is to allow agents to expose their data and as a result allow it to be used as part of a more holistic profile.

Matchmaking algorithms and assistance delivery mechanisms are, like data acquisition agents, out of scope given The focus of this research on data representation and storage. They are however required in order for the approach provided in this paper to be realised and work aimed at developing a Global Public Inclusive Infrastructure is providing research in this area (Vanderheiden & Treviranus, 2011a; Loitsch *et al.*, 2012; Atkinson, 2012).

9.5.3 Wider Achievements

Despite its potential limitations, the research described within this thesis, detailing the development and evaluation of the framework, has resulted a number of achievements that result in it making a valid contribution to current knowledge. The research has intentionally taken a broad view of the accessibility landscape, similar to the approach taken by current projects such as the Global Public Inclusive Infrastructure (GPII) and Virtual User Modelling (VUMS) cluster. However unlike both of the above, the framework can be used to construct models representing both people and technology and it allows those models to be created in different levels of details. While the importance of user modelling cannot be underestimated, by reacting to the need for transportability of data, both of the projects that have been mentioned fail to address the underlying cause of this need: the variability of technology modelling approaches. By focusing solely on the user, existing projects miss the opportunity to simplify the technology modelling process.

The approach proposed in this thesis addresses this underlying problem and rather than creating a static modelling format, a framework has been proposed to guide the creation of models that can be used as the basis for the collection of data, which can then be stored in transportable profiles. By modelling technology in the same way as users, profiles are directly comparable and observation of the flow of information from one to another can be used to identify the functional reasons for any accessibility barriers that exist between them. Although the current approach already provides an ability to improve the comparability of different technology configurations for a particular user, it is also possible for direct comparisons to be made between the suitability of different pieces of software that may be used within a single device.

9.5.4 Impact

This chapter has presented an evaluation of the framework developed throughout the thesis, in terms of both its practical utility and theoretical validity. A data collection study was used to provide data from real users, which allowed the prototypical system developed in chapter 8 to be assessed against the functional contributions identified in chapter 2. The use of the framework within a practical implementation provided evidence to support its feasibility as an approach for use in the assessment of accessibility and subsequent provision of appropriate assistance. The findings of this assessment were then used as part of an evaluation of the framework against a series of inter-related validity criterion, providing an assessment of the underlying approach against the thesis' aims. The next chapter will present a summary of the contributions and impact of the work contained within this thesis as a whole.

Chapter 10

Conclusion

This chapter will highlight the contributions and impact of the work contained within this thesis. The individual contribution that each chapter has made will be identified in terms of their impact on the thesis as a single coherent piece of research. The outputs of the thesis as whole will then be described in terms of the contribution it makes to the wider body of research. Following this, the potential for extensions to the work will be discussed and a summary of the achievements of the thesis provided.

10.1 Contributions

This thesis presents the culmination of a number of strands of research that have been combined to present a holistic investigation of the problems surrounding the modelling of accessibility.

10.1.1 Individual Chapters

Each of the individual chapters represents a contribution to the thesis as a whole in terms of either defining and providing evidence to support the existence of a problem or developing, implementing and evaluating a potential solution.

Chapter 2 identified the problems associated with the provision of assistance to people with multiple dynamic impairments and the subsequent need for a framework that guides the modelling of accessibility as a tool for sustaining their technology use.

Chapter 3 provided a method through which the research was carried out.

Chapter 4 combined existing approaches to developed a novel, flexible approach based on the use of standard elements, connected by semantic relationships, which provide functional accessibility assessments at varying (appropriate) levels of fidelity, while maintaining the transportability of data.

Chapter 5 provided a theoretical framework (based on the previous approach) that facilitates the collection and storage of data from various sources, the creation of profiles and the construction of accessibility models in turn.

Chapter 6 compared the accuracy of older people and semi-automated data collection agents in terms of their ability to provide the quality of data required to drive a system based on the design developed in chapter 5. In doing so it highlighted the potential for a series of factors to be used as an indication of varying degrees of quality in older people's subjective judgements and the need for meditation processes to be developed in order to facilitate the practical implementation of the design.

Chapter 7 expanded the framework to address the variability of data accuracy and the need for permission-based profile controls, by providing threshold-based mediation of contextual meta-data.

Chapter 8 demonstrated the implementability of the theoretical framework as both a means of validating its feasibility and a vehicle to enable further evaluation to take place.

Chapter 9 evaluated the ability of the framework to satisfy the aims of the thesis and its resulting validity in terms of providing a solution to the problems identified in the first chapter.

10.1.2 Objectives

Through the contributions of the individual chapters, the thesis' objectives were all met:

- "A search of relevant literature will be used to better define the problem and identify existing approaches that provide partial solutions." Chapters 2 and 4 provided a search of existing literature and approaches.
- "The potential solutions will be combined to create a novel approach from which a design can be developed." Chapter 5 described the resulting design.
- "An investigation of the ability of older people to provide the data required to drive the approach will be conducted." Chapter 6 described the study that was performed and the resulting additions that were required to the design.

- "The ethical implications of automated data collection will be explored, resulting in the identification of a mechanism to allow mediation between data from different sources." Chapter 7 provided a discussion that lead to the development of the required mechanism.

- "The design and data mediation mechanism will be integrated and a prototypical system will be implemented." The implementation was described in chapter 8.

- "Data collection and evaluation studies, to provide data from real users for use within the implementation to evaluate the effectiveness of the approach in solving the problem." In chapter 9 both of the studies are described, providing evidence of the ultimate fulfilment of the thesis' aims through the demonstration of the functional contributions within the implementation.

10.1.3 Thesis Contributions

As a result of meeting the objectives, the thesis makes a number of contributions to the wider body of research as follows:

- A review of the challenges involved in the provision of assistance to people with multiple dynamic impairments; with an emphasis on the collection, storage and use of data for holistically modelling accessibility.

- Development of a flexible approach to representing interaction, that supports the functional comparison of users and technology at varying (appropriate) levels of fidelity.

- Proposal of a framework to facilitate the collection, storage and comparison of data from a number of different sources, through the use of standard elements and semantic relationships.

- Evidence to suggest a series of factors for use in identifying the accuracy of older adults' self-reported ability data.

- The ability to mediate between data based on its accuracy and contextual appropriateness.

- Support for the use of speculative augmentation to simulate the effects of different forms of assistance (e.g. device, software, AT and adaptation) on the interactions between users and technology.

10.1.4 Aims

Through meeting the objectives, and delivering the contributions above the thesis' aims were met, as can be seen in the corresponding sections in the evaluation:

Aim 1: "Functional Assessment of Accessibility" – Section 9.4.1.

Aim 2: "Variability Between and Within Individuals" – Section 9.4.2.

Aim 3: "Variety Within and Interaction Between Accessibility Barriers" – Section 9.4.3.

Aim 4: "Variety Within and Interaction Between Accessibility Solutions" – Section 9.4.4.

Aim 5: "Variety of Agents Producing and Using Data" – Section 9.4.5.

Aim 6: "Variability of Data Quality" – Section 9.4.6.

10.2 Practical Reuse of the Research

The research contained within this thesis is able to stand alone and provides its own contribution to the field. This section will however describe the work that is needed to enable the reuse of different aspects of the thesis.

10.2.1 The Approach

The approach that has been taken throughout this thesis has benefited from both theoretical and practical development work. In order to enable its future use, further work is required in the areas described in section 9.5.2, particularly the greater availability of data acquisition agents. In practice the approach has also been designed on the assumption that a greater degree of personalisation will be available and interfaces will therefore be required that have the potential to support adaptation. The approach is currently best suited to for use in analysing interaction and therefore canonical models of people and technology will also be beneficial.

10.2.2 The Implementation

In order to enable further use of the practically implemented aspects (e.g. the prototypical system developed in chapter 8) a degree of ruggedisation will be required. The code has been developed in a modular fashion, however, it contains minimal error checking or optimisation and in addition the code-base currently exists solely as a series of Prolog rule repositories. While the repositories can be reused 'as-is' in future Prolog-based projects, packaging them into a standalone reasoner (potentially interpreted via a more common language) would increase the "accessibility" of the reasoning engine for use in an internet focused domain.

10.3 Extensions and Further Research

Although the research presented in this thesis constitutes a valid contribution to current knowledge in itself (and could benefit from futher research in the areas described above), it also has the potential to generate derivative work, extending its potential scope. Section 9.5 identified at the end of the previous chapter acknowledged the reliance of this research on technologies and techniques that are still under development. There is however the potential for further research focusing on the framework itself.

10.3.1 Data Aggregation

Chapter 7 identified a range of stakeholders that might wish to exploit the data that could be produced by systems that were based on the framework. The widespread collection and storage of user data provides the potential for aggregation functions to be developed as a means of exposing appropriate data.

10.3.2 Additional Functional Layers

The framework has been designed with three functional layers that address the common concerns of current accessibility models. Each layer builds on top of the one 'below' it and provides additional information that in turn results in a greater level of functionality. The modular structure of the framework allows the potential for an incremental implementation and as such extensions to the framework could be developed separately.

10.3.3 Context

Context can have a large effect on the usability of data and at present the framework is able to both store it and then use it to determine the likelihood that a match is accurate. Where there is a lack of data to describe an actor's capabilities in a given context it is currently not possible to infer it (other than finding the closest possible match or using a stereotype). This concern could be addressed through the use of human capabilities and the framework's ability to model the environment as an actor.

As interference (e.g. background light or noise) is stored as a capability in a format compatible with the user's abilities to perceive them, it should be possible to determine the effect that a particular contextual capability has on an actor's capabilities. Known capabilities and contexts could then be used to extrapolate capabilities in other contexts. Techniques for reasoning based on environmental constraints have been explored in (Atkinson, 2012) but are out of the scope of the research described in this thesis.

10.3.4 Standardisation

Various elements of the research have drawn inspiration from standards that have been formally recognised in their own right such as: the Common Accessibility Profiles (ISO/IEC 24756, 2009), the Dublin Core (DCMI, 2012) and the majority of the Semantic Web stack. Standardisation presents a natural extension to the research and there are a number of national and international bodies that register standards; including the International Organization for Standardization (which is mainly concerned with proprietary standards) and the W3C (who publish open web-related standards).

While the intention of the research has not been to promote a rigidly prescriptive format, the use of a framework (or reference model) represents an effort to guide the storage of data its resulting comparison processes as a means of increasing their transportability. The use of standard elements joined together by standard semantic relationships, provides enough structure to warrant the creation of a standard (which could be developed from chapters 5, 7 and 8).

The standardisation process would provide both a dissemination exercise and a method of creating wider debate on the suitability of the framework. In promoting the framework as a standard, it would be subject to critique by experts from both the academic and industrial communities and while further work may be needed before this was possible it remains an avenue for further investigation.

10.3.5 Use Within Alternative Domains

Through abstracting out the approach that has been developed, the framework could be applied to a number of different scenarios outside of those on which the thesis is based. Hierarchical task analysis has already been applied to a number of different disciplines and has a history that stems from the desire to streamline industrial processes (Crystal & Ellington, 2004). The use of channels based on higher level work related skills and capabilities as a means of matching employees to internal vacancies and identifying training needs was investigated during a project detailed in Appendix D.3 ("Matching People to Jobs").

10.4 Summary

An adaptive capability profiling framework proposed within this thesis as a means of both harmonising the collection and storage of user data and providing a dynamic approach to the comparison of users and technology, in order to identify their resulting accessibility.

While the approach is new, it builds on a range of existing techniques and technologies that have been selected the individual benefits they can provide. Though there is the potential for further research and development, the approach has already provided a contribution in terms of its inclusion within the Sus-IT project and a W3C Research and Development Working Group Symposium on User Modelling for Accessibility.

References

2001. *International classification of functioning disability and health (ICF)*. (used on p 12 and 35.)

Abadi, Daniel J., Marcus, Adam, Madden, Samuel R., & Hollenbach, Kate. 2007. Scalable Semantic Web Data Management Using Vertical Partitioning. *Pages 411–422 of: Proceedings of the 33rd International Conference on Very Large Data Bases*. VLDB '07. VLDB Endowment. (used on p 211.)

Abel, Fabian, Henze, Nicola, Krause, Daniel, & Plappert, Daniel. 2008. User Modeling and User Profile Exchange for SemanticWeb Applications. *Pages 4–9 of: LWA*, vol. 448. Citeseer. (used on p 90.)

Ackermann, Philip, Velasco, Carlos A., & Power, Christopher. 2012. Developing a semantic user and device modeling framework that supports UI adaptability of web 2.0 applications for people with special needs. *Pages 12:1–12:4 of:* Vigo, Markel, Abascal, Julio, Lopes, Rui, & Salomoni, Paola (eds), *Proceedings of the International Cross-Disciplinary Conference on Web Accessibility*. W4A '12. New York, NY, USA: ACM. (used on p 51 and 79.)

Adair, John G. 1984. The Hawthorne effect: A reconsideration of the methodological artifact. *Journal of applied psychology*, **69**(2), 334. (used on p 137.)

Agarwal, Anshu, & Meyer, Andrew. 2009. Beyond usability: evaluating emotional response as an integral part of the user experience. *Pages 2919–2930 of: CHI EA '09: Proceedings of the 27th international conference extended abstracts on Human factors in computing systems*. New York, NY, USA: ACM. (used on p 41 and 139.)

Amundson, Ron. 2000. Against normal function. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, **31**(1), 33–53. (used on p 10.)

Annett, John, & Duncan, Keith D. 1967. Task analysis and training design. *Occupational Psychology*, **41**, 211–221. (used on p 85.)

Arif, Ahmed Sabbir, & Stuerzlinger, Wolfgang. 2009. Analysis of text entry performance metrics. *Pages 100–105 of: Science and Technology for Humanity (TIC-STH), 2009 IEEE Toronto International Conference*. IEEE. (used on p 140.)

