

MX 0100344 5



# Multi-Modal Usability Evaluation

Submitted in partial fulfillment of the requirements for a Doctor of Philosophy degree

May 2001

Joanne Krystia HYDE

School of Computing Science

Middlesex University

Bounds Green Road

London

N11 2NQ

# **ABSTRACT**

Research into the usability of multi-modal systems has tended to be device-led, with a resulting lack of theory about multi-modal interaction and how it might differ from more conventional interaction. This is compounded by a confusion over the precise definition of modality between the various disciplines within the HCI community, how modalities can be effectively classified, and their usability properties. There is a consequent lack of appropriate methodologies and notations to model such interactions and assess the usability implications of these interfaces. The role of expertise and craft skill in using HCI techniques is also poorly understood. This thesis proposes a new definition of modality, and goes on to identify issues of importance to multi-modal usability, culminating in the development of a new methodology to support the identification of such usability issues. It additionally explores the role of expertise and craft skill in using usability modelling techniques to assess usability issues.

By analysing the problems inherent in current definitions and approaches, as well as issues relevant to cognitive science, a clear understanding of both the requirements for a suitable definition of modality and the salient usability issues are obtained. A novel definition of modality, based on the three elements of sense, information form and temporal nature is proposed. Further, an associated taxonomy is produced, which categorises modalities within the sensory dimension as visual, acoustic and haptic. This taxonomy classifies modalities within the information form dimension as lexical, symbolic or concrete, and classifies the temporal form dimension modalities as discrete, continuous, or dynamic. This results in a twenty-seven cell taxonomy, with each cell representing one taxon, indicating one particular type of modality. This is a faceted classification system, with the modality named after the intersection of the categories, building the category names into a compound modality name.

The issues surrounding modality are examined and refined into the concepts of modality types, properties and clashes. Modalities are identified as belonging to either the system or the user, and being expressive or receptive in type. Various properties are described based on issues of granularity and redundancy. The five different types of clashes are described.

Problems relating to the modelling of multi-modal interaction are examined by means of a motivating case study based on a portion of an interface for a robotic arm. The effectiveness of five modelling techniques, STN, CW, CPM-GOMS, PUM and Z, in representing multi-modal issues are assessed. From this, and using the collated definition, taxonomy and theory, a new methodology, Evaluating Multi-modal Usability (EMU), is developed. This is applied to a previous case study of the robotic arm to assess its application and coverage. Both the definition and EMU are used by students in a case study to test the definition and methodology's effectiveness, and to examine the leverage such an approach may give. The results shows that modalities can be successfully identified within an interactive context, and that usability issues can be described.

Empirical video data of the robotic arm in use is used to confirm the issues identified by the previous analyses, and to identify new issues. A rational re-analysis of the six approaches (STN, CW, CPM-GOMS, PUM, Z and EMU) is conducted in order to distinguish between issues identified through craft skill, based on general HCI expertise and familiarity with the problem, and issues identified due to the core of the method for each approach. This is to gain a realistic understanding of the validity of claims made by each method, and to identify how else issues might be identified, and the consequent implications. Craft skill is found to have a wider role than anticipated, and the importance of expertise in using such approaches emphasised. From the case study and the re-analyses the implications for EMU are examined, and suggestions made for future refinement.

The main contributions of this thesis are the new definition, taxonomy and theory, which significantly contribute to the theoretical understanding of multi-modal usability, helping to resolve existing confusion in this area. The new methodology, EMU, is a useful technique for examining interfaces for multi-modal usability issues, although some refinement is required. The importance of craft skill in the identification of usability issues has been explicitly explored, with implications for future work on usability modelling and the training of practitioners in such techniques.



# ACKNOWLEDGEMENTS

This work would not have been possible without the support and encouragement of a large number of people. All the credit goes to them, and all the errors remain my own.

It was David Duke, from the Department of Computer Science, University of York, who started me on this path, and encouraged me to believe that the research was worthwhile.

Ann Blandford has been a wonderful supervisor, and has kept me on the straight and narrow despite my best attempts to wander off-track! Gordon Rugg has supervised me with sensitivity and assurance, and channeled my enthusiasm into more beneficial areas. Howard Goodman provided me with valuable technical input at a critical stage of my research.

Thanks is especially due to Thomas Green and Bernard Parsons for providing two of the case studies used in the thesis. I would like to make a special mention to those Middlesex students who took module COM3210 in the autumn semester 1998, for allowing my methodology to form part of their coursework.

This PhD research has been funded by the School of Computing Science, Middlesex University. I would like to take this opportunity to thank the School for providing financial assistance, both for the three years of the original grant, and for a six month extension to allow completion.

The staff and students of Middlesex University have provided immeasurable support to me in my time here. Special mention must go to fellow members of 'The Aviary': namely Serengul Smith, Yin Leng Theng and Cecile Rigny; and the rest of the PhD sufferers. Harold Thimbleby had enough courage both to accept me onto the PhD programme and to allow me to stay on it. Various members of staff have given me helpful feedback.

My friends have had to put up with a lot from me over the past few years. My thanks to the Rangelagh quiz team for letting me let off steam with them. All the Exiles have made me feel secure and confident in myself. Without you lot, I don't know how I would have coped. Karen and James provided me with a welcome escape hatch whenever the research became too scary and I needed to regain perspective.



My family have been very proud of what I've been doing, even though they haven't been entirely sure of what exactly it is that I'm researching. My thanks to my Mum especially, for frequently saying that I would finish, when I was none too sure myself. This thesis is dedicated to my friends and family: you make it all worthwhile.

# CONTENTS

<b>1.</b>	<b>INTRODUCTION .....</b>	<b>1</b>
1.1	THE RESEARCH PROBLEM .....	1
1.1.1	<i>Background to the research problem.....</i>	<i>1</i>
1.1.2	<i>Motivation.....</i>	<i>3</i>
1.2	RESEARCH AIMS.....	3
1.2.1	<i>Thesis Summary.....</i>	<i>4</i>
1.3	SUMMARY .....	6
<b>2.</b>	<b>EVALUATING MULTI-MODAL USABILITY: A PRELIMINARY STUDY .....</b>	<b>7</b>
2.1	INTRODUCTION .....	7
2.2	MULTI-MODAL INTERACTION.....	7
2.2.1	<i>Possible benefits of multi-modal interaction.....</i>	<i>8</i>
2.2.2	<i>Possible problems with multi-modal interaction .....</i>	<i>10</i>
2.3	COMPARATIVE STUDY OF FIVE TECHNIQUES APPLIED TO A MULTI-MODAL INTERFACE .....	11
2.3.1	<i>The Robotic Arm Interface.....</i>	<i>12</i>
2.3.2	<i>The Five Approaches .....</i>	<i>13</i>
2.3.3	<i>Analysis.....</i>	<i>16</i>
2.4	SUMMARY .....	26
<b>3.</b>	<b>MODALITY DEFINITIONS AND PROPERTIES.....</b>	<b>28</b>
3.1	INTRODUCTION .....	28
3.2	MODALITY DEFINITIONS.....	28
3.2.1	<i>Sense .....</i>	<i>30</i>
3.2.2	<i>Device .....</i>	<i>31</i>
3.2.3	<i>Mode .....</i>	<i>31</i>
3.2.4	<i>Form .....</i>	<i>32</i>
3.2.5	<i>Other Definitions .....</i>	<i>33</i>
3.2.6	<i>Derived Definition Elements.....</i>	<i>35</i>

3.3	MULTI-MODAL PROPERTIES .....	37
3.3.1	<i>Modality Theory</i> .....	37
3.3.2	<i>The Multi-Sensori-Motor Framework</i> .....	38
3.3.3	<i>Multimodal Multimedia Design Space</i> .....	40
3.3.4	<i>CARE Properties</i> .....	41
3.3.5	<i>Taxonomy of Representational Systems</i> .....	42
3.3.6	<i>Derived Multi-modal Properties</i> .....	45
3.4	COGNITIVE CONSTRAINTS .....	50
3.4.1	<i>Four Cognitive Models</i> .....	51
3.4.2	<i>Properties of Senses</i> .....	52
3.4.3	<i>Representation Issues</i> .....	53
3.4.4	<i>Combination Issues</i> .....	54
3.4.5	<i>Temporal Issues</i> .....	56
3.4.6	<i>Derived Cognitive Issues</i> .....	57
3.5	SUMMARY .....	58
4.	<b>MODALITY DEFINITION AND THEORY</b> .....	<b>59</b>
4.1	INTRODUCTION.....	59
4.2	A NEW DEFINITION OF MODALITY .....	59
4.2.1	<i>Human Sense</i> .....	60
4.2.2	<i>Information Form</i> .....	60
4.2.3	<i>Temporal Form</i> .....	61
4.2.4	<i>New Definition</i> .....	62
4.3	MODALITY TAXONOMY .....	63
4.3.1	<i>Sensory Dimension</i> .....	64
4.3.2	<i>Information Form Dimension</i> .....	65
4.3.3	<i>Temporal Dimension</i> .....	66
4.3.4	<i>Analysis of Taxonomy</i> .....	67
4.3.5	<i>Example Of Taxonomy Use</i> .....	69
4.4	MODALITY TYPES AND PROPERTIES .....	71
4.4.1	<i>User and System Modality Types</i> .....	71
4.4.2	<i>Expressive and Receptive Modality Types</i> .....	72
4.4.3	<i>Composite and Atomic Modalities</i> .....	73
4.4.4	<i>Dependent Modalities</i> .....	74
4.4.5	<i>Mismatch</i> .....	74
4.4.6	<i>Redundancy</i> .....	76



4.5	MODALITY CLASHES .....	76
4.5.1	<i>Physical clashes</i> .....	76
4.5.2	<i>Lexical clashes</i> .....	77
4.5.3	<i>Temporal Clashes</i> .....	77
4.5.4	<i>Semantic clashes</i> .....	78
4.5.5	<i>Clash unless expert</i> .....	78
4.6	SUMMARY .....	79
<b>5.</b>	<b>DEVELOPMENT OF EMU, A MULTI-MODAL METHODOLOGY .....</b>	<b>81</b>
5.1	INTRODUCTION.....	81
5.2	IMPORTANT PROPERTIES .....	82
5.2.1	<i>Properties from Comparative Case Study</i> .....	82
5.2.2	<i>Properties from existing theories of modality</i> .....	82
5.2.3	<i>Cognitive Properties</i> .....	83
5.2.4	<i>Properties relating to the new definition</i> .....	83
5.3	ROBOTIC ARM EXAMPLE.....	84
5.4	TELEPHONE EXAMPLE.....	86
5.5	REPRESENTATION.....	87
5.5.1	<i>Tabular Representation</i> .....	87
5.5.2	<i>Diagrammatic Representations</i> .....	89
5.5.3	<i>Notational Representations</i> .....	90
5.5.4	<i>Analysis</i> .....	97
5.6	STRUCTURE.....	98
5.6.1	<i>Stage 1. Define the task that is to be analysed</i> .....	100
5.6.2	<i>Stage 2. Modality lists</i> .....	101
5.6.3	<i>Stage 3. Define the user, system and environment profiles</i> .....	101
5.6.4	<i>Stage 4. Profiles compared to modality listings</i> .....	102
5.6.5	<i>Stage 5. Interaction sequence listing</i> .....	102
5.6.6	<i>Stage 6. Add in clashes, etc</i> .....	102
5.6.7	<i>Stage 7. Assess the use of modalities</i> .....	103
5.6.8	<i>Stage 8. Final report</i> .....	103
5.6.9	<i>Analysis</i> .....	103
5.7	EMU ANALYSIS OF ROBOTIC ARM.....	105
5.7.1	<i>Usability issues</i> .....	107
5.7.2	<i>Application issues</i> .....	108
5.7.3	<i>Discussion</i> .....	109
5.8	SUMMARY .....	110

<b>6.</b>	<b>APPLICATION OF DEFINITION AND METHODOLOGY .....</b>	<b>112</b>
6.1	INTRODUCTION .....	112
6.2	METHOD.....	113
6.2.1	<i>Domain</i> .....	113
6.2.2	<i>Subjects</i> .....	113
6.2.3	<i>Procedure</i> .....	113
6.3	IDENTIFICATION OF MODALITIES .....	116
6.3.1	<i>Results</i> .....	117
6.3.2	<i>Discussion</i> .....	129
6.4	UNDERSTANDING OF INTERACTION .....	132
6.4.1	<i>Results</i> .....	132
6.4.2	<i>Discussion</i> .....	138
6.5	IDENTIFICATION OF USABILITY ISSUES.....	140
6.5.1	<i>Results</i> .....	140
6.5.2	<i>Discussion</i> .....	149
6.6	GENERAL DISCUSSION.....	153
6.6.1	<i>Identification of Modalities</i> .....	153
6.6.2	<i>Understanding of Interaction</i> .....	154
6.6.3	<i>Identification of Usability Issues</i> .....	156
6.7	SUMMARY .....	157
<b>7.</b>	<b>EVALUATION OF APPROACHES .....</b>	<b>159</b>
7.1	INTRODUCTION.....	159
7.2	VIDEO DATA ANALYSIS.....	160
7.2.1	<i>The users</i> .....	161
7.2.2	<i>Comparison with previous analyses</i> .....	162
7.2.3	<i>Discussion</i> .....	167
7.3	SYSTEMATIC RE-ANALYSIS OF METHODS .....	168
7.3.1	<i>Scope of approaches</i> .....	169
7.3.2	<i>Role of expertise</i> .....	175
7.3.3	<i>Accuracy and completeness</i> .....	176
7.3.4	<i>Representation form</i> .....	178
7.3.5	<i>Discussion</i> .....	180
7.4	SUMMARY .....	181