Atkinson, Matthew. 2012. *Collaborative Adaptive Accessibility and Human Capabilities*. Ph.D. thesis, Loughborough University. (used on p xix, 36, 194, 236, 261, 262, 269, 309, and 310.)

Atkinson, Matthew, Li, Yunqiu, Machin, Colin, & Sloan, David. 2010. Towards accessible interactions with pervasive interfaces, based on human capabilities. *Pages 162–169 of:* Miesenberger, Klaus, Klaus, Joachim, Zagler, Wolfgang, & Karshmer, Arthur (eds), *Computers Helping People with Special Needs*. Lecture Notes in Computer Science, vol. 6179. Springer Verlag Berlin / Heidelberg. (used on p 36, 53, and 72.)

Atkinson, Matthew Tylee, Machin, Colin H. C., & Sloan, David. 2008. Making Accessibility Accessible. *In:* Petrie, Helen (ed), *Proceedings of Accessible Design in the Digital World 2008*. University of York. (used on p 29 and 191.)

Atkinson, Matthew Tylee, Bell, Matthew J., & Machin, Colin H. C. 2012. Towards ubiquitous accessibility: capability-based profiles and adaptations, delivered via the semantic web. *Pages 14:1–14:4 of:* Vigo, Markel, Abascal, Julio, Lopes, Rui, & Salomoni, Paola (eds), *Proceedings of the International Cross-Disciplinary Conference on Web Accessibility*. W4A '12. New York, NY, USA: ACM. (used on p 174 and 316.)

Balasubramanian, Vidhya, & Venkatasubramanian, Nalini. 2003. Adaptation of Multimedia Data for Cross-Disability Access. *revision, ACM Multimedia Systems Journal*. (used on p 72.)

Ballantine, Joan A., Larres, Patricia McCourt, & Oyelere, Peter. 2007. Computer usage and the validity of self-assessed computer competence among first-year business students. *Computers & Education*, **49**, 976–990. (used on p 41, 129, and 130.)

Bandura, Albert. 1978. Self-efficacy: Toward a unifying theory of behavioral change. *Advances in Behaviour Research and Therapy*, **1**(4), 139 – 161. Perceived Self-Efficacy: Analyses of Bandura's Theory of Behavioural Change. (used on p xvi, 40, 126, 127, and 128.)

Bandura, Albert. 1995. Exercise of personal and collective efficacy in changing societies. *In: Self-Efficacy In Changing Societies*. (used on p 127 and 175.)

Bandura, Albert. 2006. *Self-Efficacy Beliefs of Adolescents*. Vol. 5. Information Age Publishing Inc. Chap. 15 – Guide for constructing self-efficacy scales, pages 307–337. (used on p 139.)

Basser, L.A., & Jones, M. 2002. The Disability Discrimination Act 1992 (CTH): A Three-Dimensional Approach to Operationalising Human Rights. *Melb. UL Rev.*, **26**, 254. (used on p 8.)

Baxter, Joseph, Hungerford, Bruce Carter, & Helms, Marilyn Michelle. 2011. Predicting success in the introduction to computers course: GPA vs. student's self-efficacy score. *Information Systems Education Journal*, **9**(2), 75. (used on p 129, 130, and 131.)

Bell, M. J., Machin, C., & Stone, R. 2014. Using a Common Semantic Structure to Provide Comparable Contextual Models of Users and Technology. *In: Universal Access in Human-Computer Interaction*. Springer. (Accepted for publication). (used on p 94 and 316.)

Bell, Matthew J., & Machin, Colin H. C. 2013. Increasing the Flexibility of Accessibility Modelling Through the Use of Semantic Relationships. *Page article 6 of:* Yehya Mohamad, Christos Kouroupetroglou (ed), *W3C WAI RDWG Symposium on User Modeling for Accessibility*. (used on p 316.)

Bell, Matthew J., & Machin, Colin H.C. 2009. A framework for adaptive communication design. *Pages 45–50 of: Proceedings of the 27th ACM international conference on Design of communication*. SIGDOC '09. New York, NY, USA: ACM. (used on p 316.)

Bell, Matthew J., Sloan, David, Li, Yunqiu, Atkinson, Matthew T., & Machin, Colin. 2010 (October). The Benefits and Potential Pitfalls of User Monitoring. *Pages 7—8 of: The AAATE Workshop: Assistive Technology — Technlogy Transfer*. (used on p 181 and 316.)

Benyon, D. R. 1993. Accommodating Individual Differences through an Adaptive User Interface. *In:* Schneider-Hufschmidt, M., Kühme, T., & Malinowskicaven, U. (eds), *Adaptive User Interfaces - Results and Prospects*. North-Holland, Amsterdam: Elsevier Science Publications. (used on p 36.)

Benyon, David, & Murray, Dianne. 1993. Applying user modeling to human-computer interaction design. *Artificial Intelligence Review*, **7**(3-4), 199–255. (used on p 16, 37, 50, and 51.)

Berrueta, Diego, & Polo, Luis. 2011 (December). *Measurement Units Ontology*. (used on p 114.)

Bettini, Claudio, Brdiczka, Oliver, Henricksen, Karen, Indulska, Jadwiga, Nicklas, Daniela, Ranganathan, Anand, & Riboni, Daniele. 2010. A survey of context modelling and reasoning techniques. *Pervasive and Mobile Computing*, **6**(2), 161 – 180. Context Modelling, Reasoning and Management. (used on p 20.)

Bickenbach, Jerome E, Chatterji, Somnath, Badley, E.M, & Ustun, T.B. 1999. Models of disablement, universalism and the international classification of impairments, disabilities and handicaps. *Social Science & Medicine*, **48**(9), 1173 – 1187. (used on p 12.)

Biswas, P., & Robinson, P. 2010. A brief survey on user modelling in HCI. *In: International Conference on Intelligent Human Computer Interaction (IHCI)*. (used on p 53, 86, and 87.)

Biswas, Pradipta, & Robinson, Peter. 2013. Evaluating interface layout for visually impaired and mobility-impaired users through simulation. *Universal Access in the Information Society*, **12**(1), 55–72. (used on p 235 and 297.)

Biswas, Pradipta, Kaklanis, N., Mohamad, Y., Peissner, M., Langdon, Pat, Tzovaras, D, & Jung, Chrisoph. 2013. An Interoperable and Inclusive User Modeling Concept for Simulation and Adaptation. *Pages 195–236 of: A Multimodal End-2-End Approach to Accessible Computing*. Springer London. (used on p xv, 49, 50, 51, and 52.)

Bliss, J., Chandra, P., & Cox, M. 1986. The introduction of computers into a school. *Computers & Education*, **10**(1), 49–54. (used on p 134.)

Bohman, Paul Ryan, & Anderson, Shane. 2005. A conceptual framework for accessibility tools to benefit users with cognitive disabilities. *Pages 85–89 of: W4A '05: Proceedings of the 2005 International Cross-Disciplinary Workshop on Web Accessibility (W4A)*. New York, NY, USA: ACM. (used on p 15 and 26.)

Böhme, Rainer, & Köpsell, Stefan. 2010. Trained to Accept?: A Field Experiment on Consent Dialogs. *Pages 2403–2406 of: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '10. New York, NY, USA: ACM. (used on p 183.)

Bougie, Theo, & Heerkens, Yvonne. 2009 (December). *A revised version of N19rev based on ISO 9999, version 2007*. Tech. rept. Dutch Normalisation Institute and Dutch WHO FIC Collaborating Centre. (used on p 35 and 68.)

Broome, Travis, & Havelka, Douglas. 2011. Determinants of computer anxiety in business students. *Review of Business Information Systems (RBIS)*, **6**(2), 9–16. (used on p 139.)

Browne, Dermot, Totterdell, Peter, & Norman, Mike (eds). 1990. *Adaptive User Inferfaces.* London, UK, UK: Academic Press Ltd. (used on p 37.)

Brusilovsky, Peter, & Millán, Eva. 2007. The adaptive web. Berlin, Heidelberg: Springer-Verlag. (used on p 120 and 293.)

BS 4274-1:2003, British Standards Institution. 2003. *Visual acuity test types. Test charts for clinical determination of distance visual acuity.* (used on p 236 and 309.)

BS 8878, British Standards Institution. 2010. *Web Accessibility. Code of Practice.* (used on p 32.)

Burchardt, Tania. 2004. Capabilities and disability: the capabilities framework and the social model of disability. *Disability & Society*, December, 735–751. (used on p 11.)

Burzagli, Laura, Emiliani, Pier Luigi, & Gabbanini, Francesco. 2009. Design for All in action: An example of analysis and implementation. *Expert Systems with Applications*, **36**(2, Part 1), 985 – 994. (used on p 15 and 36.)

Caldwell, Ben, Cooper, Michael, Reid, Loretta Guarino, & Vanderheiden, Gregg. 2008. *Web Content Accessibility Guidelines (WCAG) 2.0.* (used on p 313.)

Capgemini. 2012. *Evaluating the work of the UK Digital Champion and Race Online 2012.* Tech. rept. Capgemini Consulting. (used on p 54.)

Card, S., Moran, T., & Newell, A. 1986. The model human processor- An engineering model of human performance. *Handbook of perception and human performance.*, **2**, 45–1. (used on p 141 and 294.)

Card, Stuart K., Newell, Allen, & Moran, Thomas P. 1983. *The Psychology of Human-Computer Interaction.* Hillsdale, NJ, USA: L. Erlbaum Associates Inc. (used on p 86.)

Carroll, John M., Rosson, Mary Beth, & Zhou, Jingying. 2005. Collective efficacy as a measure of community. *Pages 1–10 of: CHI '05: Proceedings of the SIGCHI conference on Human factors in computing systems.* New York, NY, USA: ACM. (used on p 128.)

Carter, Jim, & Fourney, David. 2004a. *The Common Accessibility Profile.* Tech. rept. Tech. Rep. No. TR2004-02. Department of Computer Science, University of Saskatchewan, Saskatoon, Saskatchewan, Canada. (used on p 51.)

Carter, Jim, & Fourney, David. 2004b. Using a universal access reference model to identify further guidance that belongs in ISO 16071. *Universal Access in the Information Society*, **3**(1), 17–29. (used on p xv, 14, 15, 17, 18, 19, 20, 21, 75, 95, 254, and 255.)

Chandler, Daniel. 1994. *The Transmission Model of Communication*. Available at: http://www.aber.ac.uk/media/Documents/short/trans.html. (used on p 13 and 14.)

Chang, Mei-Mei, & Ho, Chiung-Mei. 2009. Effects of locus of control and learner-control on web-based language learning. *Computer Assisted Language Learning*, **22**(3), 189–206. (used on p 127.)

Charness, N., & Boot, W.R. 2009. Aging and Information Technology Use Potential and Barriers. *Current Directions in Psychological Science*, **18**(5), 253–258. (used on p 29 and 130.)

Chen, G., Kotz, D., *et al.* 2000. *A survey of context-aware mobile computing research*. Tech. rept. Technical Report TR2000-381, Dept. of Computer Science, Dartmouth College. (used on p 19 and 20.)

Cheng, Lung-Pan, Liang, Hsiang-Sheng, Wu, Che-Yang, & Chen, Mike Y. 2013. iGrasp: Grasp-based Adaptive Keyboard for Mobile Devices. *Pages 3037–3046 of: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '13. New York, NY, USA: ACM. (used on p 262.)

Chua, Siew Lian, Chen, Der-Thanq, & Wong, Angela F. L. 1999. Computer anxiety and its correlates: a meta-analysis. *Computers in Human Behavior*, **15**(5), 609 – 623. (used on p 151 and 153.)

Compeau, Deborah R., & Higgins, Christopher A. 1995. Computer self-efficacy: development of a measure and initial test. *MIS Q.*, **19**(2), 189–211. (used on p 40, 128, and 139.)

Crow, Kevin L. 2008. Four Types of Disabilities: Their Impact on Online Learning. *TechTrends*, **52**(1), 51–55. (used on p 24.)

Crystal, Abe, & Ellington, Beth. 2004. Task analysis and human-computer interaction: approaches, techniques, and levels of analysis. *Pages 1–9 of: Proceedings of the Tenth Americas Conference on Information Systems, New York*. Citeseer. (used on p 85, 86, and 269.)

Cullen, Kevin, McAnaney, Donal, Dolphin, Ciaran, Delaney, Sarah, & Stapleton, Philomena. 2012. *Research on the provision of Assistive Technology in Ireland*

and other countries to support independent living across the life cycle. Tech. rept. Work Research Centre, Dublin. (used on p 35 and 36.)

Cunningham, Laura M., Nugent, Chris D., Moore, George, Finlay, Dewar D., & Craig, David. 2012. Computer-based assessment of movement difficulties in Parkinson's disease. *Computer Methods in Biomechanics and Biomedical Engineering*, **15**(10), 1081–1092. (used on p 190 and 236.)

Damodaran, L. 2014. *Sus-IT project publications and selected outputs (2009-2013) http://sus-it.lboro.ac.uk/publications.html*. Tech. rept. Sus-IT project. (used on p 2.)

Damodaran, L., Olphert, C. W., & Sandhu, J. 2012. *A Toolkit for Engaging Older People in Research and Design.* Tech. rept. Loughborough University. (used on p 133, 142, 173, 179, 180, 181, and 183.)

Damodaran, L, Olphert, CW, & Sandhu, J. 2013. Falling Off the Bandwagon Exploring the Challenges to Sustained Digital Engagement by Older People. *Gerontology*, **60**(2), 163–173. (used on p 43, 138, 174, and 195.)

Data Protection Act, 1998. 1998. *HMSO*. London. (used on p 179, 184, 187, and 189.)

DCMI. 2012 (June). *Dublin Core Metadata Terms*. http://dublincore.org/documents/dcmi-terms/ Accessed 19/3/14. (used on p 197 and 269.)

Dickinson, A., Arnott, J., & Prior, S. 2007. Methods for human–computer interaction research with older people. *Behaviour & Information Technology*, **26**(4), 343–352. (used on p 180.)

Dieterich, Hartmut, Malinowski, Uwe, Kühme, Thomas, & Schneider-Hufschmidt, Matthias. 1993. State of the art in adaptive user interfaces. *Human factors in information technology*, **10**, 13–13. (used on p 39 and 41.)

Edlin-White, Robert, D'Cruz, Mirabelle, Floyde, Anne, Johann Riedel, Sue Cobb, Broadley, Sylvia, Marsá, Juan Carlos Virginia Suárez, Hernández, José Alberto, Gacimartín, Carlos, Bruikman, Hester, Peissner, Matthias, Häbe, Dagmar, Olmo, Alberto, Allen, Ian, & Takacs, Barnabas. 2010. *Requirements for User Interface Adaptation.* Tech. rept. MyUI: Mainstreaming Accessibility through Synergistic User Modelling and Adaptability. (used on p 39.)

Edwards, ADN. 2002. Multimodal interaction and people with disabilities. *Multimodality in Language and Speech Systems*, 73–92. (used on p 18.)

Eisma, R., Dickinson, A., Goodman, J., Syme, A., Tiwari, L., & Newell, A.F. 2004. Early user involvement in the development of information technology-related products for older people. *Universal Access in the Information Society*, **3**, 131–140. (used on p 33 and 175.)

EN ISO 8596:2009, International Organization for Standardization. 2009. *Ophthalmic optics – Visual acuity testing – Standard optotype and its presentation.* (used on p 236 and 309.)

Field, Andy. 2009. *Discovering statistics using SPSS*. Sage publications. (used on p 111, 114, 132, 139, 145, 147, 148, 157, 208, and 221.)

Findlater, Leah, & McGrenere, Joanna. 2004. A comparison of static, adaptive, and adaptable menus. *Pages 89–96 of: CHI '04: Proceedings of the SIGCHI conference on Human factors in computing systems*. New York, NY, USA: ACM. (used on p 37.)

Fink, Josef, Kobsa, Alfred, & Nill, Andreas. 1996. User-oriented Adaptivity and Adaptability in the AVANTI Project. *In: Designing for the Web: empirical studies*. (used on p 36 and 44.)

Fitts, Paul M. 1954. The information capacity of the human motor system in controlling the amplitude of movement. *Journal of experimental psychology*, **47**(6), 381. (used on p 140.)

Fleishman, Edwin A., Quaintance, Marilyn K., & Broedling, Laurie A. 1984. *Taxonomies of Human Performance*. Academic Press, INC. (used on p 72.)

Forbes, Paula, Gibson, Lorna, Hanson, Vicki L., Gregor, Peter, & Newell, Alan F. 2009. Dundee user centre: a space where older people and technology meet. *Pages 231–232 of: Proceedings of the 11th international ACM SIGACCESS conference on Computers and accessibility*. Assets '09. New York, NY, USA: ACM. (used on p 42 and 43.)

Fourney, David, & Carter, Jim. 2006a. A Standard Method of Profiling the Accessibility Needs of Computer Users with Vision and Hearing Impairments. *In: Conference & Workshop on Assistive Technologies for People with Vision & Hearing Impairments*. (used on p 53.)

Fourney, David W. 2007. *Using a Common Accessibility Profile to Improve Accessibility*. M.Phil. thesis, University of Saskatchewan, Saskatoon. (used on p xv, xix, 76, and 81.)

Fourney, DW, & Carter, JA. 2006b. *Ergonomic accessibility standards*. Tech. rept. University of Saskatchewan. (used on p 30 and 32.)

Franks, John R. *Hearing measurement*. Retrieved from http://www.who.int/occupational_health/publications/noise8.pdf on 28/04/2014. (used on p 308.)

Freitas, Ricardo. 2009. Scientific Research Methods and Computer Science. *MAP-I Seminars Workshop*. (used on p 58 and 59.)

Gajos, Krzysztof Z., Czerwinski, Mary, Tan, Desney S., & Weld, Daniel S. 2006. Exploring the design space for adaptive graphical user interfaces. *Pages 201–208 of: AVI '06: Proceedings of the working conference on Advanced visual interfaces*. New York, NY, USA: ACM. (used on p 72.)

Ghahramani, Zoubin. 2013. Bayesian non-parametrics and the probabilistic approach to modelling. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, **371**(1984), 20110553. (used on p 208.)

Gilbertson, Teresa D., & Machin, Colin H. C. 2012. Guidelines, icons and marketable skills: an accessibility evaluation of 100 web development company homepages. *Pages 17:1–17:4 of: Proceedings of the International Cross-Disciplinary Conference on Web Accessibility*. W4A '12. New York, NY, USA: ACM. (used on p 32.)

Gitlow, Lynn, Dininno, Destiny, Choate, Lindsay, Luce, Raechell A., & Flecky, Kathleen. 2011. The Provision of Assistive Technology by Occupational Therapists Who Practice in Mental Health. *Occupational Therapy in Mental Health*, **27**(2), 178–190. (used on p 42.)

Godoy, Daniela, & Amandi, Analia. 2005. User profiling in personal information agents: a survey. *The Knowledge Engineering Review*, **20**(11), 329–361. (used on p 262 and 296.)

Golbreich, Christine, Zhang, Songmao, & Bodenreider, Olivier. 2006. The foundational model of anatomy in OWL: Experience and perspectives. *Web Semantics: Science, Services and Agents on the World Wide Web*, **4**(3), 181 – 195. Semantic Web for Life Sciences. (used on p 204.)

Gooding, C. 2000. Disability Discrimination Act: from statute to practice. *Critical Social Policy*, **20**(4), 533–549. (used on p 8.)

Gross, B.M. 2009. Names of our lives. *Pages 747–752 of: Computational Science and Engineering, 2009. CSE'09. International Conference on*, vol. 4. IEEE. (used on p 192.)

Gruber, T.R. et al. 1993. A translation approach to portable ontology specifications. *Knowledge acquisition*, **5**(2), 199–220. (used on p 82.)

Guiard, Yves, Olafsdottir, Halla B., & Perrault, Simon T. 2011. Fitt's Law As an Explicit Time/Error Trade-off. *Pages 1619–1628 of: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '11. New York, NY, USA: ACM. (used on p 101 and 102.)

Hanson, Vicki L. 2011. Technology skill and age: what will be the same 20 years from now? *Universal Access in the Information Society*, **10**(4), 443–452. (used on p 28, 32, 33, and 34.)

Hawthorn, D. 2000. Possible implications of aging for interface designers. *Interacting with computers*, **12**(5), 507–528. (used on p 27 and 134.)

Heckmann, D., Schwartz, T., Brandherm, B., Schmitz, M., & von Wilamowitz-Moellendorff, M. 2005. Gumo–the general user model ontology. *Pages 428–432 of: User modeling 2005*. Springer. (used on p 82 and 189.)

Heinssen, Robert K., Glass, Carol R., & Knight, Luanne A. 1987. Assessing computer anxiety: Development and validation of the Computer Anxiety Rating Scale. *Computers in Human Behavior*, **3**(1), 49 – 59. (used on p 139.)

Henry, John W., & Stone, Robert W. 1995. Computer self-efficacy and outcome expectancy: the effects on the end-user's job satisfaction. *SIGCPR Comput. Pers.*, **16**(4), 15–34. (used on p 131.)

Hertzum, Morten, & Hornbæk, Kasper. 2010. How Age Affects Pointing With Mouse and Touchpad: A Comparison of Young, Adult, and Elderly Users. *International Journal of Human-Computer Interaction*, **26**(7), 703–734. (used on p 100.)

HMSO. 1982. *Supply of Goods and Services Act*. London. (used on p 180.)

HMSO. 2010. *UK Equality Act*. Accessed via http://www.legislation.gov.uk/ukpga/2010/15/contents. (used on p 10 and 11.)

Hoffman, Mark E., & Vance, David R. 2007. Gender difference trends in computer literacy of first-year students. *Pages 405–409 of: SIGCSE '07: Proceedings of* *the 38th SIGCSE technical symposium on Computer science education*. New York, NY, USA: ACM. (used on p 131.)

Hong, W., Thong, J.Y.L., Wong, W.M., & Tam, K.Y. 2002. Determinants of user acceptance of digital libraries: an empirical examination of individual differences and system characteristics. *Journal of Management Information Systems*, **18**(3), 97–124. (used on p 128.)

Horvitz, Eric, & Apacible, Johnson. 2003. Learning and reasoning about interruption. *Pages 20–27 of: ICMI '03: Proceedings of the 5th international conference on Multimodal interfaces*. New York, NY, USA: ACM. (used on p 296.)

Horvitz, Eric, Breese, Jack, Heckerman, David, Hovel, David, & Rommelse, Koos. 1998. The Lumiere Project: Bayesian User Modeling for Inferring the Goals and Needs of Software Users. *In: 14th Conference on Uncertainty in Artificial Intellegence*. (used on p 296.)

Hunter, Audrey, Sayers, Heather, & McDaid, Liam. 2007 (September). An Evolvable Computer Interface for Elderly Users. *In: Workshop "Supporting Human Memory with Interactive Systems"*. BCS. (used on p 29.)

Hurst, Amy. 2009. Automatic assessment and adaptation to real world pointing performance. New York, NY, USA: ACM. (used on p 262.)

Hurst, Amy, Hudson, Scott E., & Mankoff, Jennifer. 2007. Dynamic detection of novice vs. skilled use without a task model. *Pages 271–280 of: CHI '07: Proceedings of the SIGCHI conference on Human factors in computing systems*. New York, NY, USA: ACM. (used on p 297.)

Iglesias-Perez, A. 2010. A context-aware semantic approach for the effective selection of an assistive software. *In: IV International Symposium of Ubiquitous Computing and Ambient Intelligence*. (used on p 14, 79, 90, 252, and 295.)

ISO 226:2003, International Organization for Standardization. 2003. *Acoustics – Normal equal-loudness level countours*. (used on p 246.)

ISO 9241-11, International Organization for Standardization. 1998. *Ergonomics requirements for office work with visual display terminal (VDTs) – Part 11: Guidance on usability*. (used on p 20.)

ISO 9241-171, International Organization for Standardization. 2008. *Ergonomics of Human-System interaction – Part 171: Guidance on software accessibility*. (used on p 7 and 15.)

ISO 9241-20, International Organization for Standardization. 2008. *Ergonomics of Human-System interaction – Part 20: Accessibility guidelines for information/communication technology (ICT) equipment and services.* (used on p 7.)

ISO 9241-400, International Organization for Standardization. 2007. *Ergonomics of Human-System Interaction – Part 400: Principles and requirements for physical input devices.* (used on p 141.)

ISO 9241-9, International Organization for Standardization. 2000. *Requirements for Non-Keyboard Input Devices.* (used on p 236 and 310.)

ISO 9999, International Organization for Standardization. 2011. *Assistive products for persons with disability - Classification and terminology.* (used on p 35 and 36.)

ISO/IEC 24751-1, International Organization for Standardization / International Electrotechnical Commission. 2008. *Information technology – Individualized adaptability and accessibility in e-learning, education and training – Part 1: Framework and reference model.* (used on p 74.)

ISO/IEC 24751-2, International Organization for Standardization / International Electrotechnical Commission. 2008. *Information technology – Individualized adaptability and accessibility in e-learning, education and training – Part 2: "Access for all" personal needs and preferences for digital delivery model.* (used on p 74 and 82.)

ISO/IEC 24751-3, International Organization for Standardization / International Electrotechnical Commission. 2008. *Information technology – Individualized adaptability and accessibility in e-learning, education and training – Part 3: "Access for all" digital resource description.* (used on p 74.)

ISO/IEC 24756, International Organization for Standardization / International Electrotechnical Commission. 2009. *Information technology – Framework for specifying a common access profile (CAP) of needs and capabilities of users, systems, and their environments.* (used on p 75, 81, 118, 269, and 295.)

ISO/IEC 7498-1, International Organization for Standardization / International Electrotechnical Commission. 1994. *Information Technology–Open Systems Interconnection–Basic ReferenceModel: The Basic Model.* (used on p 88.)

ISO/IEC 80000, International Organization for Standardization / International Electrotechnical Commission. 2009. *Quantities and Units.* (used on p 114.)

Jakab, Z, & Wenzel, K. 2004. Detecting tetrachromacy in human subjects. *Perception 33 ECVP Abstract Supplement*. (used on p 25.)

Jefferson, Luke, & Harvey, Richard. 2006. Accommodating Color Blind Computer Users. *Pages 40–47 of: Proceedings of the 8th International ACM SIGACCESS Conference on Computers and Accessibility*. Assets '06. New York, NY, USA: ACM. (used on p 25.)

John, Bonnie. 1998 (June). Cognitive Modeling in Human-Computer Interaction. *Pages 161–167 of: Graphics Interface*. (used on p 45.)

Kaklanis, Nikolaos, Moschonas, Panagiotis, Moustakas, Konstantinos, & Tzovaras, Dimitrios. 2011. A Framework for Automatic Simulated Accessibility Assessment in Virtual Environments. *Pages 302–311 of:* Duffy, Vincent G. (ed), *Digital Human Modeling*. Lecture Notes in Computer Science, vol. 6777. Springer Berlin Heidelberg. (used on p 73.)

Kaklanis, Nikolaos, Moschonas, Panagiotis, Moustakas, Konstantinos, & Tzovaras, Dimitrios. 2012. Virtual user models for the elderly and disabled for automatic simulated accessibility and ergonomy evaluation of designs. *Universal Access in the Information Society*, 1–23. (used on p 295 and 298.)

Kammer, Dietrich, Wojdziak, Jan, Keck, Mandy, Groh, Rainer, & Taranko, Severin. 2010. Towards a formalization of multi-touch gestures. *Pages 49–58 of: ACM International Conference on Interactive Tabletops and Surfaces*. ITS '10. New York, NY, USA: ACM. (used on p 71.)