<b>8.</b>	<b>IMPLICATIONS FOR EMU .....</b>	<b>183</b>
8.1	INTRODUCTION.....	183
8.2	SCOPE.....	183
8.3	OMISSIONS.....	185
8.4	REPRESENTATION .....	186
8.5	LENGTH.....	188
8.6	DISCUSSION.....	189
8.7	SUMMARY .....	190
<b>9.</b>	<b>CONCLUSIONS.....</b>	<b>191</b>
9.1	INTRODUCTION.....	191
9.2	THESIS SUMMARY.....	191
9.2.1	<i>Multi-modal theory</i> .....	<i>191</i>
9.2.2	<i>Application of theory</i> .....	<i>193</i>
9.2.3	<i>Scoping of usability modelling techniques</i> .....	<i>195</i>
9.3	CONTRIBUTIONS AND LIMITATIONS.....	197
9.4	CURRENT RESEARCH .....	199
9.5	FUTURE WORK.....	200
<b>10.</b>	<b>BIBLIOGRAPHY.....</b>	<b>203</b>
<b>11.</b>	<b>APPENDIX A: A SURVEY OF FOUR COGNITIVE MODELS .....</b>	<b>214</b>
11.1	HUMAN INFORMATION PROCESSING MODEL .....	214
11.2	MODEL HUMAN PROCESSOR .....	215
11.3	INTERACTING COGNITIVE SUBSYSTEMS.....	216
11.4	EXECUTIVE PROCESS-INTERACTIVE CONTROL .....	217
<b>12.</b>	<b>APPENDIX B: TAXONOMY AS APPLIED TO ROBOTIC ARM .....</b>	<b>219</b>



<b>13.</b>	<b>APPENDIX C: METHODOLOGY APPLIED TO ROBOTIC ARM .....</b>	<b>230</b>
13.1	STAGE 1. DEFINE THE TASK THAT IS TO BE ANALYSED.....	230
13.2	STAGE 2. MODALITY LISTS .....	230
13.2.1	<i>User modalities:</i> .....	230
13.2.2	<i>System modalities:</i> .....	231
13.2.3	<i>Granularity:</i> .....	231
13.3	STAGE 3. DEFINE THE USER, SYSTEM AND ENVIRONMENT PROFILES.....	232
13.3.1	<i>User Profile:</i> .....	232
13.3.2	<i>System Profile:</i> .....	232
13.3.3	<i>Environment Profile:</i> .....	233
13.4	STAGE 4. PROFILES COMPARED TO MODALITY LISTINGS .....	233
13.4.1	<i>User profile compared with System modalities:</i> .....	233
13.4.2	<i>System profile compared with User modalities:</i> .....	233
13.4.3	<i>Environment profile compared with User and System modalities:</i> .....	233
13.5	STAGES 5 AND 6. INTERACTION SEQUENCE LISTING AND CLASHES ETC. ....	234
13.6	STAGE 7. ASSESS THE USE OF MODALITIES .....	282
13.7	STAGE 8. FINAL REPORT .....	282
<b>14.</b>	<b>APPENDIX D: THE HEATING CONTROL PANEL WEBSITE.....</b>	<b>284</b>
<b>15.</b>	<b>APPENDIX E: TASK STEP SUMMARY FOR TUBE TICKET MACHINE TASK.....</b>	<b>285</b>
<b>16.</b>	<b>APPENDIX F: STUDENT PACK.....</b>	<b>286</b>
<b>17.</b>	<b>APPENDIX G: QUESTIONNAIRES .....</b>	<b>314</b>
17.1	FIRST QUESTIONNAIRE QUESTIONS.....	314
17.2	FIRST QUESTIONNAIRE RESULTS TABLES .....	315
17.3	SECOND QUESTIONNAIRE QUESTIONS .....	319
17.4	SECOND QUESTIONNAIRE RESULTS TABLES .....	321

<b>18.</b>	<b>APPENDIX H: VIDEO DATA ANALYSIS .....</b>	<b>329</b>
18.1	THE USERS.....	329
18.2	EXCERPT ONE .....	330
18.2.1	<i>Length of interaction examined.....</i>	<i>330</i>
18.2.2	<i>Listing of what happened .....</i>	<i>330</i>
18.3	EXCERPT TWO .....	331
18.3.1	<i>Length of interaction examined.....</i>	<i>331</i>
18.3.2	<i>Listing of what happened .....</i>	<i>331</i>
18.4	EXCERPT THREE .....	332
18.4.1	<i>Length of interaction examined.....</i>	<i>333</i>
18.4.2	<i>Listing of what happened .....</i>	<i>333</i>
18.5	EXCERPT FOUR.....	334
18.5.1	<i>Length of interaction examined.....</i>	<i>334</i>
18.5.2	<i>Listing of what happened .....</i>	<i>334</i>
18.5.3	<i>Speech from excerpt .....</i>	<i>335</i>
18.6	EXCERPT FIVE.....	336
18.7	EXCERPT SIX.....	336
<b>19.</b>	<b>APPENDIX I: THE RE-ANALYSES.....</b>	<b>338</b>
19.1	STN RE-ANALYSED .....	338
19.2	DISCUSSION OF STN RE-ANALYSIS.....	342
19.3	CW RE-ANALYSIS.....	345
19.4	DISCUSSION OF CW RE-ANALYSIS.....	348
19.5	CMN AND CPM GOMS RE-ANALYSES.....	350
19.6	DISCUSSION OF CMN AND CPM GOMS RE-ANALYSIS.....	353
19.7	PUM RE-ANALYSIS.....	355
19.8	DISCUSSION OF PUM RE-ANALYSIS .....	358
19.9	Z RE-ANALYSIS.....	360
19.10	DISCUSSION OF Z RE-ANALYSIS.....	364
19.11	EMU RE-ANALYSIS .....	366
19.12	DISCUSSION OF EMU RE-ANALYSIS .....	369
<b>20.</b>	<b>APPENDIX J: A COMPARISON OF FIVE TECHNIQUES FOR INVESTIGATING GENERAL AND MULTI-MODAL USABILITY ISSUES .....</b>	<b>372</b>

# **1. INTRODUCTION**

## **1.1 The Research Problem**

One aim of human-computer interaction (HCI) is to enable the positive use of computer systems, by providing the underlying theory to design systems that are usable, in that they can support the tasks which users wish to accomplish, and to provide the appropriate means of assessing the usability of such systems once they are designed. One particular form of interaction involving multiple forms of information, known as multi-modal interaction, is imperfectly understood, due to a limited understanding of both what a modality actually is, and the relevant usability properties. Additionally, there is a lack of appropriate methodologies which can specifically examine multi-modal usability, and the role that the skill of the analyst plays in applying a methodology is imperfectly understood. This thesis seeks to address these problems.

### **1.1.1 Background to the research problem**

The issue of effective interaction has become of critical importance as computers become more widespread. Computers are no longer found only at the desk, but in a wide variety of other situations, and are not necessarily passive but active, as for instance in the realm of robotics, where robots are capable of accepting input through sensors to the environment as well as directly from a user (Anzai, 1994). Not only are computers mobile, they are also immersive. The growth of virtual environments and the synthetic experiences which they offer give HCI new challenges in terms of three dimensional manipulation and associated issues of navigation, orientation and feedback (Mills & Noyes, 1999).

The pace of technological change has broadened the range of devices that can be used to facilitate interaction. No longer limited to the keyboard, users can now use devices such as a mouse, stylus, touch screen, data-glove and foot-pedal. The development of graphical user interfaces means that vision is currently the most commonly used means of information presentation by a computer system, due to the large amount of information that can be displayed and retained on a screen for inspection by the user. Audio interaction is growing in importance, and is now used in interfaces both as a



means of inputting information, e.g. by direct speech, and for outputting information in the form of either words or sound. Touch-based, or haptic, interaction tends to be used by the user rather than by the computer. The user can manipulate a mouse, operate a keyboard, or touch a screen, or handle any one of an increasing number of touch-based input devices.

Indeed, it seems as though interaction devices are now limited more by the parameters of the human body than by technical considerations. Given that current interfaces now have to deal with a very wide range of tasks and an extremely large potential user group, the onus is on new interface technology to try to make interaction more straightforward, by drawing on humans natural communication skills. A great deal of interest has been addressed towards systems which incorporate more than one means of input or output at a time: so-called 'multi-modal' interfaces. A multi-modal interface is one which allows users to interact with a system by means of several input and output devices. Such systems attempt to capitalise on people's innate ability to utilise several means of communicating information at a time.

However, most research into multi-modal interaction is device-led (see Dowell et al (1994), Alty (1991), Roth & Hefley (1993) amongst others), with interest focused on the application of novel interaction technologies rather than on the development of a broader theory of multi-modal usability. Furthermore, there is confusion within the area of human-computer interaction over what constitutes a modality, with psychologists defining modality in terms of sensory channels (Krapichler et al, 1999), and computer scientists taking a device-centred view (Verjans, 1995). Inevitably, this means that such theory as does exist does not address the relevant issues, and is complicated by a lack of appropriate methodologies to support the effective analysis of multi-modal usability.

Indeed, as Cockton (1998) points out, little is known about how methodologies can be used together to help successful design. Existing methodologies are hampered by their dependence on the existing skill of the practitioner. This is acknowledged rarely (Blandford et al, 1998a). Thus, an understanding of how the skill of the practitioner affects the application of a methodology is as essential as understanding what the scope of the methodology might be.

### **1.1.2 Motivation**

I first became interested in the issues surrounding multi-modality and its representation whilst researching a dissertation for an MSc in Information Processing and the Man-Machine Interface at the University of York under the supervision of Dr. David Duke. The dissertation investigated aspects of a software artifact for combining multi-modal inputs by means of specifying the artifact using the Z notation, then reasoning about the usability of the specification and proposing changes to the structure of the original artifact (Hyde, 1995). The software artifact was based on MATIS, a Multimodal Airline Travel Information Service (see Nigay et al (1993) and Nigay & Coutaz (1995) for a full description) developed at the Laboratoire de Genie Informatique, University of Grenoble, France.

The in-depth study of the MATIS system revealed various problems regarding multi-modal interaction, as for example when two different types of input were used in conjunction with each other, such as when the user is typing and speaking different data at the same time, or other problems where users did not use the multi-modal capabilities of the system and instead used only one modality at a time (Duke et al, 1998).

I discovered that there were various interpretations of the term “modality” current within the HCI community, and that different interpretations were presented within the same papers on occasion. Researchers often disagreed fundamentally on the interpretation of the term, with the result that what research existed suffered from the lack of a common vocabulary of accepted terms. The use of Z to explore aspects of usability stimulated an interest in the use of modelling techniques generally and their limitations for assessing usability, with particular reference to multi-modal usability.

Together, this prompted an interest in multi-modal usability, both in terms of the definition of modalities and the associated usability properties, and in the use and limitations of existing approaches for assessing the usability of such interfaces.

## **1.2 Research Aims**

The main area of research presented in this thesis concerns multi-modal usability, involving the precise and useful definition of modality, the investigation of multi-modal usability properties, and the development of a methodology to apply this theoretical



understanding. An associated area is the importance of examining the level of skill needed in applying methodologies.

### 1.2.1 Thesis Summary

The introduction chapter presents the area of research and the aims of this thesis: the main aim of addressing the lack of theoretical understanding of multi-modal usability; the aim of addressing the lack of appropriate approaches to apply that theoretical understanding effectively; and the role of skill in applying such methodologies. The motivating factors and the approach adopted are discussed, and the thesis is summarised to provide an overview of the research and how it has been conducted.

In chapter two the research is motivated by a brief discussion on the nature of multi-modal systems, and the lack of theoretical understanding of multi-modal usability is identified as a cause for concern. In order to provide a starting point for examining the issues surrounding the theoretical understanding of multi-modality, as well as issues affecting the usability modelling of such systems, a case study involving the application of five modelling approaches (STN, CW, CPM-GOMS, PUM and Z) to a portion of an interface of a robotic arm is presented. This highlights the need to clearly identify and represent modalities, and their constraints. The study also highlights general issues related to the application of usability theory, including issues relating to representation, guidance, and coverage.

Chapter three critically assesses research in the definition of modality and understanding of multi-modal issues. By examining existing definitions of modality their problems are exposed and discussed. The requirements for a definition of modality within an HCI context are investigated, and the need for a new definition, using the human senses in combination with other elements, established. Existing approaches to the usability of multi-modal and multi-media systems are examined and assessed. Four areas of critical relevance to multi-modal usability are identified. Certain mental and physical capabilities of the user are investigated by examining several cognitive models. From this, cognitive restrictions applicable to multi-modal usability are identified.