Karavidas, Maria, Lim, Nicholas K., & Katsikas, Steve L. 2005. The effects of computers on older adult users. *Computers in Human Behavior*, **21**(5), 697 – 711. (used on p 41 and 131.)

Karim, Shuaib, & Tjoa, A.Min. 2006. Towards the Use of Ontologies for Improving User Interaction for People with Special Needs. *Pages 77–84 of:* Miesenberger, Klaus, Klaus, Joachim, Zagler, Wolfgang L., & Karshmer, Arthur I. (eds), *Computers Helping People with Special Needs*. Lecture Notes in Computer Science, vol. 4061. Springer Berlin Heidelberg. (used on p 89.)

Keates, S., & Clarkson, P.J. 2003. Countering design exclusion: bridging the gap between usability and accessibility. *Universal Access in the Information Society*, **2**(3), 215–225. (used on p 32.)

Keates, Simeon, Adams, Ray, Bodine, Cathy, Czaja, Sara, Gordon, Wayne, Gregor, Peter, Hacker, Emily, Hanson, Vicki, Kemp, John, Laff, Mark, Lewis, Clayton, Pieper, Michael, Richards, John, Rose, David, Savidis, Anthony, Schultz, Greg, Snayd, Paul, Trewin, Shari, & Varker, Philip. 2007. Cognitive and learning difficulties and how they affect access to IT systems. *Univers. Access Inf. Soc.*, **5**(4), 329–339. (used on p 26, 34, and 69.)

Keith, Suzette. 2010. Diversity in age: the challenges of reaching the 'hard to reach'. *Pages 34–45 of: UNITECH 2010: International Conference on Universal Technology: proceedings.* Trondheim, Norway: Tapir Academic Press. Conference held on 19-20th May 2010 Oslo Norway. Same paper subsequently presented at 'Research Elders Conference', Middlesex University, June 2010. (used on p 138.)

Kelly, Brian, Sloan, David, Brown, Stephen, Seale, Jane, Petrie, Helen, Lauke, Patrick, & Ball, Simon. 2007. Accessibility 2.0: people, policies and processes. *Pages 138–147 of: Proceedings of the 2007 international cross-disciplinary conference on Web accessibility (W4A).* W4A '07. New York, NY, USA: ACM. (used on p 72.)

Kelly, Diane, & Belkin, Nicholas J. 2002. A user modeling system for personalized interaction and tailored retrieval in interactive IR. *Proceedings of the American Society for Information Science and Technology*, **39**(1), 316–325. (used on p 22.)

Klyne, Graham, Reynolds, Franklin, Woodrow, Chris, Ohto, Hidetaka, Hjelm, Johan, Butler, Mark H, & Tran, Luu. *Composite Capability/Preference Profiles (CC/PP): Structure and Vocabularies 1.0. W3C Recommendation, W3C, January 2004.* (used on p 20.)

Kobsa, Alfred, Koenemann, Jürgen, & Pohl, Wolfgang. 2001. Personalised hypermedia presentation techniques for improving online customer relationships. *Knowl. Eng. Rev.*, **16**(2), 111–155. (used on p 37.)

Kovacs Burns, K., & Gordon, G.L. 2010. Analyzing the impact of disability legislation in Canada and the United States. *Journal of Disability Policy Studies*, **20**(4), 205. (used on p 8.)

Kurniawan, S. H., King, A., Evans, D. G., & Blenkhorn, P. L. 2006. Personalising web page presentation for older people. *Interact. Comput.*, **18**(3), 457–477. (used on p 22.)

Larman, Craig, & Basili, Victor R. 2003. Iterative and Incremental Development: A Brief History. *Computer*, **36**(6), 47–56. (used on p 59.)

Larres, Particia McCourt, Ballantine, Joan A., & Whittington, Mark. 2003. Evaluating the validity of self-assessment: measuring computer literacy amoung entry-level undergraduates within accounting degree programmes at two UK universities. *Accounting Education*, **12**(June), 97–112. (used on p 128 and 129.)

Lenker, J.A., & Paquet, V.L. 2003. A review of conceptual models for assistive technology outcomes research and practice. *Assistive Technology*, **15**(1), 1–15. (used on p 42.)

Li, Yunqiu, Sloan, David, Atkinson, Matthew T., Bell, Matthew J., & Machin, Colin. 2010 (October). Ethical Considerations of How Monitoring Data Is Stored and Used. *Pages 7—8 of: The AAATE Workshop: Assistive Technology — Technlogy Transfer*. (used on p 173, 181, and 316.)

Liffick, Blaise W., & Zoppetti, Gary. 2007. You can take it with you: profile transportability. *Pages 106–110 of: Proceedings of the 12th annual SIGCSE conference on Innovation and technology in computer science education*. ITiCSE '07. New York, NY, USA: ACM. (used on p 70 and 188.)

Loitsch, Claudia, Zimmermann, Gottfried, Stiegler, Andreas, Strobbe, Christophe, Votis, Kostas, Clark, Colin, & Basman, Antranig. 2012. *D204.1. Profile Matching Technical Concept*. Tech. rept. Cloud4All. (used on p 33, 37, 79, 262, and 293.)

Macefield, Ritch. 2009. How to specify the participant group size for usability studies: A practitioner's guide. *Journal of Usability Studies*, **5**(1), 34–45. (used on p 234.)

MacKenzie, I Scott, & Isokoski, Poika. 2008. Fitts' throughput and the speed-accuracy tradeoff. *Pages 1633–1636 of: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM. (used on p 101, 140, and 242.)

MacKenzie, I Scott, Kauppinen, Tatu, & Silfverberg, Miika. 2001. Accuracy measures for evaluating computer pointing devices. *Pages 9–16 of: Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM. (used on p 140, 236, and 310.)

Madrid, Jaisiel, Peinado, Ignacio, & Koutkias, Vassilis. 2012. *D101.1. Cloud4all Priority applications and User Profile Ontology*. Tech. rept. Cloud4All. (used on p 78.)

Marakas, George M., Yi, Mun Y., & Johnson, Richard D. 1998. The Multilevel and Multifaceted Character of Computer Self-Efficacy: Toward Clarification of the Construct and An Integrative Framework for Research. *Info. Sys. Research*, **9**(2), 126–163. (used on p 40.)

Martin, A., Cechich, A., Gordillo, S., & Rossi, G. 2007. A Three-Layered Approach to Model Web Accessibility for Blind Users. *Pages 76–83 of: Web Conference, 2007. LA-WEB 2007. Latin American.* (used on p 50 and 51.)

Masuwa-Morgan, K.R. 2008. Introducing AccessOnto: Ontology for Accessibility Requirements Specification. *Pages 33–38 of: Ontologies in Interactive Systems, 2008. ONTORACT '08. First International Workshop on.* (used on p 89.)

Maurer, Todd J., & Andrews, Kimberly D. 2000. Traditional, Likert, and Simplified Measures of Self-Efficacy. *Educational and Psychological Measurement*, **60**(6), 965–973. (used on p 140.)

McDonald, Sharon, & Petrie, Helen. 2013. The Effect of Global Instructions on Think-aloud Testing. *Pages 2941–2944 of: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* CHI '13. New York, NY, USA: ACM. (used on p 311.)

Messick, S. 1995. Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *The American Psychologist*, **50**(9), 741–749. (used on p xiii, 260, and 313.)

Mitra, Sophie. 2006. The Capability Approach and Disability. *Journal of Disability Policy Studies*, **16**(4), 236–247. (used on p 11 and 12.)

Mohamad, Yehya, Dangelmaier, Manfred, Peissner, Matthias, Biswas, Pradipta, & Velasco, CarlosA. 2012. Virtual User Models for Designing and Using of Inclusive Products: Introduction to the Special Thematic Session. *Pages 284–287 of:* Miesenberger, Klaus, Karshmer, Arthur, Penaz, Petr, & Zagler, Wolfgang (eds), *Computers Helping People with Special Needs.* Lecture Notes in Computer Science, vol. 7382. Springer Berlin Heidelberg. (used on p 293.)

Moller, Aage R. 2006. *Hearing: anatomy, physiology, and disorders of the auditory system.* Academic press. (used on p 251.)

Molnar, A.R. 1997. Computers in Education: A Brief History. *THE Journal*, **24**(11), 63–68. (used on p 134.)

Moores, Trevor T., & Chang, Jerry Cha-Jan. 2009. Self-efficacy, overconfidence, and the negative effect on subsequent performance: A field study. *Inf. Manage.*, **46**(March), 69–76. (used on p 41, 129, 130, 133, and 186.)

Mori, G., Paterno, F., & Santoro, C. 2004. Design and development of multidevice user interfaces through multiple logical descriptions. *Software Engineering, IEEE Transactions on*, **30**(8), 507 – 520. (used on p 71.)

Newell, A.F., & Gregor, P. 2002. Design for older and disabled people – where do we go from here? *Universal Access in the Information Society*, **2**, 3–7. 10.1007/s10209-002-0031-9. (used on p 28.)

Obrenovic, Zeljko, Abascal, Julio, & Starcevic, Dusan. 2007. Universal accessibility as a multimodal design issue. *Commun. ACM*, **50**(5), 83–88. (used on p 18.)

OED, Oxford English Dictionary. Accessed 11/4/14. *"accessibility, n."*. http://www.oed.com. Accessed 11/4/14. (used on p 6.)

OED, Oxford English Dictionary. Accessed 21/3/14. *"abstraction, adj."*. http://www.oed.com. (used on p 83.)

Olphert, C. W., Damodaran, L., & May, A. J. 2005 (August). Towards digital inclusion - engaging older people in the 'digital world'. *In: Accessible Design in the Digital World Conference 2005*. (used on p 28, 29, 41, 42, 131, 150, and 151.)

Oppermann, Reinhard. 1994. Adaptively supported adaptability. *International Journal of Human-Computer Studies*, **40**(3), 455 – 472. (used on p 37.)

Papoulis, Athanasios, & Pillai, S Unnikrishna. 2002. *Probability, random variables, and stochastic processes*. Tata McGraw-Hill Education. (used on p 102.)

Parahoo, Kader. 2006. *Nursing research: principles, process and issues (2nd Edition)*. Palgrave Macmillan. (used on p 180.)

Parasuraman, R. 2000. Designing automation for human use: empirical studies and quantitative models. *Ergonomics*, **43**(7), 931–951. (used on p 44.)

Parsons, Talcott. 1975. The sick role and the role of the physician reconsidered. *The Milbank Memorial Fund Quarterly. Health and Society*, 257–278. (used on p 10.)

Paterno, Fabio, Santoro, Carmen, & Spano, Lucio Davide. 2009. MARIA: A universal, declarative, multiple abstraction-level language for service-oriented applications in ubiquitous environments. *ACM Transactions on Computer-Human Interaction (TOCHI)*, **16**(4), 19. (used on p 71.)

Peissner, Matthias, Dangelmaier, Manfred, Biswas, Pradipta, Mohamad, Yehya, Jung, Christoph, Wolf, Peter, Gonzalez, Mari Feli, & Kaklanis, Nick. 2011 (June). *Interim Report on VUMS Cluster Standardisation*. Tech. rept. VUMS Cluster. (used on p 47 and 66.)

Peissner, Matthias, Häbe, Dagmar, Janssen, Doris, & Sellner, Thomas. 2012a. MyUI: generating accessible user interfaces from multimodal design patterns. *Pages 81–90 of: Proceedings of the 4th ACM SIGCHI symposium on Engineering interactive computing systems*. EICS '12. New York, NY, USA: ACM. (used on p 297.)

Peissner, Matthias, Dangelmaier, Manfred, Biswas, Pradipta, Jung, Yehya Mohamad Christoph, & Kaklanis, Nick. 2012b (June). *White Paper: Virtual User Modelling Public Document*. Tech. rept. 1.0. VUMS Cluster. (used on p 46, 77, 79, 86, and 191.)

Peter, Christian, & Herbon, Antje. 2006. Emotion representation and physiology assignments in digital systems. *Interacting with Computers*, **18**(2), 139 – 170. (used on p 41.)

Petrie, Helen, & Kheir, Omar. 2007. The relationship between accessibility and usability of websites. *Pages 397–406 of: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '07. New York, NY, USA: ACM. (used on p 7, 10, 21, and 22.)

Prentice, Kristen J, Appelbaum, Paul S, Conley, Robert R, & Carpenter, William T. 2007. Maintaining informed consent validity during lengthy research protocols. *IRB: Ethics & Human Research*, 1–6. (used on p 184.)

Project, Sus-IT. 2013. *Project Briefing Papers*. http://sus-it.lboro.ac.uk/SusITBriefingDocs.pdf. (used on p 138.)

Quesenbery, Whitney. 2013. Personas can tell the story behind the model. *Page article 7 of:* Yehya Mohamad, Christos Kouroupetroglou (ed), *W3C WAI RDWG Symposium on User Modeling for Accessibility*. (used on p 69.)

Rao, C.R. 1997. *Statistics and Truth: Putting Chance to Work*. World Scientific. (used on p 293 and 296.)

Razmerita, Liana, Angehrn, Albert, & Maedche, Alexander. 2003. Ontology-Based User Modeling for Knowledge Management Systems. *Pages 213–217 of:* Brusilovsky, Peter, Corbett, Albert, & Rosis, Fiorella (eds), *User Modeling 2003*. Lecture Notes in Computer Science, vol. 2702. Springer Berlin Heidelberg. (used on p 82.)

Razmerita, Liana, Nabeth, Thierry, & Kirchner, Kathrin. 2012. User Modeling and Attention Support: Towards a Framework of Personalization Techniques. *Pages 27–33 of: CENTRIC 2012, The Fifth International Conference on Advances in Human-oriented and Personalized Mechanisms, Technologies, and Services.* (used on p 37 and 39.)

RDWG, & W3C WAI Research and Development Working Group (eds). 2012. *W3C WAI Symposium on User Modeling for Accessibility.* W3C WAI Research and Development Working Group (RDWG) Symposia. W3C Web Accessibility Initiative (WAI). (used on p 50 and 94.)

Rettie, R. 2002. Net generation culture. *Journal of Electronic Commerce Research*, **3**(4), 254–264. (used on p 42 and 134.)

Rijgersberg, Hajo, van Assem, Mark, & Top, Jan. 2013. Ontology of units of measure and related concepts. *Semantic Web*, **4**(1), 3–13. (used on p 114.)

Ritter, Frank E., & Young, R. M. 2001. Embodied models as simulated users: introduction to this special issue on using cognitive models to improve interface design. *International Journal of Human Computer Studies*, **55**(1), 1–14. (used on p 45 and 46.)

Sala, Pilar, Fernandez, Carlos, Mocholí, Juan Bautista, Presencia, Pablo, & Naranjo, Juan Carlos. 2011. Implementation of the ISO/IEC 24756 for the Interaction Modeling of an AAL Space. *Pages 210–219 of:* Stephanidis, Constantine (ed), *Universal Access in Human-Computer Interaction. Context Diversity.* Lecture Notes in Computer Science, vol. 6767. Springer Berlin Heidelberg. (used on p 72, 252, and 295.)

Schaumburg, Heike. 2001. Computers as Tools or as Social Actors? – The Users' Perspective on Antrhopomorphic Agents. *International journal of cooperative information systems*, **10**(01n02), 217–234. (used on p 40.)

Shabajee, P. 2006. Informed Consent on the Semantic Web-Issues for Interaction and Interface Designers. *In: 3rd International Semantic Web User Interaction Workshop.* (used on p 182.)

Shakespeare, Tom. 1996. Disability, identity and difference. *Exploring the divide: Illness and disability*, 94–113. (used on p 24.)

Shannon, Claude E, & Weaver, Warren. 1949. The mathematical theory of communication (Urbana, IL). *University of Illinois Press*, **19**(7), 1. (used on p xv, 13, 14, 15, and 66.)

Shneiderman, Ben. 2000. Universal Usability. *Commun. ACM*, **43**(May), 84–91. (used on p 22, 30, 31, and 36.)

Sieverding, Monika, & Koch, Sabine C. 2009. (Self-)Evaluation of computer competence: How Gender Matters. *Computers & Education*, **52**(3), 696–701. (used on p 128 and 131.)

Sloan, David, & Sayago, Sergio. 2012. *Older people and their use of the internet*. Tech. rept. Nominet Trust. (used on p 28.)

Sloan, David, Atkinson, Matthew Tylee, Machin, Colin, & Li, Yunqiu. 2010. The potential of adaptive interfaces as an accessibility aid for older web users. *Pages 35:1–35:10 of: Proceedings of the 2010 International Cross Disciplinary Conference on Web Accessibility (W4A)*. W4A '10. New York, NY, USA: ACM. (used on p 2, 38, 181, and 256.)

Sosnovsky, Sergey, & Dicheva, Darina. 2010. Ontological technologies for user modelling. *Int. J. Metadata Semant. Ontologies*, **5**(1), 32–71. (used on p 79, 190, 197, and 253.)

Steinfeld, Edward. 2008. *Universal Design/Design for All: Practice and Method*. John Wiley & Sons, Inc. Chap. 44, pages 803–818. (used on p 31.)

Sutcliffe, Alistair, Fickas, Steve, Sohlberg, McKay Moore, & Ehlhardt, Laurie A. 2003. Investigating the usability of assistive user interfaces. *Interacting with Computers*, **15**(4), 577 – 602. (used on p 22.)

Taveira, Alvaro D., & Choi, Sang D. 2009. Review Study of Computer Input Devices and Older Users. *International Journal of Human-Computer Interaction*, **25**(5), 455–474. (used on p 135 and 136.)

Thatcher, Jim, Rutter, Richard, Lauke, Patrick H, Waddell, Cynthia, Henry, Shawn Lawton, Lawson, Bruce, Kirkpatrick, Andrew, Heilmann, Christian, Burks, Michael R, & Regan, Bob. 2006. *Web Accessibility: Web Standards and Regulatory Compliance*. Apress. Accessed via: http://uiaccess.com/understanding.html#others. (used on p 21.)

Trewin, Shari. 2000. Configuration agents, control and privacy. *Pages 9–16 of: CUU '00: Proceedings on the 2000 conference on Universal Usability*. New York, NY, USA: ACM. (used on p 41 and 44.)

Trewin, Shari, Zimmermann, Gottfried, & Vanderheiden, Gregg. 2002. Abstract user interface representations: how well do they support universal access? *SIGCAPH Comput. Phys. Handicap.*, June, 77–84. (used on p 38.)

Trewin, Shari, Keates, Simeon, & Moffatt, Karyn. 2006. Developing steady clicks: a method of cursor assistance for people with motor impairments. *Pages 26–33 of: Assets '06: Proceedings of the 8th international ACM SIGACCESS conference on Computers and accessibility*. New York, NY, USA: ACM. (used on p 15 and 262.)

(UN), United Nations. 2006. *World Population Aging 1950–2050*. Tech. rept. United Nations. http://www.un.org/esa/population/publications/worldageing19502050/. (used on p 27.)

Vanderheiden, Gregg, & Treviranus, Jutta. 2011a. Creating a Global Public Inclusive Infrastructure. *Pages 517–526 of:* Stephanidis, Constantine (ed), *Universal Access in Human-Computer Interaction. Design for All and eInclusion*. Lecture Notes in Computer Science, vol. 6765. Springer Berlin Heidelberg. (used on p 47 and 262.)

Vanderheiden, Gregg, & Treviranus, Jutta. 2011b. Creating a Global Public Inclusive Infrastructure. *Pages 517–526 of:* Stephanidis, Constantine (ed), *Universal Access in Human-Computer Interaction. Design for All and eInclusion*. Lecture Notes in Computer Science, vol. 6765. Springer Berlin Heidelberg. (used on p 188 and 194.)

Vanderheiden, Gregg C. 2008. Ubiquitous Accessibility, Common Technology Core, and Micro Assistive Technology: Commentary on "Computers and People with Disabilities". *ACM Trans. Access. Comput.*, **1**(2), 10:1–10:7. (used on p 177.)

Vanderheiden, Gregg C., Treviranus, Jutta, Gemou, Maria, Bekiaris, Evangelos, Markus, Kasper, Clark, Colin, & Basman, Antranig. 2013. The Evolving Global Public Inclusive Infrastructure (GPII). *Pages 107–116 of:* Stephanidis, Constantine, & Antona, Margherita (eds), *Universal Access in Human-Computer Interaction. Design Methods, Tools, and Interaction Techniques for eInclusion*. Lecture Notes in Computer Science, vol. 8009. Springer Berlin Heidelberg. (used on p xv, 46, and 48.)

Vigo, Markel, Kobsa, Alfred, Arrue, Myriam, & Abascal, Julio. 2007. User-tailored web accessibility evaluations. *Pages 95–104 of: Proceedings of the eighteenth conference on Hypertext and hypermedia*. HT '07. New York, NY, USA: ACM. (used on p 33.)

W3C, World Wide Web Consortium. 2008. *Web Content Accessibility Guidelines 2.0 (WCAG 2.0)*. (used on p 8.)

Wagner, Nicole, Hassanein, Khaled, & Head, Milena. 2010. Computer use by older adults: A multi-disciplinary review. *Computers in Human Behavior*, **26**(5), 870 – 882. Advancing Educational Research on Computer-supported Collaborative Learning (CSCL) through the use of gStudy CSCL Tools. (used on p 27, 28, 131, 151, 153, and 164.)

Waller, Annalu. 2013. Public Policy Issues in Augmentative and Alternative Communication Technologies a Comparison of the U.K. And the U.S. *interactions*, **20**(3), 68–75. (used on p 180.)

Weaver, W. 1949. Recent contributions to the mathematical theory of communication. *The mathematical theory of communication*, 1–28. (used on p 14 and 120.)

WHO, World Health Organisation. 2013 (Sept). *Fact Sheet No. 334*. http://www.who.int/mediacentre/factsheets/fs334/en/ – Accessed 5th May 2014. (used on p 153.)

Wold, Erling, Blum, Thom, Keislar, Douglas, & Wheaten, James. 1996. Content-based classification, search, and retrieval of audio. *MultiMedia, IEEE*, **3**(3), 27–36. (used on p 101.)

Wolf, Peter, Strnad, Oliver, Haller, Heiko, Schmidt, Andreas, Peissner, Matthias, & Hernandez, Jose Alberto. 2011. *Context Ontology, User Modelling Concept and Context Management Architecture*. Tech. rept. MyUI. (used on p 78, 79, and 89.)

Yesilada, Yeliz, Brajnik, Giorgio, & Harper, Simon. 2009. How much does expertise matter?: a barrier walkthrough study with experts and non-experts. *Pages 203–210 of: Proceedings of the 11th international ACM SIGACCESS conference on Computers and accessibility*. Assets '09. New York, NY, USA: ACM. (used on p 42 and 69.)

Appendix A

Inference

Transformations between hierarchical levels take place using a process called inference. Inference is the combination of evidence (information) with reasoning (transformation rules) to come to a conclusion. Inference can move in two directions (Rao, 1997):

Deductive (Top-Down) inference involves using general statements to draw specific conclusions.

Inductive and Abductive (Bottom-Up) inference involves using specific examples to reach a general conclusion.

The two directions can be discussed in terms of the relationship between the two types of model identified in Brusilovsky & Millán (2007): stereotype- and feature-based. With its top-down approach, deductive inference describes the way that stereotype-based models can be used to predict the individual features of a user within a stereotypical group. The bottom-up approach used by induction and abduction has parallels with the way that feature-based modelling can be used to take an individual's characteristics and identify their inclusion within a stereotypical group. Finally it is possible for both forms of inference to be combined. Induction can be used to match existing features to a stereotype from which other expected features can be predicted. Alternately deduction can be used to decompose a stereotype into its component features from which can be used to induce similar stereotypes.

Inference has been widely used in online recommendation and e-learning systems which combined a user profile and inference engine to make suggestions (Mohamad *et al.*, 2012). The matchmaking work-package of the GPII project (Loitsch *et al.*, 2012) is focused on the use of inference to match between the needs and abilities of users, technology and accessibility solutions. It identifies two generic scenarios where inference can be used in matchmaking:

Inferring a preference for a target context: Where a new device or environment is encountered the user's existing settings can be used to infer their preferred settings on the new device.

Responding to a user action by recommending new preferences: Where a user initiates a change in their settings, the information can be used to infer other settings that may be of use to the user.

In response to the general scenarios, four potential approaches are identified which fall into two categories:

Firstly translation between application-unique and common preferences can be achieved through the use of either top-down or bottom-up inference. Translation from specific to generic preferences equates to moving from a low-level representation to a higher-level one. Inversely, translation from generic to specific preferences requires a move down the hierarchy.

Secondly, rules can be employed to provide matchmaking and selection of solutions based on semantic similarities. Simple matchmaking would use both of the forms of inferencing. By using deduction to describe the user's existing preferences and solutions in terms of their effect on lower-level interaction-focused abilities, induction can then be used to identify semantic links with potential new solutions. Semantic information can then be exploited to decide on the best solution amongst semantically similar proposals.

A.1 Deductive (Top-Down) Inference

Deductive inference moves from general statements to specific conclusions. In terms of the actor models discussed above, this involves using higher-level stereotypes to predict the existence of lower-level characteristics. For example, where there is insufficient data available to make a decision, stereotypes can be used to provide the data necessary by using existing data that is similar.

In the context of matchmaking, deduction alone is equivalent to using the stereotypes themselves as features are provided by the stereotypes. More useful is the use of deduction coupled with the induction of the features to allow comparison of higher-level models based on underlying features.

This form of reasoning is widely used in modelling for prediction. Models use a corpus of collected data to create a mathematical representation of expected performance. The Model Human Processor Card *et al.* (1986) provides an early example of this kind of model, by predicting performance based on a general description of a task and an estimated time for the completion of each element thereof. More modern variants allow the actions of different groups to be predicted, through the decomposition of higher-level tasks into lower-level constituent actions (Kaklanis *et al.*, 2012; Iglesias-Perez, 2010; Sala *et al.*, 2011).

The VAALID project (Sala *et al.*, 2011) provides evidence of the practical implementation of ISO 24756 and the hierarchical representation of actors in terms of their underlying abilities. The developed application takes the form of an integrated development environment that requires an expert user to build situations based on a user, system and environment model (assistive technologies are not considered). The IDE requires an expert user and is focused on the modelling of stereotypical situations.

The INREDIS project Iglesias-Perez (2010) uses a similar approach to the CAP structure to provide universal remote control services to a variety of devices. Due to criticism of the way abilities are provided as external services in ISO/IEC 24756 (2009) the model focuses on the use of transmitters and receivers with abilities to transmit and receive data being matched across a channel of communication. By deconstructing users and target devices into their abilities to transmit and receive data via a central controller, interaction is made accessible through the provision of services that are able to deliver the transformations necessary. The INREDIS project is focused on the matching of user-to-device and as such simply indicates the applicability of the approach to individual users without testing the implementation.

The VERITAS model (Kaklanis *et al.*, 2012) describes interaction in terms of a task model which is related to more complex abstract actions, such as driving. Abstract actions are then broken down into simpler tasks (steering) and primitive tasks (grasping). Within the VERITAS model, primitive tasks are the only common reference between the different models and the only tasks that are implemented bio-mechanically. With higher-level tasks dependent on primitives, any combination of primitives can be supported without extra implementation effort. Different virtual user models are then be used to represent the effects of disabilities on the primitive tasks, with induction used to infer success or failure at the high-level task.

The issue with deduction is linked to the use of stereotypes as average profiles which may not represent the abilities of the actual user (or device) in their current context. The VERITAS project is restricted by its reliance on bio-mechanical models which require the accurate collection of data. The proposed use of stereotypical disability models will not provide the data necessary to provide prediction for individuals.