Chapter four addresses the original aim of identifying and clarifying theoretical issues relevant to multi-modal usability by developing a new definition of modality, a new taxonomy, and the clarification of other properties and clashes relevant to modality. A new three-part definition of modality is proposed, based on the elements of sense,



information form, and temporal nature. The new definition is supported by a taxonomy providing a structure so that modalities can be identified according to their attributes. The final taxonomy has twenty-seven cells, each cell representing one modality. The work on the four issues identified in chapter three is applied to the new definition of modality to clearly draw out those properties of importance. Modalities are identified as User or System, and Receptive or Expressive. The granularity of modalities is addressed by their definition as composite, atomic and dependent. Issues of mismatch and redundancy are explored. The cognitive restrictions established in chapter three are formalised into five different types of clashes: physical, relating to sensory processing restrictions; lexical, relating to language processing restrictions; temporal, relating to temporal processing constraints; semantic, relating to clashes of meaning due to the automatic cognitive processes involved; and clash unless expert, relating to the constraints associated with levels of expertise.

In chapter five the theory proposed in chapter four is developed into a methodology: Evaluating Multi-modal Usability (EMU). Three areas are considered: which properties and elements should be incorporated; what form of representation should be used within the methodology; and the overall structure of the methodology, guiding its application and the kind of issues to be addressed. An examination of the robotic arm analysed previously in chapter two, using the new definition, is used to gain insight into matters relating to representation and structure. An example case study of answering a telephone is also used. The final representation takes a notational form in order to clearly represent issues of ordering, combination, preconditions, and choice. Issues of structure are then examined, incorporating relevant aspects of other approaches where appropriate, resulting in a methodology of eight stages (EMU). The robotic arm is then analysed using EMU in order to examine its use by a practitioner.

Chapter six evaluates the usability and usefulness of the new definition, taxonomy and methodology, through a case-study in which computer science students assessed the usability issues associated with two tasks performed on two multi-modal interfaces. The definition and taxonomy were found to be both usable and useful with some reservations regarding types, preconditions and the use of AND/OR. Very few subjects identified any clashes or properties. Overall, the subjects are found to have mixed feelings regarding the methodology; usability issues are identified, and additional issues

with no direct link to previous stages of the analysis are also included in the final reports.

In chapter seven video data of the robotic arm in use is used to empirically assess the usability issues identified by the previous analyses. The six analyses (STN, GOMS, CW, PUM, Z and EMU) are then reanalysed in order to determine the scope of the approaches and the use of craft skill in their application.

Chapter eight assesses the implications of the previous research for EMU. Various improvements that could be made to the methodology, in terms of its scope, content, representation and length, are discussed, and suggestions proposed.

The concluding chapter brings together the research presented in this thesis and assesses it with regards to the original aims as presented in chapter one. The strengths and weaknesses of the work are examined, and the resulting contribution to knowledge assessed. The research is placed within the context of current work in the field, and future work based on this research is discussed.

### **1.3 Summary**

This chapter has introduced the research problem, and the motivating factors behind the research, and presented the approach adopted and the overall thesis structure. The next chapter will explore the issues surrounding multi-modal interaction in more depth by means of a motivating case study of five approaches for modelling usability as applied to a portion of a multi-modal interface.



## **2. EVALUATING MULTI-MODAL USABILITY: A PRELIMINARY STUDY**

### **2.1 Introduction**

Multi-modal systems make use of multiple ways of inputting and outputting information. However, although there are many potential benefits from such systems, there are also many potential problems. This chapter first introduces the concept of multi-modal interaction, and discusses the advantages and disadvantages of such an approach, in order to highlight existing deficiencies in our understanding of the interaction process.

To examine these issues in greater depth, a study involving the application of five modelling approaches to a portion of a multi-modal interface for a robotic arm is presented. From this areas of investigation are derived which are examined in future chapters.

### **2.2 Multi-modal interaction**

Traditional interface development has concentrated on three of the five senses with which humans interact with the environment; vision, hearing and touch. There are as yet no interfaces which routinely utilise either taste or smell and, given the complexity of these two senses and that they are as yet imperfectly understood, it seems unlikely that any such interface will be developed in the near future. Interfaces typically use the senses in isolation, using one sense for input and one for output, as in the case of personal computers, which use touch for input and vision for output, with sound taking a secondary role in the output.

However, when humans communicate, they do not use the senses individually but utilise various ways of passing information through the sensory channels, combining them into one overall stream of information (Alty, 1991). Allowing information to flow in several parallel channels at a time has been shown to enhance interaction (Krapichler et al, 1999). For example, humans normally process both voice and gesture at the same time when in



conversation, using one to disambiguate the other should confusion occur, and to support the other if the context requires clarification (McKenzie Mills & Alty, 1997). This mix of complementary and redundant information helps to ensure effective communication. Interfaces which could handle such a broad bandwidth of sensory information simultaneously would therefore potentially allow a more natural and more usable interaction with a system.

Interfaces which allow or make use of one or more sensory channel for the input or output of information are commonly known as “multi-modal”. They differ from multimedia systems, in that they encompass both input and output, whereas multimedia systems are characterised by being output oriented, concentrating the expression of information by means of vision and sound (Roth and Hefley, 1993). Indeed, multimedia systems could be considered as a limited subset of multi-modal interfaces (Coutaz et al, 1993).

### **2.2.1 Possible benefits of multi-modal interaction**

The use of more than one modality can greatly enhance the user’s understanding. By providing more ways of communicating between the system and the user, it is less likely that ambiguity will occur. Using modalities in combination not only resolves issues of ambiguity, but allows a richer, more efficient and effective type of communication with one modality supporting another, so that “the total usability thus obtained is greater than the usability of each individual modality” (Bretan & Karlgren, 1993).

There have been various interfaces designed to take advantage of multiple modalities and the benefits that they can bring. For example Shoptalk (Cohen, 1992) combines direct manipulation and natural language in a way which attempts to overcome the weaknesses of each form of interaction by using them in such a way that their strengths cancel the weaknesses out. For instance, one of the major problems with direct manipulation is that there is no way for users to describe entities, or provide temporal constraints. Natural language currently suffers from the problem that technology is as yet unable to support entirely natural dialogue, so that systems must provide checking mechanisms which slows down the interaction and can frustrate users. Cohen (1992) explicitly acknowledges the way that interaction can be made more beneficial by using both styles of interaction together to provide greater power.

Other systems include the dynamic map systems described by Oviatt (1997) which allow a range of functions and support. The CUBRICON (The CUBR Intelligent Conversationalist) interface (Neal & Shapiro, 1991) integrates speech input and output, graphics, direct manipulation and natural language text, in order to achieve “the manner in which two people naturally communicate in co-ordinated multiple modalities when working at a graphics device.” This is achieved through the automatic selection of output by the interface, working to selection rules.

Multi-modal systems are often used in aircraft cockpits, normally as a combination of visual representation of information with aural alerts to draw attention to critical situations (Palmer et al, 1993). This is particularly effective when pilots are engaged in monitoring other aspects of the flight and their attention is diverted.

Multi-modal interfaces are also seen as being helpful from a teaching perspective. Alpert et al (1995) show how interfaces can be designed to include multi-modal help components based on the use of “gurus” to assist people in the Smalltalk environment. The system utilises differing modalities in order to educate the user using animated demonstrations “not only to tell them what to do but show them how to do it.” It is asserted that this use of animation combined with voice-over avoids overloading any one input channel for the user, making the information easier to absorb.

The growth of multi-modal systems supporting virtual environments which allow users the ability to manipulate synthetic objects in a three-dimensional space have various advantages and applications in the fields of medicine, education and entertainment, and obvious advantages in the training of users in safety-critical situations, allowing as they do the mimicking of complex operations in a risk-free environment (Mills & Noyes, 1999). The multi-modal nature of such worlds which integrate several forms of output and input can potentially enhance human performance greatly, and research is underway to take advantage of this (Mills & Noyes, 1999).



### 2.2.2 Possible problems with multi-modal interaction

Although multi-modal systems can potentially facilitate human-computer interaction there are serious problems with their effective development. If the goal of multi-modal systems is to increase usability, then an understanding of the usability characteristics of such systems is essential. The usability of ordinary systems is still, after several decades, imperfectly understood, so it is hardly surprising that multi-modal systems, with their more complex issues, have been ignored or treated in a simplified fashion. The development of virtual reality has accentuated this trend, making it more critical that these issues are addressed and resolved. This lack of usability research on multi-modal systems means that their usability properties are not fully understood, making them potentially less usable, and requiring more resources in the design life-cycle for empirical research. This problem is compounded by most multi-modal systems currently being more technology driven than user-driven. This is similar to the verification/validation problem in software (see Mellor (1994) for a particularly succinct description of how it applies in the field of aeronautic catastrophe), and as Alty (1991) observes: “Very often we work outwards from the technology asking ‘what might the user *be able to do* with this new technology?’ rather than ‘what might the user *want to do* with this new technology?’ ”. What is needed is an approach which asks “can the multi-modal interface be used?” rather than “have we built an interface which uses multi-modality?”.

Multi-modal systems are not necessarily more usable because they include more modalities. If the modalities are not combined in a complementary manner the interaction will be impeded rather than facilitated; as (Coutaz et al, 1995) have stated, “without a sound scientific framework to assess the compatibility between system properties and users’ preferences, improved usability through multimodality remains just another vacuous and potentially dangerous slogan.”

Johnson & Nemetz (1998) cite the lack of detailed design knowledge as a contributing factor in the poor design and evaluation of multimedia systems, and point out that the introduction of a wider spectrum of media not only allows for richer and more effective communication, but also for ambiguity, confusion and contradiction within the interaction. Bearne et al (1994) observe that the provision of more information in more forms places



new and increased demands on the attentional capacity of system users. Indeed, the overload on attentional capacity has been shown to be a contributing factor in several aircraft disasters such as the Habsheim disaster in France, June 1988 (Mellor, 1994), as well as in many more minor incidents (Palmer et al, 1993).

The variety of information available needs to be effectively combined into formats which support rather than hamper the user (Roth and Hefley, 1993). Without guidance on how this might be achieved, systems with too many, or poorly integrated, input and output options run a risk of fragmenting user behaviour and causing cognitive overload (Oviatt, 1997). They can also impact on shared workloads, causing a degradation in communication with colleagues and impinging on their ability to maintain a common frame of reference or shared situation assessment, which can have important safety implications when engaged in safety-critical situations such as aircraft landings (Woods et al, 1994). It has also been shown that where a variety exists, users do not always make use of it effectively, because of confusion caused by the many options, resulting in a lack of proficiency in methods of interaction which thus hinder effective communication with the interface (Woods et al, 1994). Not only must these systems support user capabilities, they must do so relative to the task undertaken (Roth et al, 1997).

There is therefore a need for a thorough understanding of the properties of multi-modal interaction, in order that the usability of such systems can be assessed and improved.

### **2.3 Comparative study of five techniques applied to a multi-modal interface**

The previous section has introduced the concept of multi-modal interaction and identified the lack of a theoretical understanding of multi-modal usability. One area in which such theory is crucial is that of usability assessment. In this section the results of a comparative study of five existing modelling techniques as applied to a portion of a multi-modal interface are presented. For a full description of the interface and the application of the techniques see (Hyde et al, 1998) in Appendix J.

This motivating case study was conducted at an early stage of the research, and had two aims. The primary aim was to investigate how well these approaches could deal with multi-

modal usability issues, in order to demonstrate practically that multi-modal usability issues cannot be identified satisfactorily by many existing techniques, due to a lack of multi-modal theory. Future chapters will seek to clarify the understanding of multi-modal usability. In addition, there was the important secondary aim of determining what kind of general usability issues the approaches are able to identify, and the issues surrounding their application. In this way the important influences of representation and methodology can be discussed. This will then be used in future chapters in the development of an approach which is able to handle issues identified as important to multi-modal usability.

### **2.3.1 The Robotic Arm Interface**

The techniques were applied to a portion of an interface of a robotic manipulator for use by wheelchair-bound people (see Parsons et al, 1995), intended to be used in a domestic context for everyday tasks such as feeding and grooming. The arm consists of eight joints, powered by motors, which can move either individual joints, or the whole arm at once, via input devices using gesture or speech. In the portion of the interface examined, the user is able to set the direction and speed of the arm, choose whether to manipulate all or part of the arm, and initiate and stop arm movement, by means of menu-based software. At the time the analyses were conducted, none of the input or output devices for the arm controller had been implemented. The interface was to be implemented as a Windows application, with menu options selected from a screen using either voice recognition or gesture.

The gesture input system is based on a baseball cap with two sensors and is presently implemented so that a cursor moves along underneath the menu options continuously in turn, and if the correct gesture is made when the cursor is underneath a particular option, then that option is selected. The voice recognition system allows direct menu option selection simply by saying the menu option out loud. The output is not yet implemented, but will take the form of a small LCD display attached to the base of the arm showing the menus.

Since three of the five techniques under investigation involve analysis of predefined user tasks, one typical but non-trivial task was specified for analysis, that of moving the arm from rest to a certain position. The task was thought of in terms of the device task: moving the gripper to position XYZ from rest; and the user task: turning on a light switch. This is



the kind of simple task that the robotic arm is designed to support, and can therefore shed considerable light on the usability issues inherent in this interface.

### **2.3.2 The Five Approaches**

The five techniques chosen cover a wide spectrum, in order to test a range of formal system and user approaches, and include a diagrammatic representation (State Transition Networks), a natural language goal-based method (Cognitive Walkthrough), a hierarchical goal-based technique (GOMS), a means-end planning based technique (Programmable User Model), and one based on set theory and first-order predicate logic (Z). Cognitive Walkthrough (CW) and GOMS are user centred, while Programmable User Model (PUM) analyses both the user domain tasks and the device tasks, and the Z and State Transition Network (STN) representations are more system-oriented. Z and STN are not standard usability modelling techniques, being generally used in software engineering to describe the specification and functionality of a system. They were included to see what leverage well-known techniques with no explicit usability analysis support could give to the understanding of the interface, against which other usability-specific techniques could be compared.

The State Transition Network representation of the interface was originally created to clarify understanding of the robotic arm in terms of the structure of the interaction rather than for any usability assessment. STNs are a popular and well-established way of diagrammatically representing an interface (Dix et al, 1993) and can take various forms. For simple interaction sequences STNs can clearly illustrate the flow of interaction and allow redundant cycles to be identified. Not only are they easy to use, requiring little training and being quick to learn, they are also quick to write, uncomplicated, and allow an overall visual representation of the interface to be communicated to the analyst. The simplest type, as used here, has each state of the system represented by a circle, linked by lines, or transitions, which correspond to the actions necessary to move from that state to another.

Cognitive Walkthrough (Wharton et al, 1994) is a cognitively based method aimed at uncovering usability issues by following the sequence of actions a user would take to perform a specific task, and by analysing at each stage how successful the user would be in performing the action correctly. The method takes a task-oriented perspective, in that it



considers the goal structure and the way goals are addressed in completing the task. At every stage the interface is evaluated by set questions to determine whether or not it provides the necessary information for the user to successfully continue with the task, and what feedback the interface provides to the user. The analysis of user actions is done in terms of success and failure stories. Cognitive Walkthroughs concentrate primarily on ease of learning, on the grounds that if something is easy to learn then it will also be easy to use and thus have fewer usability problems. This perspective is justified by the fact that users tend to learn features of an interface as they need to, rather than all at once. Therefore, ease of learning is seen as essential to interface usability. This can be very effective in showing the kind of problems that a user may have at a given point in the interaction.

GOMS (Card et al, 1983) is a cognitive modelling technique that is based on the idea of the human as an information processor. GOMS stands for Goals, Operators, Methods, and Selection and is based on the premise that a user achieves goals by breaking them down into sub-goals which can then be separately achieved. The Operators are the ways available to accomplish the goals, Methods are defined sequences of operators and goals, and Selection rules determine how to choose between more than one method (John and Kieras, 1996). Tasks are seen as something that people do to achieve goals. The emphasis is not just on the physical aspects of interaction, but also on mental processes — for example, what the user has to know or remember. Since there is more than one way of applying the GOMS model, it can be very versatile and useful. Varieties of GOMS address goal hierarchies, working memory loads, schedule tasks, lists of operators, and production systems. GOMS is not a complete answer to cognitive modelling, and has various drawbacks. The method depends on the interaction being very structured, having easily identifiable goals and methods, with clear selection rules, and there is no guarantee that all user goals will be identified. It is best at dealing with rational, expert and error-free behaviour, so is less useful for dealing with novice situations. In design it tends to be reactive, in that it assesses the quality of a design, rather than being proactive in guiding the design.

The interface to the robotic arm was first analysed using CMN-GOMS which is based on the Keystroke Level Model developed by Card, Moran and Newell (1983). This version of GOMS was chosen for its simplicity of learning, and has a strict goal hierarchy, with

methods represented in a program form as series of steps which must be performed in sequence. The analysis was then taken down to a CPM-GOMS level in order to examine more fully the cognitive, motor and perceptual aspects of the interaction. CPM stands for both Cognitive-Perceptual-Motor, and also Critical Path Method, and is based on the assumption that tasks needing different processes within the Model Human Processor as put forward by Card et al (1983) can be performed in parallel. To use CPM-GOMS, a CMN-GOMS analysis is first done to determine the goal hierarchy and methods in order to obtain the basic perceptual, cognitive and motor operators, which are then expressed using schedule charts.

In contrast to the methods discussed above, PUM is still at the development stage, and is therefore not widely used. It attempts to bring together aspects of user modelling and system modelling into one representation (Blandford & Young, 1996). A description of the knowledge that the user needs to operate the interface successfully is written in an Instruction Language (IL), which is then compiled by the cognitive model in order to give an indication of problems that might arise. The IL description is a formal description consisting of: the conceptual objects that the user manipulates; relations between defined object types; a device description including commands, the initial state, and information displayed to the user; and user knowledge in terms of conceptual operators, initial knowledge, and user task.

Z (Spivey, 1989) is a formal specification notation based on set theory and first order predicate logic, using schemas, which are collections of named objects with relationships specified by axioms. These schemas can be built up to define large specifications. Z is one of the most widely used formal notations in software development, is currently undergoing international standardisation, and is the most formal of the five techniques studied here. The mathematical base of Z means that it can be considered to be unambiguous, which makes it a powerful notation for communicating ideas and concepts, and can allow the designer to gain an insight into the structures and relationships that are of importance, and to manipulate those relationships and examine the implications of change.

The five approaches represent some of the more formal modelling techniques, and are not intended to be definitive. Indeed, other approaches, for instance Petri Nets (e.g. Bastide &



Palanque, 1990) and Task-Action Grammar (e.g. Schiele & Green, 1990), would have been equally applicable. Other techniques such as syndetics (Duke et al, 1998) and Cognitive Task Analysis (May & Barnard, 1994) were not used, since there is little or no published information on their application to interface analysis.

### **2.3.3 Analysis**

For a full description of the final analyses, please refer to Appendix J. By examining the techniques in terms of their overall coverage and application, as well as their ability to identify issues relating to multi-modal usability, it is possible to assess the importance of the underlying representations and methods. A method differs from a representation (Hinchey & Bowen, 1995) in that a representation is merely a way of representing what is happening in an interaction, whereas a method is the way in which a process is conducted, and may incorporate a representation or notation which is applied according to defined procedures. Thus it would be expected that a purely representational technique would uncover fewer issues than one with a method, since the method would provide guidance on how to apply the representation for best results, as well as guidance on what kinds of issues to look out for. Different representations and notations would also be able to represent different issues, thus affecting their coverage, and differing methods would take differing approaches, again affecting the number and type of issues uncovered.

These results are, inevitably, subjective, given that they are confined to one interface, and the analyses performed mainly by one person. However, the controlled nature of this case study has allowed important qualitative conclusions to be drawn and, by confining the study to one interface and one analyst, the difficulties highlighted by (Gray and Salzman, 1998) in other comparative studies can be minimised. Previous studies have suffered in that the analyses have been conducted by people expert in their use. This study, in contrast, involved an analyst who was a comparative novice in the use of all the techniques investigated. Thus issues relating to the learnability and usability of these techniques can also be considered.

It is not possible to compare how long each analysis took to complete due to the particular nature of the study. Continuous exposure to the interface meant greater familiarity with



certain areas of potential concern, and the need for consultation with an expert on aspects of the PUM and Z analyses also affected the overall times.

### *2.3.3.1 Multi-modal Issues*

The analyses of the interface using the five different approaches show that they were generally unable to identify and represent multi-modal issues such as the possible conflicts between voice and gesture input.

Overall the STN diagram was task-based and represented the task in a sequential manner concentrating more on the states of the interaction than the process or means of interaction, or user capabilities or constraints. This constrained the issues it could address, in that it could look at input and output issues only with regard to sequence and steps, rather than the modalities involved, modality combinations, and issues relating to the duration or ordering of modalities.

The CW's emphasis on goal-structure and previous knowledge, rather than actual means and method of interaction, meant that this analysis did not cover important aspects of multi-modality such as modality selection. The system is only looked at from the user's perspective, and on a staged basis. Furthermore, only novice users are considered, and the system is assumed to be non-dynamic, with time for the user to choose options without the system state changing. Therefore it seems that the CW appears more concerned with the user's reaction and the next state than with the actual mechanics of the interaction, or the problems of actually moving the interaction forwards.

The CMN-GOMS part of the analysis could not deal with multi-modal issues at all, due to its hierarchical and sequential structure, and was unable to do much more than identify input methods and the sequence of operations needed to use them. CPM-GOMS was able to provide a fine grain of detail for reasoning about precise operations in a particular modality, and to represent and examine modalities at a very low level of detail. The schedule charts clearly bring out the differences between the methods for the voice and gesture input devices, and are thus useful for comparing them against each other. .

PUM was unable to shed much light on multi-modal usability issues since it dealt with interaction in a procedural and knowledge-based manner. There was no way of showing

complexity of interaction or of defining the decision process between two choices of input device, except at the most basic of levels. The filters describing that the use of the gesture or the voice devices was inappropriate could be expanded to show the exact reasoning behind the choice of one input device over another. However the increased complexity of interaction when two input devices are available and the user can use one or other or both was not clearly shown, and PUM could not allow for timing issues.

Z was similar to PUM in its inability to handle multi-modal interaction except at a trivial level, since it is based on states of operation rather than wider issues involving choice and complexity. The multi-modal input devices proved to have little impact on the usability analysis, since the operations were merely distinguishing between the two available forms of input, and describing how to use them. The Z notation is better able to represent the states and movement between the states rather than the modalities which convey the information necessary to change the state, is unable to give insight into issues of choice and selection between the devices, and can not handle issues of timing. However, since Z is purely a notation rather than a method for analysing any usability issues, let alone those of a multi-modal nature, this is hardly surprising.

All of these techniques were goal-based to some extent, and were concerned more with the ordering of the interaction and the knowledge possessed by the user rather than the nature of the interaction. Their limitations with regard to multi-modal usability centred around the same issues: being unable to represent in detail the actual means and method of the interaction, or deal with the increased complexity of the interaction, in terms of loading and timing. The representation used by the STN was unable to represent modalities adequately, and only CPM-GOMS was able to show the modalities in any detail. Methodologically, CPM-GOMS was best able to reason about the modalities since it concentrates on low-level interaction issues of timing and complexity. The other approaches were hampered either, in the case of STN and Z, by having no method or set of guidelines to apply, or, as in the case of CMN-GOMS, CW and PUM, because the method was looking for different kinds of issues, not directly concentrating on multi-modal issues.

In order for multi-modal usability issues to be apparent, therefore, it seems that a greater emphasis is needed on the actual identification and representation of the modalities



involved in the interaction, and that their constraints be clearly identified, in terms of timing, combination and ordering. These are issues that, as later chapters will show, are crucial to an understanding of multi-modal usability, and necessary to any approach purporting to analyse such systems. The inability of the approaches to handle such issues emphasises the current gap in techniques for assessing multi-modal usability, and emphasises the need for a thorough understanding of the relevant issues, which can then be applied to the structured usability assessment of multi-modal interfaces. This is addressed in the next chapter which closely examines what is meant by “modality” and investigates the properties relevant to multi-modal usability.

### *2.3.3.2 General Usability Issues*

All the analyses as shown in Table 2-1 revealed general usability issues to a greater or lesser extent, depending both on the level of support given and the analyst’s inherent skill. STN and Z are obviously the most heavily dependent on the analyst having relevant insights, because they are not normally applied to usability evaluation, being forms of representation rather than methods.

Although it was not developed in order to facilitate usability assessment, the STN highlighted important usability issues relating to the sequencing, for example the lack of any means of returning to the **Direction** menu without going back to the original state, and possible redundancy of operations, for example the **Continue** option, which appears to add little to the functionality of the arm and instead creates more states for a user to navigate through. This reflects the use of STN in software engineering to show the states and transitions of a system. Task-based information was not explicitly covered, and more complex usability issues were not made apparent through using this representation; in particular, possible causes of user error are not revealed.

The CW was best able to deal with issues of misleading option names, and the importance of feedback to the user, since the technique is aimed at describing the behaviour of novice users, to whom such issues are important. The main problem uncovered using this method was that of ineffective or misleading labelling of options. By assuming the perspective of a new user, the walkthrough was able to determine which labels could be considered effective and which needed changing. For example, there is a problem in determining the difference

between the options **MoveArm** and **Move**, as the user may not know at this stage which one concerns moving the arm as a whole (**MoveArm**), and which will move the individual joint named arm (**Move**). The directions of movement of the arm relate to the user, but the user will possibly be unaware of this at that time. The CW acknowledged the existence of failure explicitly, but did not follow through to consider the consequences of such failure, merely stating that failure could possibly occur at a particular point and giving reasons why. Another drawback was that the interaction was considered stage by stage rather than as a whole, thereby possibly missing certain issues relevant to the interface overall. However, overall this method uncovered the largest number of usability issues, showing that the stage-by-stage approach can be very effective.