A.2 Bottom-Up Inference

Abductive and Inductive inference both move from specific statements to general conclusion. In terms of the actor models mentioned above, bottom-up inference involves using low-level characteristics to predict conformance to one or more higher-level stereotypes. The difference between them lies in the distinction between reasoning with probabilities based on observational data (induction) and reasoning to the best explanation by intuition without base data (abduction) (Rao, 1997, 57). Both forms of reasoning are used in machine learning techniques (Godoy & Amandi, 2005) and have different uses in the terms of user modelling.

Induction has already been alluded to above. When comparing between two higher-level constructs, their underlying features can be deduced. The comparison of the features to determine similarity (where the features are not matched exactly) is induction. It is also possible to induce the similarity of a set of low-level features to a higher-level construct where the features have been identified separately. This form of induction is used in supervised learning and is often used by online recommender systems that take a user's preferences (e.g. films watched) and match them to a stereotyped profile. The stereotype can then be used to deduce other preferences (e.g. films in the same genre) that are likely to suit the user.

Abduction involves the clustering of features into stereotypes, without having the stereotype for reference. A set of features can be clustered to determine the similarities between them with the resulting groups being used in the creation of stereotypes. This form of reasoning is called unsupervised learning and is suited (according to Godoy & Amandi (2005)) to the categorisation of user information interests.

In terms of accessibility modelling, bottom-up reasoning techniques are used for the collection and categorisation of data.

The Lumiere Project (Horvitz *et al.*, 1998) used Bayesian user models to predict the problems that a user was experiencing based on a series of features. As well as the construction of the Bayesian Model, the project faced challenges including: the generation of a stream of features, transformation of the stream into variables for use in the model, development of persistent profiles to store the information generated and a lack of an architecture for an intelligent user interface. The project was shipped with Microsoft Excel [1] in 1997. Its success was hampered by the inappropriate interruption of users, and so the technique was adapted to predict the expected cost of interrupting the user (Horvitz & Apacible, 2003). The approach collected information by observing a series of low-level features such as

[1] Now part of the office suite: http://office.microsoft.com/en-gb/excel/

whether the user was paying attention to the screen, if they were typing, using the mouse, which application they were using etc.

Hurst *et al.* (2007) tackled a similar problem—the detection of novice vs. skilled use of a graphical user interface— through the use of decision trees. A number of features were chosen including low-level mouse motion characteristics (e.g. time taken to complete an action, dwell time), interaction technique (e.g. Number of opened submenues, dwell time/menues visited) and performance models (e.g. difference between KLM[2] prediction and actual time taken for an action). The combination and nature of the features identified provided an indication of the higher-level construct of skill. The features are usable across a number of applications as they are not application specific, they are however tied to the use of indirect manipulation devices.

Both of the above are examples of the use of bottom-up approaches to provide data, either relating to the task that the user is trying to complete or the skill of the use. The bottom-up approach has also been used to match the suitability of interfaces to the abilities of the user. As with the top-down approach, once adherence to a stereotype has been inferred, actions can be taken based on the stereotype including the deduction of other lower-level features.

(Biswas & Robinson, 2013) describes how user characteristics including expertise, usage time, motor and sensory impairments and user-interest can all influence the variance in task completion time. The developed model deals with motor impairments, measured via grip strength. A simulator was built using the model and used within the GUIDE project to investigate interaction problems of people with a motor impairment using a pointing device and sensory problems in terms of visual acuity and colour blindness. The simulator was able to predict task completion times for 'able-bodied', visually impaired' and 'motor-impaired' participants. The information produced by this study can be used to describe the applicability of an interface for an individual and as a result interfaces can be selected based on the abilities of the user.

The MyUI model (Peissner *et al.*, 2012a) uses an approach based on the use of patterns with different levels of abstractions. User actions are matched to existing patterns which are used to create and adapt an interface specifically for them. High-level 'device-specific' and 'individualisation' patterns describe the functionality of the intended interface and the needs of the user in their current context. Lower-level 'interaction' patterns are then applied that fit both of the higher-level patterns after which user interface components can be selected that are suitable given the interaction-patterns selected. Finally 'adaptation rendering' patterns are chosen to inform the user of the change in the interface and 'adaptation dia-

[2]See section 4.2.1.1

logue' patterns define the degree of user notification appropriate when a change is executed.

A.3 Relevance to the CAP

Examples of the use of hierarchical layering of data within accessibility modelling have been presented. Their advantages stem from an ability to use inference to move between layers, either demonstrating adherence of underlying features to a stereotype or using stereotypes to provide extra information.

The CAP used a fixed four-level structure (as seen in 4.10) to provide comparison between actors based on their underlying interaction-focused abilities. Although this provides a hierarchical structure, each of the levels is semantically different. The top level represents the overall interaction model, the second level represents the actors within the model, the third level represents the task-level capabilities of the actors and the fourth represents the interaction-level properties of the capabilities.

While this is a valid approach it is restrictive as it only allows comparison at a single hierarchical level (the lowest one). In reality actors themselves may be described at various levels of abstraction. The tasks they are able to complete may also be decomposed into sub-tasks which may be decomposed themselves as demonstrated by Kaklanis *et al.* (2012). While Kaklanis *et al.* (2012) advocates comparison at a fixed (bottom) level based on a top-down approach, the approach is dependent on the availability of data to describe those primitive tasks. A number of approaches have been described which suggest that a bottom-up combination of features can be used to provide data, however this implies the existence of at least one level below the bottom level that has been advocated.

Appendix B

Self Efficacy Study

B.1 Instruments

B.1.1 Self Efficacy Questionnaires

Participant_________________

Answer all the questions by circling a number on the 1-10 scale
where: 1= "Not at All Confident" and 10= "Highly Confident"

How well do you think you can use a mouse?

1 2 3 4 5 6 7 8 9 10

How well do you think you could play a game that was based on using a mouse?

1 2 3 4 5 6 7 8 9 10

How much would you agree with the statement – "The mouse is my favourite way to input to a computer"?

1 2 3 4 5 6 7 8 9 10

Figure B.1: Predictive Self-Efficacy Judgement Sheet

B.1.2 Computer Anxiety Questionnaire

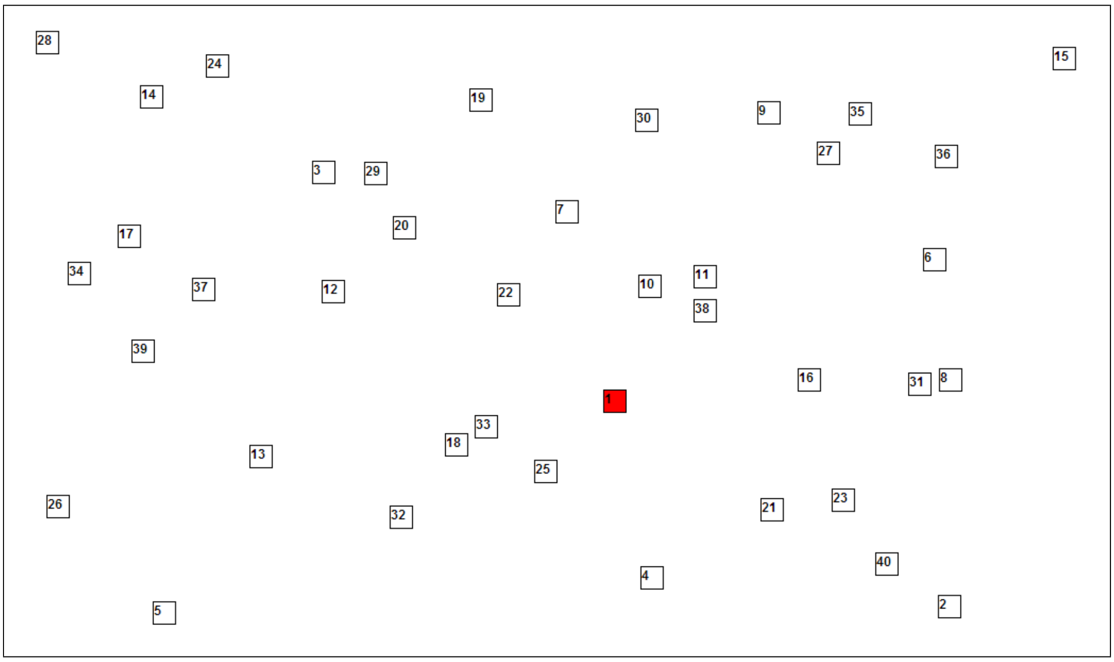
COMPUTER ANXIETY RATING SCALE

INSTRUCTIONS: Please read each item below and respond to it by choosing one of the responses on the scale from (1) to (5), where (1) = strongly disagree and (5) = strongly agree. Do not write in numbers between these choices, only numbers 1, 2, 3, 4, or 5 are options.

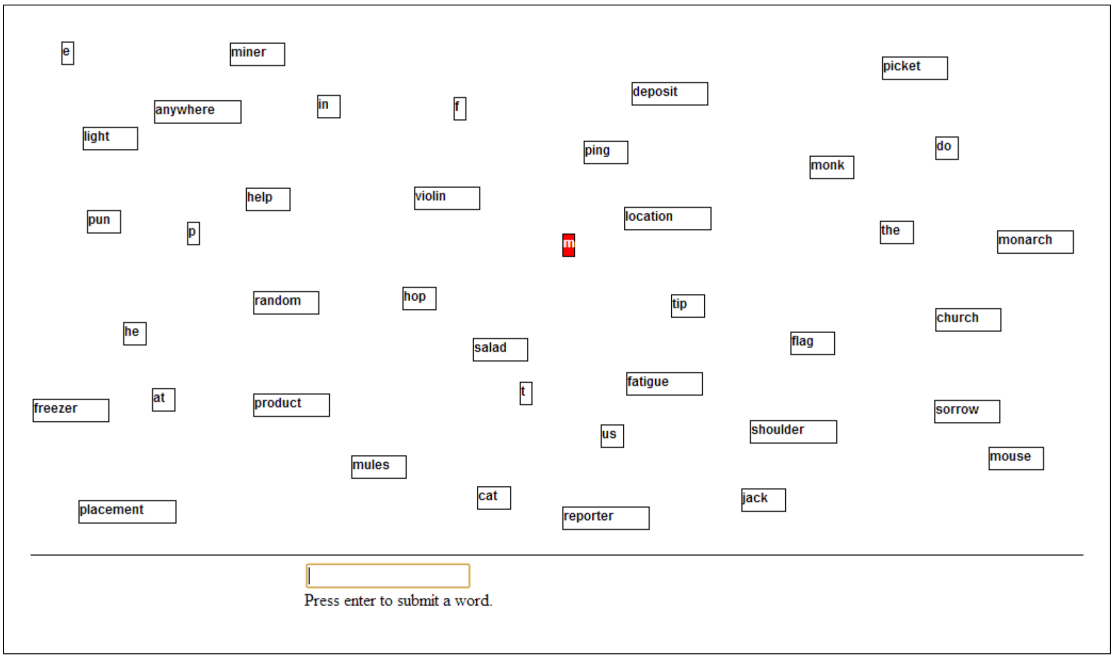
	Strongly disagree ... *Strongly agree*
I feel insecure about my ability to interpret and use a new computer application.	(1) (2) (3) (4) (5)
I look forward to using a computer.	(1) (2) (3) (4) (5)
I do not think I would be able to learn a computer programming language.	(1) (2) (3) (4) (5)
The challenge of learning computers is exciting.	(1) (2) (3) (4) (5)
I am confident I can learn computer skills.	(1) (2) (3) (4) (5)
Anyone can learn to use a computer if they are patient and motivated.	(1) (2) (3) (4) (5)
Learning to operate a computer is like learning any new skill --- the more you practice the better you become.	(1) (2) (3) (4) (5)
I am afraid that if I begin to use computers I will become dependent on them and lose some of my reasoning skills.	(1) (2) (3) (4) (5)
I am sure that with time and practice I will be as comfortable working with computers as I am working with a typewriter/basic word processing software.	(1) (2) (3) (4) (5)
I feel that I will be able to keep up with the advances happening in the computer field.	(1) (2) (3) (4) (5)
I dislike working with machines that are smarter than I am.	(1) (2) (3) (4) (5)
I feel apprehensive about using computers.	(1) (2) (3) (4) (5)
I have difficulty in understanding how a computer works.	(1) (2) (3) (4) (5)
It scares me to think that I could cause the computer to destroy a large amount of information by hitting the wrong key.	(1) (2) (3) (4) (5)
I hesitate to use a computer for fear of making mistakes that I cannot correct.	(1) (2) (3) (4) (5)
You have to be a genius to understand all the special commands contained in most computer software.	(1) (2) (3) (4) (5)
If given the opportunity I would like to learn about and use computers.	(1) (2) (3) (4) (5)
I have avoided computers because they are unfamiliar and somewhat intimidating to me	(1) (2) (3) (4) (5)
I feel computers are necessary tools in both educational and work settings.	(1) (2) (3) (4) (5)

Figure B.2: Computer Anxiety Rating Scale

B.2 Objective Task



(a) Mouse, Mousepad and Touchscreen Game



(b) Keyboard Game

Figure B.3: Example Games

B.2.1 Recorded Data

The recorded data was used to produce measures of both the amount of time taken for a participant to complete the task and number of errors made through clicking away from the intended target. The amount of data collected however demonstrated the potential for a greater level of analysis to be performed.

B.2.1.1 Keyboard Task

- Time of Action – Absolute time.
- Key Pressed – ASCII numeric code, '+' for correct submission and '-' for incorrect submission.

B.2.1.2 Touchscreen Task

- Time of Action – Absolute time.
- X co-ordinate.
- Y co-ordinate.
- 'c+' for correct submission and 'c-' for incorrect submission.

B.2.1.3 Mouse and Touchpad Task

- Time of Action – Absolute time.
- X co-ordinate.
- Y co-ordinate.
- 'm' for a cursor move, 'd' for mousedown event, 'u' for mouseup event, 'c+' for correct submission and 'c-' for incorrect submission.