The use of CMN-GOMS allowed the interface to be examined in terms of how well structured it is, the naming of options, and could identify problems relating to the length of interaction, the redundancy of operations, and the lack of short-cuts. For example, in order to move the arm to a specific position, the user has to decide whether to move the arm as a whole, or to move an individual joint. If the user decides to move just a joint, the user then has to specify which joint, the direction of movement of the joint, to adjust the speed of the joint if necessary, start the joint moving, stop the joint moving, and then decide whether or not to continue moving that same joint or to move either the whole arm or another joint. In either of the latter choices, the user would have to return to the main menu. There is no other way of going back to the selection of joint stage. However, CMN-GOMS was unable to represent behaviour other than that of expert users, and was unable to deal with failure, or users incorrectly using the interface to achieve a goal. CPM-GOMS in this analysis was unable to uncover many usability issues, concentrating as it does on the critical path of interaction, and issues of overload and complexity, which were not problems for this interface.

PUM was able to show a wide range of problems and provide reasoning as to their significance. Amongst other issues, the analysis made clear the lack of mapping between what the user wanted to accomplish, and how the device would attempt to accomplish it. The strong divergence between the user's domain task and the actual device task was clearly identified by the PUM analysis almost immediately due to the standard PUM principle of focusing on domain tasks and how they are achieved using a device. This is



mainly because the device operates in three dimensions, and there is a clear difference between how the device works and how a user might manipulate something in three dimensions. However, again, user-failure was an issue insufficiently covered, and the analysis assumed that the user would be able to apply their knowledge consistently in using the interface, which might not be a typical user's profile.

The Z specification, being mainly device oriented with little explicit representation of the user in this particular specification, was unable to cover a wide range of issues. It was, however, able to shed light on the problem of ordering, the inability of the interface to allow backtracking, and the existence of some actually unavailable choices. The **Move** and **MoveArm** commands, and **Go** and **Continue**, were shown to be sufficiently similar as to question the advisability of leaving them separate. The fact that the interface is strongly ordered, in that there is no way for a user to backtrack and change the part of arm or direction selected without having to go back to the initial menu and starting the whole process again, was brought out by the specification. Despite not covering a task in detail, Z was able to show how operations might be more effectively and simply represented.

In order to be widely acceptable, a technique has to be seen to cover a wide range of usability issues, to justify the time spent in applying it.

The CW, GOMS and PUM methods were successful in that they were able to uncover a number of usability issues, more than the STN and Z approaches which are purely representational. The methods had different focuses: PUM and GOMS are concerned with both the structure of the system as well as certain user aspects, whereas CW concentrates more on the user's perception and experience of the system, rather than the issues inherent to the overall structure.



Key

issue identified



CMN-GOMS and CPM-GOMS are included under the heading of GOMS

PROBLEM	STN	CW	GOMS	PUM	Z
Long sequence of operators to move arm					
Inability to backtrack					
Difficulty of choosing between MoveArm or Move					
Lack of short cuts					
Continue versus Go: Continue seen as redundant					
Confusion over joint called Arm					
Gesture input with twice as many operations as voice because dependent on cursor movement					
Problem if head moved to look at arm while gesture system operational may be interpreted as a command					
If user pauses in middle of saying "Move arm"...					
If user engaged in conversation...					
Lack of feedback after Move Arm selected, no indication that whole arm is to be moved					
Problems of determining left and right, especially when arm contorted					
User cannot check direction choice until arm starts to move					
Time taken to interact with system to stop arm					
Similarity between moving joint and moving whole arm					
Illegal options					
Mismatch between way that arm works and way that user would move arm					
Not clear that End returns user to main menu					
End having two meanings					

**Table 2-1: Summary of usability problems uncovered**



Both STN and Z could be used to reason successfully about some aspects of usability, and the Z analysis was able to highlight an issue missed by other analyses. These two notations were able to represent information relating to the flow of the interaction, the ordering of the system, and redundant or illegal states. Given that they are techniques used to develop the structure of systems, this is hardly surprising, but what was important was that they could also be used to successfully identify related usability issues. This illustrates to what extent any representation of an interface can be useful in usability assessment, as opposed to specific representations designed for such use. This is, however, dependent on the analyst having a thorough understanding of usability problems, and an ability to identify important issues. This implies that significant usability assessment may be possible by utilising representations intended for other uses, such as software design, and highlights the level of craft skill needed in usability analysis.

There are many factors which contribute to a successful usability analysis. That the methods uncovered more issues than the representations shows the importance of guidance and support, and that the representations could be used to reason about usability suggests the benefits to be obtained from clear representations. However, the methods and representations have other issues associated with their ease of application which have a significant impact on the ability of analysts to use them to identify usability issues. This is discussed below.

### ***2.3.3.3 Application of Approaches***

For a technique to be useful and to be exploited by designers of interactive systems, it must be clearly defined and straightforward to learn and apply. No matter how powerful a technique might be, if it is difficult and time-consuming, designers will prefer to use other methods (or none). The five techniques are very different in terms of level of support they provide for the analyst, and vary greatly in their ease of use. GOMS, PUM and CW, being designed for usability assessment, have procedures defined for usability, whereas STN and Z, being used in software engineering for purposes other than evaluating interfaces, are less constrained. Since STN and Z are representations they provide no support. However, GOMS, CW and PUM also rely significantly on the ability of the analyst to identify areas of

potential concern, demonstrating that even methods which purport to support the analyst and focus on user behaviour are dependent to some extent on the analyst's inherent skill.

STN, as a diagrammatic approach meant for other uses within software engineering, has no explicit support for the usability analyst. However, the structure of the diagrams with nodes of states and arcs of operations makes it straightforward to learn, easy to apply, and quick to assess. Only one or two iterations were required before a complete representation of the structure of the interaction was obtained. By explicitly representing the interaction choices in a diagrammatic form, it was straightforward to follow through a path of interaction corresponding to a given task, and to reason about the usability issues raised.

CW seems well-defined in that there is a clear set of questions, and that specific information about the users and task has to be provided from the beginning. Having to specify a task immediately concentrates the analyst on the purpose of the interface. The questions mean that the walkthrough is done in a structured manner and that at every stage searching questions about what the user was likely to do have to be answered. It can be done straight through in one session, avoiding lengthy iterations. However, the questions can be answered in many different ways, and quite often points were found that did not fit neatly into one question or another, but which were still necessary to the evaluation. The CW relies on the skill of the analyst in answering the questions correctly with reference to the potential user, without necessarily knowing much about that user. Thus although it is a straightforward methodology to learn, the analyst needs experience in order to determine how and when to answer the questions.

With CMN-GOMS, the notion of goals, operators, methods and selection rules means that there is a clear structure to the analysis. The analysis can be performed in a structured and constrained manner. First, high-level goals can be determined, then the analysis taken down level by level until the final operators are determined. This is an iterative approach, but one which allows for clarity and insight at all levels. The emphasis on the task allows the interface to be assessed as to how well that task is supported. However, there are problems at the CPM level, due to lack of knowledge about the internal workings of the brain. At this level, the analyst has to use best judgement since various alternatives can be considered. There is also difficulty in determining the relative merits of one particular goal structure



against another. This again is left to the judgement of the analyst. The main difficulty in learning this methodology is in learning how and under what conditions to apply it, rather than any inherent level of difficulty.

PUM uses the highly structured and constrained Instruction Language (IL) as a means of describing the interface, which helps the analyst to be consistent, accurate and complete. Having to specify prior knowledge, and what can and cannot be done with regards to a given task, means that the interface can be comprehensively covered. However, the IL can be hard to learn and to understand, and having to define everything means that the IL can be difficult to write. The IL allows for imprecise definitions, but even with this the technique is difficult to learn. There are also problems in actually starting the analysis. This methodology involved many iterations, with some of the analysis only becoming clear towards the end; a lot of effort was expended before final results were obtained.

Being a specification notation, Z has no support for usability assessment. It resembles the PUM IL, in that every property has to be specified, and suffers from the same kinds of problems: precise definitions are not always easy to determine, the analysis took several iterations before it was complete, and there was difficulty in deciding where to start the analysis. The notation requires a certain amount of training, but, given that it is an established technique used within software engineering, some analysts may already be experienced in its use.

In the light of using the five techniques assessed in this case study, STN and CW are the most straightforward to apply, due to their simple representational style, with GOMS next, and PUM and Z hampered by the need to learn a specific notation and the degree of formalism necessary. GOMS, PUM and Z required the analyst to learn the particular notations and means of application, whereas the STN and CW allowed for more freedom of representation. There is also a greater level of implicit craft skill in applying these techniques effectively than is generally acknowledged, which will be examined further in chapter seven.

This suggests that effective techniques are those which utilise a clear representational style, and allow for straightforward application, with well defined procedures and parameters. Educating the user in the types of issues to be identified would capitalise on inherent craft

skills and give the analyses more depth. These are issues that must be considered carefully in any approach developed to handle aspects of multi-modal usability, as will be discussed in future chapters.

## **2.4 Summary**

This chapter has introduced the concept of multi-modal interaction and discussed some of the proposed usability benefits. Problems relating to existing research in this area have been examined, and a lack of theoretical understanding highlighted as an issue for concern.

By means of a case study involving the application of five modelling approaches to a portion of an interface of a robotic arm, the problems of five existing techniques in modelling aspects of multi-modal usability has been shown. From this study it is apparent that modalities need to be clearly identified and represented, and their constraints, in terms of timing, combination and ordering, understood in greater detail, in order that such systems can be successfully modelled. The next chapter will investigate in more detail what is meant by multi-modal usability, drawing out those properties of most relevance.

The study has also highlighted important general issues concerning the coverage and application of approaches, including issues such as representation, guidance, and application. These will be considered in more depth in later chapters relating to the development of a methodology for assessing multi-modal usability.



**PAGE**  
**NUMBERING**  
**AS ORIGINAL**

# **3. MODALITY DEFINITIONS AND PROPERTIES**

## **3.1 Introduction**

The previous chapter discussed the fact that the use of more than one modality does not imply greater usability unless those modalities are used in a sensible manner, and identified the lack of overall theory in this area to guide designers and analysts. The growth of virtual environment (VE) systems has made this situation even more critical; Mills & Noyes (1999) detail the need for research in the benefits of redundancy of input and output, the extent to which users can manage parallel inputs, and the need to ensure that tasks are carried out by the most appropriate type of input. This chapter addresses this lack of theory, by assessing existing work and drawing out those issues of relevance to multi-modal usability.

Modality is a term that is subject to confusion and distortion in the HCI literature, with various definitions used concurrently. A selection of modality definitions are critically assessed here, in order to determine what is needed from a definition of modality within an HCI context. Defining what constitutes a modality, although helpful in aiding the understanding of the exchange of information in an interaction, is not enough. There must also be a way of applying such a definition in a meaningful way for it to be relevant and useful, and an understanding of the factors which impinge on multi-modal usability. To this end, various existing theories of multi-modal interaction are examined to determine the critical aspects of multi-modal usability. The importance of cognition is assessed by examining four cognitive models, leading to a set of important cognitive properties.

## **3.2 Modality definitions**

Before any research can be done in the area of multimodal usability, a precise definition of what exactly constitutes a modality needs to be determined. If modalities cannot be identified in the interaction, then they cannot be assessed for their usability issues. At present there are several, often competing, definitions of modality within the HCI



community, ranging from a systems-centred, device dependent perspective to a user-centred, sensory dependent perspective, with the definitions depending very much on the particular background and needs of the individual proponents. As Stenning & Tobin (1997) observe: "Psychology uses the term sensory modality for what computer science uses the term information medium." Roth & Hefley (1993), in their lucid discussion on modality terminology, state: "*modality* distinctions can be made along human sensory channels, technology used to store or present the information, or even computational techniques used to present the information".

The variety of definitions available means that when people from different backgrounds come together, such as in the design of interactive computer systems, confusion can result, since they are identifying different things by the same name. It also means that research from different disciplines cannot be integrated due to a lack of a common vocabulary or framework of ideas. As Ranganathan (1962) observes "...it is, on the whole, better for any special discipline to set up its own special, carefully defined terminology, rather than to use common terms. The emotional and other associations of words often unconsciously detract from accurate and exact thinking." Alty (1991) agrees with this view, stating: "we certainly have a capability for inventing new terms even if we do not understand what they mean or how they relate to each other (if at all). This is rather unfortunate since a clear terminology would help us to communicate our results more clearly and assess the relevance of other work to our own."

Modality is not a straightforward term to define, because of the difficulty of distinguishing between the form or mode of the information, the content of the information, and the manner or media in which the information is expressed. For example, consider the use of spoken language and text to interact with a computer system. The device used, either a microphone or a keyboard, will carry apparently the same information, but that information will be formulated in differing ways, and thus be quantifiably different, as Cohen (1984) has observed when comparing spoken and written language. Thus, different information forms are expressed and received in different ways by a human and system, and the variety of definitions for modality reflect this.

Sense-based definitions concentrate on the senses used by the user to input information or to receive output from the computer. Device-based definitions concentrate on the devices that input information into the computer and how the computer displays information. Mode-based definitions concentrate on how that information is transmitted or received, and form-based definitions concentrate on how that information is presented.

There is a distinct difference between these definitions. Mode-based definitions regard modalities as behaviours which are carried out, in contrast to sense or device-based definitions which regard modalities as physical artifacts which process information. Information form definitions of modalities regard modalities as informational artifacts, where it is not how they are physically processed but the structure of the content of the information that constitutes the modality.

Thus there is a split as to whether a modality is a sense, a device, a mode of behaviour, a form of information, or a combination of two or more of these things. This is not to say that definitions which are different to each other cannot be useful, but each has certain advantages and disadvantages when used to define modality within an HCI context. They are assessed on this basis in the review of the different types of modality definitions current within the HCI community below.

### **3.2.1 Sense**

Modality can be defined from a psychological and user-centred perspective in terms of the human sense by which it is expressed or received. For example, a modality could be considered as tactile or visual (Krapichler et al, 1999). Takeuchi & Nagao (1993) have developed an experimental system which integrates speech dialogue and facial animation in order to investigate the effectiveness of facial expressions in interaction. Within the context of this research, they define modality as “the sense used to perceive signals from the outside world”, with examples of modality including sight, hearing and touch. Purchase (1999) has examined the issue of modality from a semiotic perspective within the context of a classification system for multimedia interfaces. Modality is defined as equating to the sensory capability of the user, with Purchase’s classification concentrating on the modalities of vision and sound.



The definition of modality in terms of the human sensory capabilities has the advantage that it is very easy to define, since there are only three possible categories: visual, acoustic and haptic (ignoring taste and smell). However although this is a perfectly reasonable definition, it is less suitable for use within an HCI context since the definition is so broad that it makes any contrast or comparison between modalities essentially irrelevant.

### **3.2.2 Device**

There have been a number of studies within computer science that take a systems perspective (see (Verjans, 1995) for an in-depth discussion), with modality defined in terms of the physical device through which the communication passes. It is hardly surprising that the computer science approach to modality concentrates on the physical input and output devices through which the information exchange occurs since computer science has historically been device-centred. A keyboard, mouse, monitor or microphone could all be considered as modalities within this definition. The modality is therefore unconnected to the physical structure of the human and how the human processes or expresses information, or with the structure of the information, and is only concerned with the device through which that information is passed in the interaction between human and computer.

Defining modalities in terms of the device used has the advantage of being obvious and immediate, but is restricted in that it is very specific, any categorisation has to be adjusted to handle new devices, and it does not allow for easy generalisation over classes of device. It is also biased in terms of the system, and ignores the involvement of the human in the interaction process.

### **3.2.3 Mode**

Modality can be defined in terms of a mode of behaviour of either the system or the user as Alty (1991) recognises. Speech, gesture and typing can all be considered as modalities, since they are ways of doing something. This definition of modality can also be applied in more abstract terms such as “exploratory mode” and “task action mode” (Kaur et al, 1999). In this definition the characteristic quality of the modality is not the sense in which it is expressed or perceived, but the type of behaviour used. Dowell et al (1994) have used this kind of definition of modality, following on from the work of Frolich (1993) and Hutchins

et al (1985), within the context of a multimodal multimedia design space. They define modality in terms of the metaphor of interaction: virtual world mode, where the user is immersed within the world in the form of direct manipulation upon objects; virtual partner mode, which uses the conversational metaphor in interacting with knowledge rather than objects; and mixed mode, which uses both modes.

The definition of modalities in term of modes of action restricts the definition to particular ways of doing things. This can be difficult to classify and identify, since users use different ways of doing things depending on several factors such as motivation and experience.

### **3.2.4 Form**

Other definitions define modality in terms of the properties of the information conveyed. Stenning & Tobin (1997) propose a definition of modality within the context of a cognitive theory of modality assignment, with special reference to graphical representations. The modality is seen as a particular information form, or the manner by which the information is represented, as opposed to the medium in which the information is represented. An example is given: “Braille and printed text are the same modality of information in different media (tactile and visual), whereas embossed diagrams and Braille contrast in their modality, one being graphical and the other linguistic.” However, Stenning & Tobin (1997) acknowledge that defining modality is not straightforward.

Cohen (1992) defines modality as “the syntactic, semantic, and pragmatic properties of the signal - in other words, on how the signal functions to communicate” within the context of an integrated interface, Shoptalk, which combines the modalities of direct manipulation and natural language. A distinction is made between the medium by which the signal or modality is carried, and the content of the signal, which comprises the modality. The modality is seen as that which communicates information, rather than the means by which that information is communicated.

Other definitions have defined modality in terms of the information forms represented by the computer; for example Neal & Shapiro (1991) list the following output modalities used by CUBRICON: “color graphics/pictorial displays, tables, histograms, written NL prose, spoken natural language, and fill-in-the-blank forms.”



Bernsen (1995b) defines modality within the context of a comprehensive theory of modality as “mode or way of representing information to humans or machines in a physically realised intersubjective form.” An important distinction is made between what a modality might be and the medium by which it is conveyed, with modalities defined as “physically realised ‘ways of representing information’” (Bernsen, 1995a), thus attempting to avoid confusion between the information which is being conveyed and the method of conveying that information. It is the form of that information, as represented in a physical manner within the means of conveyance, which constitutes a modality to Bernsen. A device-centred approach appears to be taken, in that the ways of representing information are seen as either by the computer or to the computer, although Bernsen tries to abstract above that by “addressing forms of human-to-machine input representations of information irrespective of how these forms, or modalities, map onto the devices used for inputting them” (Bernsen, 1995c).

The information form definition of modality potentially avoids any system or user bias, and may be able to overcome some of the problems posed by strictly device-oriented or human-oriented definitions. However, defining which particular information forms should be used to categorise modalities may be problematic.

### **3.2.5 Other Definitions**

Definitions have also been proposed which combine one or more elements of the definitions above into a compound definition. Coutaz et al (1993, 1994b, 1995) have investigated modality and multi-modal usability exhaustively, and as a consequence have defined modality several times, in different ways.

In the description of the experimental development of MATIS, a Multimodal Airline Travel Information System developed at the University of Grenoble (Nigay et al, 1993), modality is defined as a combination of device and form. This is an interesting definition because it makes a clear distinction between the physical device used, and the actual language used through the device, and states that it is the two together which constitutes the modality.

Within the description of the Multi-Sensori-Motor (MSM) Framework Coutaz et al (1993), a dimension space proposed to assist reasoning about multi-modal systems, modalities are

defined as “the temporal, virtual, or physical link that makes the exchange of information possible between communicating entities...” Communication is through sensors and effectors, the sensors being devices that take information in, and effectors being devices that transmit information. These sensors and effectors may be grouped together into clusters associated with particular processing units. Therefore, the definition is concerned not with the information itself, but with the pathway through which this information is processed. Coutaz et al (1993) give as an example of a human communication channel the retina (physical input device) and the associated visual subsystem for processing the input. It is the two together, the input device and the processing device, which define the channel. An example of a channel in a computer would be a mouse taking input, which is then processed by the central processing unit. This is therefore a definition which combines sense and cognitive resource.

Within the context of the CARE properties (Complementarity, Assignment, Redundancy and Equivalence) for assessing the usability of multi-modal interaction, Coutaz et al (1994), modality is defined as “an interaction method that an agent can use to reach a goal” with what is meant by interaction method being left undefined. This gives little insight into what a modality actually is, rather than what it is used to do. However, in a later, related, paper (Coutaz et al, 1995) another definition has been proposed where modality is defined as either “a human sensory capability or...a computer physical device...At a higher level of abstraction, a modality is viewed as a representational system...A modality may also denote the coupling of a device with a particular representational system”. This definition is combining elements of the previous definitions into one which attempts to make a modality four different things, depending on the level of abstraction: a sense, a device, a form of information, or a form/device pairing.

In their work on the Event Model of human-system interaction, Duke & Harrison (1995) use interactors as abstract structures to link rigorous aspects of software development with more user-oriented aspects of interaction, and modality is defined as “the means by which interaction takes place between user and system”. This is refined into a two-pronged definition similar to the CARE definition proposed by Coutaz et al (1995), where “modalities correspond to human capabilities such as speech, vision and gesture, but in system development we may also wish to classify interaction techniques based on devices



and languages”, with modality defined in terms of both sense used and as a device/language couple.

A definition that couples the device and language makes a clear separation between the device and the language, allowing the content of the information and the device to be distinguished. This kind of definition can lead to problems because it is very specific, related to a particular device and a particular language, making it difficult to generalise over different devices except at a very high level of abstraction. This potentially makes it less useful with regards to assessing combinations of modalities and it may be biased more towards the input side of the interaction.

A definition of modality which combines the device and the associated cognitive processing path is an attractive idea, since it combines elements of both the user and the system. However, it is too specific in the form presented, and is biased towards the output of a system rather than being able to deal with both the input and output parts of the interaction.

Defining modality in several different ways in combination is unsatisfactory for several reasons. One problem is that it appears to imply that a human sensory capability can be considered as equivalent to a physical computer device. Obviously, this depends very much on whether human sensory capabilities are defined in general terms such as in gesture, speech or touch, or more specifically, as in eyeball-tracking, keyboard input and mouse operation. However it seems unreasonable to regard a sensory capability as having the same attributes and characteristics as a device. Another problem with this kind of definition is that distinctly different things appear to be regarded as a modality, depending on the level of abstraction. For example, defining a modality as a device, or a representational system, or a combination of the two, and also as a human sensory capability, appears to confuse the way of expressing information with the way in which information might be perceived or communicated.

### **3.2.6 Derived Definition Elements**

In order for a definition of modality to be useful within the field of HCI it must aid understanding of the interaction and allow insight into usability issues surrounding the interaction. Since the interaction between a user and a system has many different

components, a definition that relies on narrow categories, either in terms of human sensory capability or device, is limiting. The human sensory channels are so broad that defining a modality in these terms will be of little practical use. A definition that relies solely on technology is incapable of generalisation, and can quickly become out of date and inflexible. There is thus a need for a way of defining modality which is more specific than that provided by psychology, and less device-dependent than that used by computer science, in order to convey the richness of the interaction in a flexible and usable manner.

It is the human for whom modality, whether for input or output, holds significance. To the computer, the question of how the information is received is largely irrelevant, once technical problems associated with the processing of the data is overcome. Modality is therefore primarily a human concern, as is usability, and hence any definition of modality should include the human sensory capabilities, since it is with these that a human interacts with a computer. Since by themselves these are too narrow to be of effective use, it should also include other elements, following the approach taken by other definitions, which have attempted to classify modalities in terms of combinations of elements. Combining elements is an attractive way of resolving some of the problems involved with defining what a modality is, since it can overcome some of the limitations of single element definitions, however there is a question as to which elements should be included.

The physical input and output devices are too limiting and inflexible to be used successfully as elements for a definition of modality. Behavioural definitions are too wide-ranging, and depend too much on the individual users.

The form of information, as expressed through a device, is a suitable candidate, since it allows for the types of information that are processed by the human senses to be differentiated. This will, however, depend on how the information might be categorised, which might take several differing forms.

Therefore it has been established that current definitions of modality are unsatisfactory for several reasons, and that any new definition must include the human sensory channels as well as other elements, possibly including information form. This chapter continues by examining properties of importance to multi-modal usability. A new definition of modality will be presented in the next chapter.



### **3.3 Multi-modal properties**

The previous section has investigated various issues surrounding the definition of a modality. Whatever might constitute a modality, there is a need for an understanding of the usability properties of multi-modal systems. In this section, five different approaches to reasoning about the usability of multi-modal and multi-media systems are examined and assessed in order to establish the problems and confusions with current theories. These are then used to derive critical properties of relevance to multi-modal usability.

#### **3.3.1 Modality Theory**

Bernsen has investigated modality with the following objectives (Bernsen, 1995a): establishing a taxonomy and a set of related categorisations of generic output modalities; establishing sound foundations for describing and analysing modality outputs; establishing sound foundations for analysing input modalities and computer interfaces; developing a methodology for applying the results of the above; and using this work to build design tools to support usability engineering.

Although this work is as yet incomplete, Bernsen has postulated two modality taxonomies, one dealing with input modalities (Bernsen, 1995c) and the other dealing with output modalities (Bernsen, 1995b). The output taxonomy will be examined in some depth, since this was the first part of the taxonomy to be compiled, and clearly shows the potential problems associated with this particular approach.

Bernsen's output taxonomy (Bernsen, 1995b) is effectively a table, where modalities are classified according to four pairs of attributes and three media of representation. The four pairs of attributes are: linguistic/non-linguistic, with linguistic representations allowing the communication of general concepts; analogue/non-analogue, where analogue representations make use of "aspects of similarity between the representation and what it represents"; arbitrary /non-arbitrary, where non-arbitrary values are those which accord to an established scheme of representation; and static/dynamic, where static representations offer "freedom of perceptual inspection" (Bernsen, 1995b). The three media of

representation are graphic, acoustic and haptic, corresponding to the three human senses most often used in interaction. Forty-eight potential modalities are generated from these criteria.