B.3 Test Procedure

Welcome - Informed Consent: Participant is welcomed by the investigator and asked to sit at the table where the study will take place. Informed consent is acquired through the presentation of a verbal and written description of the nature of the study using the Participant Information Sheet.

Questionnaires: The Sus-IT Digital Inclusion Questionnaire is verbally administered by the investigator after which the Computer Anxiety Rating Scale is administered.

Ability Testing: The order in which devices are used is pre-determined to avoid learning bias with self-efficacy questionnaires arranged according to the order in which devices are to be used. The procedure for each device follows a common format:

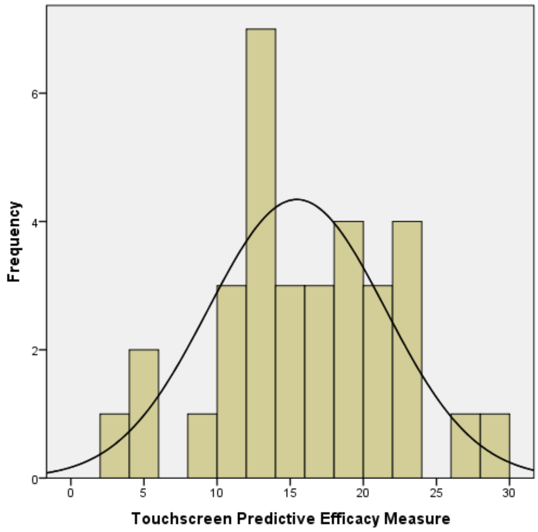
- Introduce the device and demonstrate the game.
- Administer the predictive questionnaire.
- Ask the participant to use the practise game.
- Ask the participant to use the real game.
- Administer the evaluative questionnaire.

Debrief: Thank the participant for their participation and ask if they have any questions. Ensure they have a copy of investigator contact details and ask them no to speak to other potential participants about the study until after they have taken part.

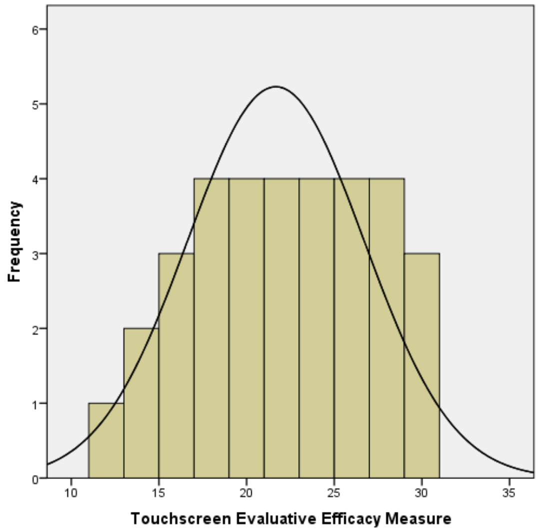
B.4 Additional Results Tables and Subjective Measure Histograms

Table B.1: Records Excluded Due to Data Loss

ID	Group	T	M	P	K
18	2	X	X		
19	2	X	X		
20	2			X	
21	2	X	X		
27	2			X	
32	3			X	

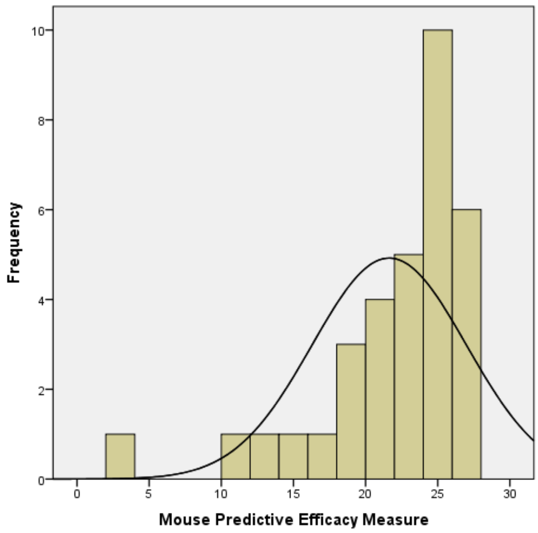


(a) Predictive

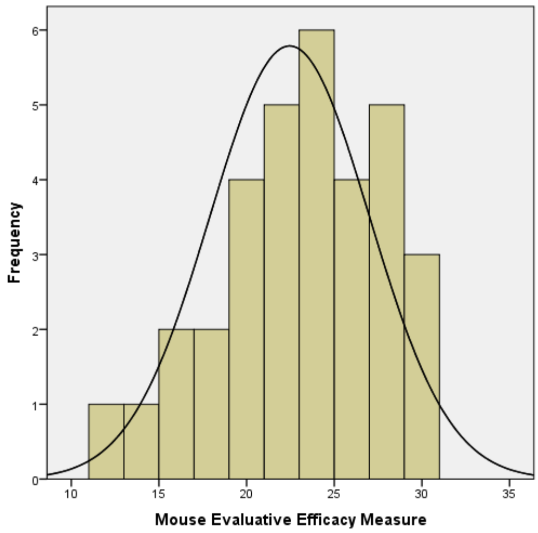


(b) Evaluative

Figure B.4: Touchscreen Self-Efficacy Judgement Histograms

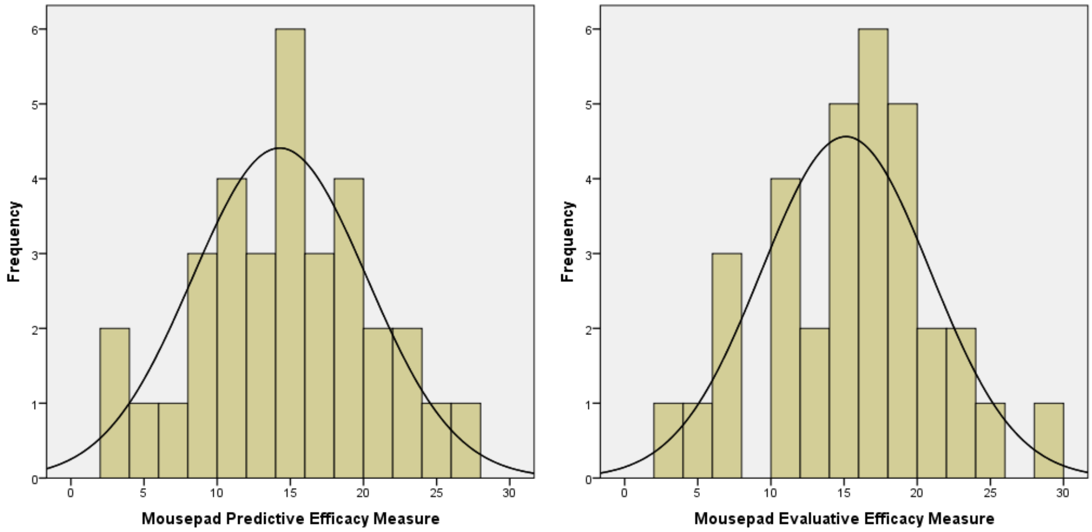


(a) Predictive



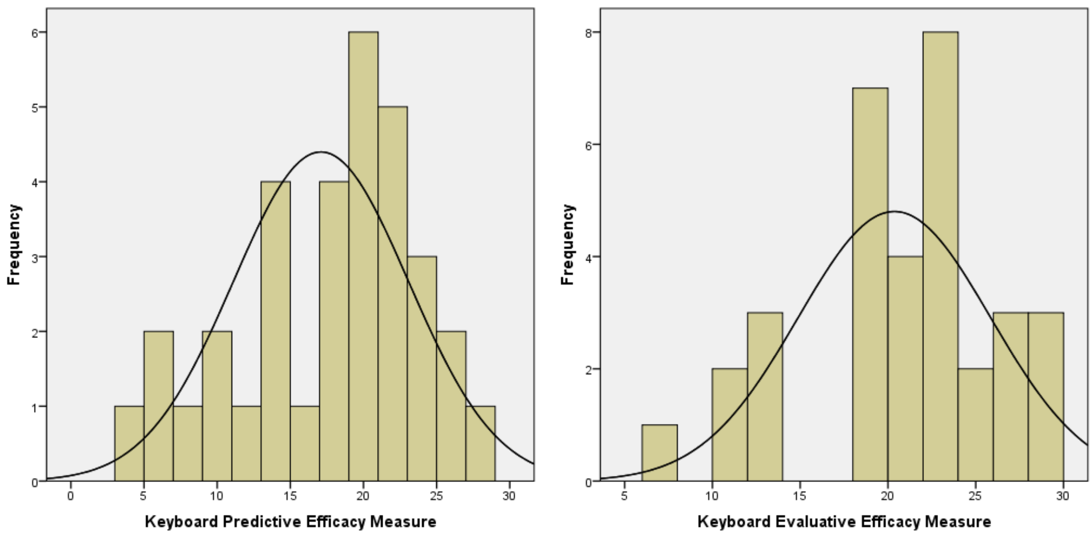
(b) Evaluative

Figure B.5: Mouse Self-Efficacy Judgement Histograms



(a) Predictive (b) Evaluative

Figure B.6: Mousepad Self-Efficacy Judgement Histograms



(a) Predictive (b) Evaluative

Figure B.7: Keyboard Self-Efficacy Judgement Histograms

Table B.2: Changes in Descriptive Statistics as a Result of Removing Records

		n=33	n=39
Age	Min	52	52
	Max	88	88
	Mean	66.15	67.1
	SD	10.6	10.6
Computer Anxiety	Min	21	21
	Max	66	66
	Mean	44.06	45.18
	SD	10.761	11.128
Technology Inclusion	Min	3	3
	Max	30	30
	Mean	19.52	18.79
	SD	6.413	6.546

Table B.3: Shapiro-Wilk Test of Normality for Subjective Measures

	Statistic	df	Sig.
Touchscreen Predictive	.978	33	.711
Touchscreen Evaluative	.960	33	.259
Mouse Predictive	.821	33	.000
Mouse Evaluative	.951	33	.140
Mousepad Predictive	.989	33	.975
Mousepad Evaluative	.986	33	.944
Keyboard Predictive	.953	33	.162
Keyboard Evaluative	.929	33	.033

Table B.4: Shapiro-Wilk Test of Normality for Objective Measures

	Statistic	df	Sig.
Touchscreen Objective Measure	.729	33	.000
T_O_Lg	.901	33	.006
Mouse Objective Measure	.497	33	.000
M_O_Lg	.835	33	.000
Mousepad Objective Measure	.641	33	.000
P_O_Lg	.941	33	.072
Keyboard Objective Measure	.839	33	.000
K_O_Lg	.948	33	.119

Appendix C

Evaluation Study Materials

C.1 Data Collection Study

C.1.1 Selection Criteria

Nine participants between the ages of 61 and 82 ($\bar{x} = 70.3$, $s.d. = 6.5$) were recruited from the Loughborough area to take part in cross-sectional data collection sessions. Participants were recruited through contacts at the Charnwood U3A (University of the Third Age) and the Leslie Edwards Trust[1] (a local charity focusing on the provision of lip-reading classes to aid people with hearing impairments). Participants were given a small food-based reward after completing the study, however this was not however advertised beforehand in order to avoid biasing recruitment. The participants exhibited a range of impairments, and many used a hearing aid and/or glasses. Participation was restricted to people that were regular computer users and all participants used some form of ICT every day. As with the study described in chapter 6, participants were allowed to use any 'assistive technologies' that they used when interacting with technology at home.

C.1.2 Procedure

Data collection took the form of single sessions that were completed in the presence of a single investigator. Sessions were conducted in a room that allowed environmental conditions to be controlled and were conducted on a one-to-one basis in order to avoid distractions. On arrival, informed consent was collected from participants and then a short introductory interview was conducted to allow both time for acclimatisation and the collection of additional data that would be used to inform the analysis of the result. Sessions were conducted based on a prescribed

[1] http://leslieedwardstrust.btck.co.uk/

procedure (detailed in appendix C) in order to avoid any variance between data collection sessions.

The order in which mini-games were performed was controlled in order to reduce the variance potentially caused by learning effects within a session. After the introductory interview, the hearing mini-game was performed as it required the lowest cognitive and motor skills with a single button (the enter key) to be pressed. The vision mini-games were performed second, one of which required increased co-ordination in the form of choosing between the two arrow keys. Finally the fine motor dexterity mini-game required the use of a mouse and therefore the highest level of hand-eye co-ordination. By gradually increasing the motor requirement, participants were given an opportunity to gradually perform more taxing motor tasks, allowing them to perform at the best in the final mini-game.

Once all of the mini-games were completed, the evaluation task described at the start of the session was re-described and attempted in four variations.

C.1.3 Introductory Interview

On arrival, a structured interview was used to acquire data about participants' personal awareness with regards to the task they were about to complete. Given the quantitative approach taken in the self-efficacy study demonstrated a general inability to predict performance, a qualitative approach was used with three questions being posed. After being presented with a description of the task they were about to complete (in writing that was read aloud) participants were asked:

1. To describe any problems they might have in completing the task.
2. Whether they considered themselves to have any disabilities.
3. What technology experience they had.

C.1.4 Mini-Games

This section describes the construction of each of the mini-games. The mini-games are described in terms of: (1) the capability that they assess, (2) the method of assessment, (3) the standardised test on which they were based and (4) the data they produced.

Audio Mini-Game

While standards exist for the measurement of hearing they are often intended to provide a level of accuracy that is unobtainable given the resources and environment that the mini-game will be operating within (Franks, n.d.). A number of elements will however be used to create a test that is suitable for use. Five tones were used to cover a range of 200–12,000Hz, based on available MP3 files[2] in order to produce a low-resolution audio-gram that was stored as a series of frequency-focused capabilities. The accuracy of the audio files and the resulting tones they produced through the available hardware was not verified. However as the same files and hardware were used for both the mini-games and evaluation tasks verification was not necessary. Depending on the desired transportability of data, future use of this mini-game could be preceded by a calibration, performed with appropriate equipment.