Bernsen does not justify in any depth the elicitation of these attributes of modalities, instead regarding them as self-evident. It is therefore difficult to decide whether there are more attributes which can be considered to have importance in this area, and what leverage these particular attributes give to multi-modal usability.

Given that the modalities are generated from the attributes rather than by any other method, it follows that if the attributes are not correct or complete, then the validity of the taxonomy as a whole is questionable (Pankhurst, 1978), (Bailey, 1994). Bernsen claims that there are certain rows in this taxonomy which are unusable, which raises interesting questions about the correctness of the taxonomy overall. Bernsen removes the non-arbitrary representations of arbitrary analogue and linguistic representations from the table since they cannot be used, except at the expense of comprehension (Bernsen, 1995b). The taxonomy is therefore reduced to thirty modalities from the original forty-eight.

These thirty modalities are then reduced to twenty, (Bernsen, 1995a) by combining static and dynamic acoustic modalities, and static and dynamic haptic modalities, since at present Bernsen claims that they are not well defined interface features, and if they become more popular in the future they can be easily reinstated into the taxonomy. Bernsen is therefore taking a pragmatic view of what this taxonomy can achieve, and is trying to make it relevant to interface design according to what interfaces can currently achieve, and allowing room for new developments. However, this particular approach inhibits extensibility.

### **3.3.2 The Multi-Sensori-Motor Framework**

The Multi-Sensori-Motor (MSM) framework as postulated by Coutaz et al (1993) is a dimension space which assists reasoning about MSM systems through the analysis of their behaviour according to six dimensions. The framework takes a systems-oriented approach, defining multi-modal systems in relation to six dimensions, but integrates sections of the Interacting Cognitive Subsystems (ICS) (see Barnard and Teasdale, 1991) psychological model in order to relate the framework to the cognitive processes and attributes of the user.



These dimensions are: channel direction, number of channels along one direction, abstraction, context, fusion, and granularity of concurrency. The first two dimensions are connected to the idea of a “communication channel”, or groupings of input and output physical devices, in terms of what the channel does and how many there are. The other four dimensions are concerned with interpretation, as in the receiving and understanding of input, and rendering, or the communicating of output information .

The idea of abstraction is central to the MSM framework, since it is concerned with the processing of data from one level of meaning to another level of meaning.

Context determines how input is processed and interpreted and is deemed to be essential to the correct interpretation of inputs and outputs when more than one is being used at a time.

Fusion and its opposite, fission, are defined by Coutaz et al (1993): “Fusion refers to the combination of several chunks of information to form new chunks. Fission refers to the decomposition phenomenon.” They can be defined at various levels of abstraction: for example, the blending of information from one particular input device, up to the blending of information from several input devices, or vice versa. Fusion is complicated in that the trigger function has to rely on several variables such as temporal and structural proximity. There is also a distinction made between “eager” and “lazy” fusion, with eager fusion trying to “combine inputs as soon as criteria are met at the lowest levels of abstraction” (Coutaz et al, 1993), whereas lazy fusion postpones fusion until all the data is collected. The advantage of eager fusion is that it can quickly generate feedback, but it suffers from the necessity to backtrack to correct inappropriate fusion of inputs.

Concurrency, or parallelism, is the capacity of the system to handle multiple inputs or outputs whether this is at the physical level in terms of multiple input devices, or at the task level, where multiple commands can be issued simultaneously at varying levels of granularity. This is closely related to issues of fusion and fission.

The main criticism of the MSM framework is also its greatest strength: abstraction. The framework deals with some very important concepts with relation to multi-modal interaction, namely, the importance of fusion and fission, levels of abstraction of information, concurrency, granularity and the importance of context, but these are considered at such an abstract level that it is difficult to reason about actual modalities in

terms of which combinations are acceptable and which are not. The framework does not handle interference, except at a high level of abstraction, in that for example it does not say: “do not use mouse and keyboard together under these circumstances because of these problems”. Therefore, although the framework is dealing with issues essential to multimodal usability, this is at such a high level of abstraction that they become difficult to practically apply.

### **3.3.3 Multimodal Multimedia Design Space**

Dowell et al (1994) have developed a Multimodal Multimedia Design Space (MMDS) as a framework for reasoning about modality issues, composed of two components: a problem statement and a design space. The problem statement is made of four models: the user model, profiling the user of a system, which may also include device constraints such as how accurate a speech recognition device should be; the domain model, looking at the user’s task in terms of objects and their states, relations, and domain goals; the task model, describing the tasks to be performed for the user to achieve their goals, expressed in terms of the transformation of domain objects and the relations between those transformations; and the performance constraints, detailing how well the interaction is required to perform in terms of successful task completion and work system cost.

The design space is made up of a set of dimensions which characterise multimodal user interfaces, on a task and dialogue level. The task level uses several concepts in its design space to characterise multi-modal systems: distance, modality, medium, and cross-media integration. Distance has two associated concepts; semantic distance, or the correspondence of the task domain with its representation at the user interface; and referential distance, or how well users understand what is meant by what is expressed. Modality has three categories: virtual world mode, where the user is immersed within the world in the form of direct manipulation upon objects; virtual partner mode, which uses the conversational metaphor in interacting with knowledge rather than objects; and mixed mode, which uses both modes. Medium is defined as a way of representing information, for example speech, text and gesture. Cross-media integration is concerned with communication happening either sequentially or in parallel, using either one or more modalities.



The dialogue level has two dimensions. The dimension of the locus of initiative for the dialogue will rest with either the human or the computer, and can be driven by either or both depending on the system. The parsing dimension of the design space at the dialogue level looks at the interaction agents in terms of their semantic, syntactic and lexical parsing.

Further work (Dowell et al, 1995) has refined the framework by adding elements of the Stimulus-Central processing Response compatibility model (SCR) as proposed by Wickens et al. (1983). Essentially, this translates into four recommendations for use in multi-modal systems, which are summarised below:

1. multimodal presentation of information is better than solely visual presentation.
2. priming reinforcement is better than verbatim reinforcement.
3. generally, verbal information presented by priming reinforcement is more compatible with spoken responses than with manual responses.
4. generally, spatial information presented by verbatim reinforcement is more compatible with manual responses than with spoken responses.

The framework makes important points about integrating user models, domain models, task models and performance constraints, in order to provide comprehensive coverage of issues affecting multi-modal interaction. However, the dimensions and concepts which relate directly to multi-modal interaction, primarily in the task level of the design space are too vague and are not defined clearly enough. Cross-media integration in particular needs separating into its component issues so that matters relating to time, ordering and complexity can be better examined. The cognitive recommendations, although useful, are too narrow in focus.

### **3.3.4 CARE Properties**

The CARE properties (Coutaz et al, 1995) are used to assess the properties of multi-modal interaction and to direct the design of such systems. They are divided into System CARE, which are used to assess system aspects, and User CARE, which are used to assess user aspects. The aim of the CARE properties is to guide usability by providing a framework

within which any given design can be examined by predicting or explaining user behaviour, or by helping to elicit user requirements through the use of experiments.

The properties are: complementarity, where modalities can be used together to achieve a goal; assignment, where there is only one modality used for a specific task, either because there is no modality alternative or the system imposes a specific choice; redundancy, where modalities can be used at or near the same time to convey the same information; and equivalence, where no one modality is dominant and any modality available can be used by the user without restriction. These properties can be either system or user related, and it is stressed that there must be compatibility between the user properties and the system properties in order for a system to meet user needs appropriately.

Modalities are described as either simple or compound. Simple modalities are atomic at a given level of analysis, depending on what level of abstraction is used. Compound modalities are made up of more than one modality and can be either complementary or redundant, since: “complementary modalities each convey a distinct piece of information, whereas redundant modalities convey the same information” (Coutaz et al, 1995).

This theory was used to test practical considerations in the combination of modalities in the form of two built systems: MATIS, a Multimodal Airline Travel Information System; and Mediatel, a system which elicits needs and usage of modalities by future communication terminal users. These properties are useful, since the interaction is looked at in terms of both the system and the user, and important aspects of multi-modal interaction can be described. However, it suffers from being at a high level, and not dealing with issues of time, ordering or synchronisation.

### **3.3.5 Taxonomy of Representational Systems**

Purchase (1999) starts from the need to define multimedia, and to provide a sound way of defining and classifying multimedia systems in the Taxonomy of Representational Systems (TRS). A semiotic approach is taken, resulting in a three-dimensional model, in order to obtain a taxonomic framework which can then be applied to classify multimedia communication devices. The aim is to separate the data communicated from the device which does the communicating, in order to “focus on the importance of appropriate



matching of text with technology” (Purchase, 1999). Purchase defines text as the “physical realisation of a message formed in a semiotic system”, and defines a semiotic system as “an organisation of signs comprising a system of meaning”.

The three dimensions of Purchase’s model are signs, syntax and sense. The first dimension, that of signs, is defined with regard to established semiotic theory. There are three categories of sign: concrete-iconic signs, having a perceptual relationship with the concepts that they signify; symbolic signs, where there is no perceptual relationship to the concepts; and abstract-iconic signs, where the perceptual relationship is stylised.

The second dimension is that of syntax, or how objects are arranged in order to communicate their meaning. This dimension has six categories: individual, where a text (or form of information) occurs once; augmentation, where a text is given extra attributes such as colour; temporal, where the text is repeated; linear, where the text is changing and ongoing; schematic, where the text is located in a certain position relative to other texts; and network, where texts are made up of other texts and each separate text within the composite text can be followed by any other text.

The final dimension is that of modality, defined as corresponding to the human senses used in the interaction. Purchase uses only visual and aural, on the grounds that these are the most established and frequently used sensory channels currently used by multimedia systems. This is a reasonable restriction since this classification is aimed at such systems rather than at interactive systems as a whole.

From this three dimensional framework a taxonomy of thirty-six cells is developed, which Purchase relates to different media types in order to aid their classification. For example, the recording of a thunderstorm could be classified as an example of a linear aural concrete-iconic text, and a movie as an example of a linear visual concrete-iconic text could be.

Purchase’s framework is strongly influenced by the visual properties of multimedia. Established semiotic definitions are used for the definitions in the sign dimension, giving a legitimate and established base, although there may be confusion over having symbolic represent visual text as well as visual symbols, and it is not particularly clear what the distinction between abstract-iconic and symbolic sounds might be.

It is with the syntactic dimension that confusions and distortions occur, and the problems in translating a definition originally conceived for visual objects and representations into one that also works for sound becomes apparent. The main problem is that the categories are not mutually exclusive. Purchase is categorising according to time or space, rather than allowing both to be attributes of any particular text. From a visual perspective, the linear and temporal categories are more dependent on time than on position. With the linear category, interpretation is sequential over time. With the temporal category, a single concept is communicated, as in the individual category, but done over a length of time. Thus it seems that the temporal category is actually identical to the individual category, with the exception of time, and that the linear category is merely a sequence of individual occurrences.

With the augmentation category it appears that two forms of information are being represented at the same time, rather than just one, therefore suggesting that augmentation is not needed as a separate category. The strong influence of the visual properties of multimedia is apparent in the fact that the aural segment of schematic is defined in two different ways, suggesting that Purchase is as yet unsure of what is to be achieved with this category. The schematic category itself appears to be a misplaced symbolic category.

Purchase has made a good start on the classification of information forms, but is limited by being visually biased, with the current definitions of sound weak, and problems within the categories. The dimension of signs appears to have too many categories, some of which it is difficult to make meaningful distinctions between, and the dimension of syntax suffers from conflating the idea of time and spatial arrangement. This is further complicated by the inclusion of the category augmentation. As yet the framework is useful only as a way of describing and classifying texts, rather than in determining what kind of text is best suited to what kind of device. This is possibly due to the motivation for the taxonomy being the classification itself rather than the application of such a classification. Only the flow of information from a computer to a user is examined, and temporal issues are inadequately covered.



### **3.3.6 Derived Multi-modal Properties**

From the above work, four major inter-related aspects to multi-modal usability have been derived: identification and granularity, user and system, combination, and time and ordering. These themes are examined in detail below.

#### ***3.3.6.1 Identification and Granularity***

Identification and granularity are concerned with the identification of a modality in terms of how it can be defined and compared, and the level of granularity or abstraction at which it can be regarded. In the previous section it was established that modalities can be thought of in very different ways. However, regardless of what constitutes a modality, there must be a way of distinguishing between modalities, based on differing attributes and characteristics. The level of detail must also be established. What may constitute a modality at one level of analysis may not hold true for a different level, and some attributes and characteristics may depend on the granularity of the analysis. Granularity is a useful means of reasoning about concepts at an appropriate level of abstraction, allowing specific instances to be examined, but also allowing generalisations to be made. As such it is a valuable way of examining modality.