The audio mini-game was constituted from volume and frequency elements and took the form of a number of tones with different frequencies that were played at gradually decreasing volumes until the participant could no longer hear them. Tones were played sequentially at random intervals and participants had to press the enter key within one second of the tone being played. Participants were instructed to press the key as soon as they could *comfortably* hear a tone. The use of a relatively short time period reduced the potential for participants to either guess when the tone would play or retrospectively decide that they had heard a tone, increasing the likelihood that they had heard it and therefore the validity of the assessment.

Visual Mini-Games

Vision was measured in terms of visual acuity and contrast sensitivity and three mini-games were created: (1) based on the Snellen visual acuity test (BS 4274-1:2003, 2003) (2) based on a preference based measure used to assess visual acuity in Atkinson (2012) and (3) a visual search task based on the Landolt C Test (EN ISO 8596:2009, 2009). Before completing the mini-games, participants asked to sit comfortably and reminded to avoid leaning forward (providing a more objective measure of their ability).

The Snellen-based visual acuity test used sequences of six letters which the participant had to identify. As the test progressed, the letters got gradually harder to read; either becoming (1) smaller, (2) lighter or (3) a combination of both. As with the official Snellen test, a participant's score was based on the most difficult sequence from which they could correctly identified five out of the six letters. However rather than using the standard Sloan font the same san-serif font used in the final evaluative task was used, in order to improve the transportability of the results.

As a follow-up exercise participants were presented with three lists of word that

[2] 200, 1000, 5000, 10000, 12000

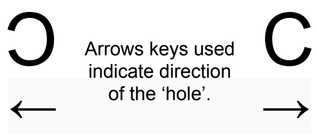


Figure C.1: Visual Mini-Games

got gradually smaller and/or lighter, with the same values as the previous Snellen-based test. This approach was used in a bootstrapping mini-game developed in Atkinson (2012) and participants were asked to choose the smallest word that they could comfortably read.

The visual search task randomly presented either a capital C or its reverse (as displayed in figure C.1) on the screen. The participant had to acquire and accurately recognise the shapes by pressing the arrow button corresponding to the side on which the hole appeared. As with the Snellen-based test three variations were created with letter becoming either: (1) smaller, (2) lighter or (3) a combination of both. Although no time limit was enforced and the test advanced at the speed at which each participant made a selection, the test was assessed similar to the audio test. Participants were given a score that reflected the lowest size and contrast that they were able to perceive, with a maximum average time of less than a second and no more than one mistake out of five attempts.

Motor Mini-Game

The motor-based mini-game provided a measure of participants' fine-motor skills in terms of their arm/wrist dexterity and finger dexterity. A joint movement and double-click test was created based on the multi-directional point-select task defined in ISO 9241-9 (2000) as used in MacKenzie *et al.* (2001). Six circular targets were arranged in a circular layout; both the diameter of the targets themselves and the diameter of the layout circle that the targets were arranged in was adaptable.

Participants were presented with a series of tests consisting of targets of different sizes and distances away from each other. Through assessing the speed and accuracy with which the mouse cursor was moved between the targets, the participant's wrist dexterity was scored. The double click speed was then used as a measure of figure dexterity.

Evaluation Task

In order to allow the framework to produce a functional assessment a computer-based tasks was created based on the one developed for the self-efficacy study in chapter 6. It required the use of all of the previously tested skills and was

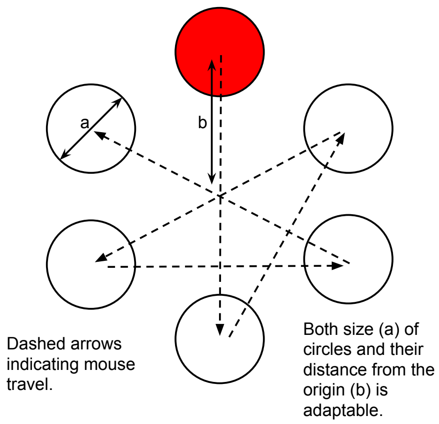


Figure C.2: Motor Skills Mini-Game

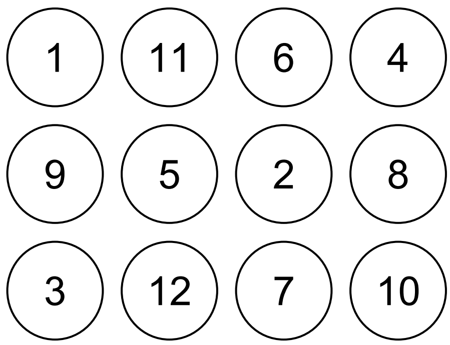


Figure C.3: Example Arrangements of the Evaluation Task

intended to simulate a complex, multi-modal task which was made up of a series of different sub-tasks. Twelve targets were randomly placed within a grid (as seen in figure C.3). The participant was required to select each target in turn by double clicking on it, after which a sound would play to acknowledge that it had been clicked and the participant could select the next target.

The task was fully adaptable, having the same adaptations as were present in the mini-games: Size of targets, distance between targets, double-click speed, label size, label contrast, tone volume and tone frequency.

Before they began the task, participants were given a standard set of instructions that had been assessed to ensure clarity and coherence by two independent experts. The instructions were presented both aurally and in written form (detailed in Appendix C and then given chance to practise on a version of the task engineered to ensure completion.

Despite its potential to provide useful data, the use of a concurrent-verbalisation technique (thinking-aloud) could have placed an extra cognitive load on participants, potentially interfering with their ability to complete the task (McDonald & Petrie, 2013). For this reason participants were not interrupted unless they were having difficulty, or could not complete a task. If a task had ended prematurely, participants were then asked to describe the problem that they had encountered.

C.1.5 Participant Instructions

Informed Consent

Before we start, have you seen the Participant Information Sheet? This study is looking at ways of matching people to adaptations that will help them when they are using computers and other technology. It will involve you using a series of mini-games designed to predict your ability to complete a short computer-based task. After playing the mini-games you will attempt the task a number of times in order to record your actual ability to complete it.

Interview

I will now describe task, while I do try to identify any problems you think you might have. The task you will be attempting involves finding, and double clicking, on 12 labelled targets, in order, as quickly and accurately as possible. The targets will be randomly laid out on the screen, in a grid. You have to find each target based on its numbered label, and then use the mouse to move the cursor over it and double click on it. Once you have clicked it successfully, a sound will play and you can move onto the next target. There will be 12 targets and you will have 30 seconds to complete the task.

Questions:

- Can you think of any problems you might have completing this task... what do you think will they be?
- Do you consider yourself to have any disabilities?
- What technology experience do you have?

You are now going to play a series of mini-games, each one is designed to test a different skill required when using computers. The games have been designed to get gradually harder so you should not worry if you don not manage to complete them all fully. Sit comfortably, do not lean in.

Audio Mini-Game

A number of sounds will be played with a variety of pitches and volumes. As soon as you comfortably hear a sound you will have to press the 'Enter' key as quickly as possible. During the game, please do not lean forward.

Visual Mini-Games

The game is very similar to a normal eye-test. 6 letters will be displayed on the screen which you must identify. As the test progresses, the size and contrast of the letters will change. During the game, please do not lean forward.

In this second eyesight game a circle with a hole (or a letter C) will be displayed on the screen, you need to press the right or left arrow depending on whether the hole is on the right- or left-hand side. If you make a mistake, just carry on. If you decide the targets are too difficult to see, let me know. During the game, please do not lean forward.

Motor Mini-Game

In this game you have to use the mouse to move the cursor and double-click on the targets in the order that they turn red. You need to move as quickly and accurately as possible. The game will measure your ability to move between the targets and to perform a double click.

Evaluation Task

Before Training and Baseline Tasks: You will now attempt the computer based task that was described at the start of this session. [Original task instructions were repeated.]

After training: You will now attempt a number of additional tasks. While all of the tasks will follow the same design as the one you have completed they have been changed in a number of ways to vary their difficulty. The differently of each of the tasks will be different, there may be some that you find easy, other more difficult and some you may be unable to complete.

C.2 Description of Validity (Messick, 1995)

This appendix provides a summary of the six inter-related aspects of validity, as described in Messick (1995). Although they have been identified in order to facilitate the validation of assessments generated as a result research in the field of psychology (with particular emphasis on testing language and mathematical ability), they are also suitable for use in the assessment of accessibility assessments. In the same way that validity is a single (but complex) construct, the accessibility of a device or situation is dependent on a number of underlying factors.[3] The

[3]E.g. "success criteria" found in WCAG 2.0 (Caldwell *et al.*, 2008) or "interacting components" from the CAP.

validity of any accessibility rating is dependent on the information used in the assessment procedure and as such, any model that provides information to an accessibility rating can also be subject to an assessment of its own validity.

C.2.1 Content Relevance and Representativeness (Content)

The content aspect looks for evidence that the test is actually measuring what it claims to be measuring. The two major sources of invalidity that are identified are construct under-representation and irrelevance, where the two concepts describe opposing ends of the same spectrum. Under-representation involves taking an overly-narrow view, which may result in missing important dimensions or facets of the construct. Irrelevance goes too far in the opposite direction, resulting in an overly-broad view that could result in noise caused by confounding variables.

C.2.2 Substantive Theories, Process Models, and Process Engagement (Substantive)

The substantive aspect assesses the substance of a measure in terms of its basis in existing theories and models. By requiring the measure to be based on empirically testable theories, there is increased confidence that the content and processes represented in the measure are correct.

C.2.3 Scoring Models As Reflective of Task and Domain Structure (Structural)

The structural aspect deals with the internal structure of the assessment, which should be consistent with that of the construct domain (structural fidelity). The structure will guide not only the selection of assessment tasks and scoring criteria, but it will also dictate the processes used to turn one into the other.

C.2.4 Generalisability and the Boundaries of Score Meaning

The generalisability aspect provides a form of contextual validity by describing the scope within which the results of the measure can be used. The generalisability of the results of a measure are dependent on the scope of the data collection methods (assessment tasks). A trade-off is described between measuring a task with sufficient validity and allowing the results to be applied to similar (but not equal) tasks or the same task in a different context. For example, the characteristics of the sample population used to calibrate a measure will dictate the characteristics of groups within the general population that the measure can validly be applied to. As well as tasks, limits of score meaning are also affected by generalisability over time, occasions and observers.

C.2.5 Convergent and Discriminant Correlations With External Variables (Convergent)

The external aspect measures the validity of a measure in terms of its comparison against existing measures. The meaning of scores can be substantiated externally by examining the extent to which correlations with existing measures are observed. Both convergent and discriminant patterns are important. Convergence should be seen when there is similarity between the constructs being measured. Discriminant evidence can be used to distinguish a measure from its rivals.

C.2.6 Consequences As Validity Evidence (Consequential)

Rather than describing the measure itself, the consequential aspect focuses on the value implications of score interpretation and use. Ideally the evidence will point to the positive impact of a measure and that negatives are not derived from test invalidity due, for example, to content under-representation or irrelevance. Low scores should be a true representation of the construct, not due to a failing of the measure to capture positive performance.

Appendix D

Research Activity

D.1 Publications and Presentations

A Framework for Adaptive Communication Design (Bell & Machin, 2009): Presented at the ACM Special Interest Group for Design Of Communication, October 2009.

The Benefits and Potential Pitfalls of User Monitoring (Bell *et al.*, 2010): Co-Hosted Workshop at AAATE, October 2010.

Ethical Considerations of How Monitoring Data Is Stored and Used (Li *et al.*, 2010): Co-author, October 2010.

Towards Ubiquitous Accessibility: Capability-Based Profiles and Adaptations, Delivered Via the Semantic Web (Atkinson *et al.*, 2012): Co-Author and presenter at the International Cross-Disciplinary Conference on Web Accessibility, April 2012.

Increasing the Flexibility of Accessibility Modelling Through the Use of Semantic Relationships (Bell & Machin, 2013): Presented at the W3C WAI RDWG Online Symposium on User Modelling for Accessibility (UM4A), July 2013.

Using a Common Semantic Structure to Provide Comparable Contextual Models of Users and Technology (Bell *et al.*, 2014): Accepted as an invited submission as within a parallel session as part of the Universal Access in Human-Computer Interaction Conference, June 2014.

D.2 Research Activity

- Organised and delivered focus groups eliciting older peoples' reaction to assistive technology and adaptive help systems in Rotherham and Loughborough.
- Organised and delivered testing sessions for self-efficacy study in Rotherham, Dundee and Mickleover.
- Organised and delivered user evaluation of Sus-IT monitoring software in Dundee.
- Delivered Sus-IT technology engagement questionnaire in Nottingham, Long Eaton and Loughborough.

D.3 Undergraduate Project Co-Supervision

- **Investigating ICT Use by Older and Disabled People:** Natalie Kassner, 2008–9.
- **Matching People to Jobs:** Nish Gopal, 2008–9.
- **Designing a DVD Player for Older People:** Stephanie Price, 2009-10.
- **Measuring Performance Using Minigames:** Matthew McGovern, 2010-11.
- **Usability Evaluation of Lipreader Software:** Louise Crocker, 2011-12.
- **Development of Attainment Standards for Lipreading and Automated Testing:** Thomas Matthews, 2011-12.
- **Training Older People in the Use of Computer Based Lipreading Tuition Software:** Alex Gray, 2011-12.

D.4 Events Attended

The Actuarial Profession – New Dynamics of Ageing Technology Showcase: Representing and promoting Sus-IT project.

Parliamentary Office of Science and Technology – ICT for Disabled People: Representing and promoting Sus-IT project.

D.5 Other Activities

Organiser Research Student Seminars, Loughborough University Computer Science Department, 2009—2011.

Lead Organiser Science Matters, Postgraduate Research Student Conference, Loughborough University School of Science, 2012.