If modalities are to be defined by elements, these elements need to be precise, comprehensive and unambiguous, so that identification can be straightforward. Bernsen's approach to the output taxonomy defines modalities according to four attributes and three media, resulting in forty-eight potential modalities. The subsequent reduction to thirty modalities by the removal of some combinations of attributes raises questions about the utility of these attributes. The MSM Framework does not specifically define or classify modalities, leaving their derivation dependent on the particular user or system configuration, thus making generalisations over classes of modalities problematical. The MMDS framework deals with modality in a very different manner, at first defining them in terms of modes of operation, then in later work (Dowell et al, 1995) seeming to imply that they are similar to human senses. There is no stronger or more precise identification of modalities or classes of modalities, so generalisations or manipulations of modalities are not possible from this particular approach. The CARE properties provide a framework for

reasoning about aspects of multi-modal interaction, but do not specifically define or classify instances of modalities, or talk about their common characteristics or elements. In the TRS, representational systems are categorised according to three dimensions, resulting in a wide coverage with thirty-six different instances identifiable, although there are problems with the elements being confused and sometimes apparently unnecessary.

A further issue is that the total modality space defined needs to be large enough to be comprehensive, but small enough to be tractable. Bernsen and Purchase have derived a space of forty-eight and thirty-six respectively, which is probably as large as could reasonably be applied to the identification of modalities at the interface.

Modalities can be made up of smaller modalities, and therefore need to be reasoned about at varying levels of abstraction, according to the needs of the analyst. Bernsen defines modalities at three levels of abstraction: a super level which is referred to only in passing and is not fully defined; a generic level which aims to set out all possible modalities; and an atomic level which distinguishes between modalities which have identical descriptions at the generic level, but which may reasonably be considered as distinct modalities on a practical basis. The atomic level is as yet not fully defined, although Bernsen does give some examples of what is considered to be a modality at this level, for instance graphical images, or graphs. However it seems that these levels of abstraction are inconsistent, in that a modality can be at one level generic, and also atomic, whereas another modality may be atomic, but share characteristics of other modalities at the generic level. This suggests a flaw in the levels of abstraction, which may make it more difficult to reason about groups or instances of modality.

The MSM framework does deal with the issue of abstraction, relying heavily on information being abstracted into more or less complex forms as it is passed up and down the processing stages, by means of fusion or fission. In MMDS, levels of abstraction are not discussed, nor are issues relating to the identification of modalities at differing levels of granularity. In the CARE properties, modalities are examined at varying levels, and are considered to be simple, or atomic, if they cannot be broken down into further modalities at the particular level of analysis. The notion of compound modalities is used for modalities which are made up of more than one atomic modality. These are seen as being



complementary, in providing additional information, or redundant, providing identical information.

The TRS does not cover issues of abstraction since Purchase has explicitly concentrated on classification issues rather than taking a wider focus. However, the category of augmentation does have some level of granularity, in that a text may have more than one attribute associated with it. This suggests that modalities may be dependent on other modalities; for example, a text modality may have the additional attribute of colour.

### *3.3.6.2 User and System*

Interaction is concerned with both the user and the system, since it is the communication and interplay between them which is of most importance in HCI. As such, issues of multi-modality should be examined from both directions, rather than purely from the user to the system or the system to the user, since users and systems have different characteristics, and therefore differing constraints. Effective communication relies on modalities being compatible between the two parties. Not only must the modalities suit the interaction, but what is being expressed by one party must also be intelligible to the other party, or else communication will fail.

In Bernsen's taxonomies, modalities are explicitly defined (Bernsen, 1995b) in relation to system properties and capabilities rather than with regard to human needs and abilities. Modalities are regarded in terms of how the information forms are represented by a computer system rather than how they are perceived by the user. One of the driving reasons for producing the theory and taxonomies was to facilitate the building of design tools to support interface development, and Bernsen has therefore approached the question of modality from an engineering perspective, in terms of how computer interfaces can be constructed. This has inevitably coloured the approach, and resulted in taxonomies which reflect existing technology. If a more user-centred approach had been taken, the taxonomies may have been found to need different attributes, to be more flexible and responsive to new forms of interaction, to be non-device specific, and related directly to a user's capabilities.

The MSM framework incorporates aspects of ICS (Barnard and Teasdale, 1991) and defines modality not only in terms of input device and processing unit, but also in terms of

human channels and cognitive processing units. However, the discussion on fusion, fission, granularity and context places more emphasis on the system or on the properties of the interaction rather than the user.

The MMDS takes a more balanced view in that the framework demands a user profile as well as device constraints, thereby explicitly recognising the importance of the user abilities and system constraints in the interaction. In the design space both parties are examined to ensure that the interaction is positive. The inclusion of the cognitive properties based on the work by Wickens et al. (1983) reinforces the linking of system properties with cognitive constraints.

The CARE properties are divided into two sets of properties, System CARE and User CARE, in order to examine the modalities from the point of view of each party.

In the TRS, Purchase considers the flow of information from the computer to the user, with only two human sensory channels, and no constraints. This is not surprising since the taxonomy is explicitly concerned only with representational systems, rather than all aspects of interaction.

If systems and users have different constraints, then the modalities used in an interaction must overcome these constraints. Only the CARE properties explicitly address this issue, stressing that there must be compatibility between the user properties and the system properties for a system to meet user needs appropriately.

Related to this notion of compatibility is the issue of interpretability. If a modality is to be effective, then it must be recognised and processed. However, although a modality may be expressed by one party, there is no guarantee that it will be recognised as such by the other. Thus information will not be exchanged and the interaction can fail. This makes the concept of modality directional, and splits it into two parts: that which is expressed by one party, that that which is received by the other. This is an issue not addressed by Bernsen, Coutaz or Purchase, and only partially addressed by Dowell and the CARE properties.

### ***3.3.6.3 Combination***

Multi-modal usability implies that modalities can be used in combination. This leads on to other issues: how many modalities can be used at once; whether the combination of



modalities giving different information can be blended together; whether the information is identical but given in differing modalities in order to overcome potential interference or ambiguity McKensie Mills & Alty (1997). It must also be recognised that modalities from one party may interfere with or complement modalities expressed by the other.

Bernsen takes the view that only by understanding the properties of single, or “uni” modalities, can multi-modal systems be understood. It is reasonable that single modalities should be fully understood before combinations of modalities can be investigated; however, it seems that Bernsen believes that the properties of multi-modality result from combining single modalities, ignoring issues of combination.

In the MSM framework single and multiple modalities are examined in a more abstract manner, differentiating between modalities used separately and modalities used in combination. Single and multiple modalities are distinguished in terms of time: whether modalities are used exclusively or alternately; and number; used concurrently or in combination. Modalities are used exclusively when only one modality is used. They are used alternately when first one modality is used, and then another. Concurrent modalities are when two modalities are in use, and synergistic modalities are when two or more modalities are used in combination.

The dimension of cross-media integration in the MMDS deals with issues of load and combination, being directly based on those issues discussed in the MSM framework. However, the design space does not go into detail, and leaves these issues open.

CARE explicitly distinguishes between the different types of combination at a high level, by means of the four properties. The properties of complementarity and redundancy are especially relevant, since they deal with modalities used in combination. Complementary modalities are modalities which are conveying separate information which can be blended together to provide overall information. Redundant modalities convey exactly the same information, without adding anything extra. Such modalities can be used to reinforce each other or to ensure correct delivery of the information.

The Taxonomy of Representational Systems is concerned more with definition than with use, and therefore does not cover issues of combination

#### ***3.3.6.4 Time and Ordering***

When dealing with more than one modality, timing issues, and associated problems of ordering, become critical. Usability issues surrounding the duration of a modality, and when it starts in relation to other modalities, need to be established. The blending of modalities is also an issue because of the fusion of information and the potential for clashes between information being received from different modalities at the same time.

Bernsen mentions time only briefly, stating that modalities can be static or dynamic, with static modalities offering freedom of perception. Issues of ordering are not examined, nor are problems relating to the blending of inputs or outputs. This is due to Bernsen's concentration on properties of single modalities, which are unaffected by such issues.

The MSM framework examines the temporal aspects of modality from the perspective of blending inputs to the system. Three forms of fusion are identified: microfusion, or the fusion of data produced in parallel; macrofusion, or the fusion of data produced within certain temporal parameters; and contextual fusion, where data is combined according to structure. Different fusion strategies are examined, with the advantages and disadvantages of "eager" and "lazy" fusion assessed.

Neither the MMDS nor the CARE properties address issues of time or fusion. In the TRS Purchase refers to time only in relation to the duration of a text, and no mention is made of blending of representations.

### **3.4 Cognitive constraints**

The previous section has concentrated on approaches dealing with multi-modality as a whole, resulting in the identification of four inter-related themes: identification and granularity, user and system, combination, and time and ordering. Some of the approaches examined included an awareness of cognitive restrictions. When examining the interaction between a user and a system, it is essential to have an understanding of the mental and physical capabilities of the user in order to understand those aspects of modality which will affect performance. This section develops this by examining four cognitive models with



regards to the previously identified themes, in order to identify cognitive restrictions applicable to multi-modal usability.

The brain is not one homogenous entity, but is a system made up of several component processors or modules, each of which have different functions, such as perceiving stimuli, processing stimuli, and reacting to stimuli. There are several different theories as to how the cognitive system is composed, but they share common features: the brain is a system for processing symbols, and that these symbols are processed and manipulated, from one representation into another; the brain has certain processing limitations, in terms of structure and resources, and these cognitive processes take a certain amount of time. Given that the brain has resource and structural parameters in the processing of information, and that this processing takes a certain amount of time, it is important to know these restrictions in order for multi-modal interfaces to make the best use of those resources that are available. Otherwise, demands will be placed on the user that may conflict with or overwhelm the cognitive system, and the interaction will suffer as a result.

#### **3.4.1 Four Cognitive Models**

Several cognitive models have been proposed, which aim to determine response times and response compatibility, by combining empirical evidence into one structure. Various approaches to multi-modal usability have tried to incorporate a psychological element to their theories, for example Coutaz et al (1993) use aspects of ICS (Barnard and Teasdale, 1991), Dowell et al (1994) uses the SCR developed by Wickens et al. (1983). Four such models, based around the idea of discrete stage processing, were examined within the context of multi-modal interaction.

The four models examined were: the Human Information Processing Model (HIPM) (Wickens et al., 1983); the Model Human Processor (MHP) (Card et al, 1983); Interacting Cognitive Subsystems (ICS) (Barnard and Teasdale, 1991); and the Executive Process-Interactive Control (EPIC) (Kieras and Meyer, 1995). For a description of each model, please refer to Appendix A.

The four models all share certain common features, namely: sensory stores, transformation of information from one form into another, the flow of information around a system of some sort, and a way of moving or changing the body state. The four models also share the potential to shed light on aspects of multi-modality, in terms of general principles of operation. On closer inspection, however, it was discovered that the models were either too task specific, in the case of MHP and EPIC; too general, in the case of ICS; or incomplete, as in the case of the HIPM, to be immediately applied.

HIPM is a model which looks at the stimuli, the type of cognitive processing needed, and the best type of response for that stimuli. As such it is not concerned with other issues of memory or load, since it can look only at single stimuli. MHP is based on various times to process information by the various processors, and can give a wide range of values depending on the parameters of the particular task. Each task has to be examined separately, and there are few overall concepts that can be used. ICS is a highly dynamic and very complex configuration of separate subsystems, which interact when achieving complicated tasks. These interdependencies make it difficult to use the model without some kind of simulation facility. There is also the issue of precision, since ICS is concerned with “patterns” of cognitive activity (Barnard and May, 1993), which again makes it difficult to apply.

However, these models did make various issues of importance to multi-modal usability more apparent, and they are examined below in more detail according to the themes of properties of senses, representation, load and combination, and time, in order to identify those cognitive issues of importance to multi-modal usability.

### **3.4.2 Properties of Senses**

The four models emphasised the differences in characteristics between the human senses, and how stimuli are perceived and attended to differently depending on the sense used. All four models regard the perception of information in similar ways, with information perceived by a particular sense, and stored in the sensory buffer connected to it. This information is then processed further, in a form that can be recognised and acted upon by the cognitive processor. Attention can be defined as “selectivity of processing”, and has



various forms: voluntary, when the concentration on a particular task is chosen; involuntary, when a reaction is caused by external stimuli; focused, when attention is given to one stimulus out of many; and divided, when attention is split between two or more stimuli at once.

It is essential to understand the factors involved in perception and attention in order to assess the implications for multi-modal usability. Research has tended to concentrate on the perceptual and attentional aspects of the senses of hearing and vision since those two senses are of most importance to HCI. Therefore an understanding of their properties is necessary to any understanding of multi-modal usability. A brief overview of some of the more important aspects of these two senses based on Eberts (1994) is presented in table 3-1.

VISION	HEARING
spatial nature	temporal nature
can be both sequential and simultaneous in presenting information	typically only sequential in presenting information
poor attention-demanding qualities: user has to be looking at display	attention-demanding: grabs the user's attention
can be kept before user: good freedom of reference	cannot be kept before user: not easy to refer to, although can be repeated
more dimensions for coding of information	fewer dimensions for coding of information
speed can be faster but more searching can be required	speed can be slower but less searching is required
less resistant to fatigue	more resistant to fatigue

**Table 3-1: Properties of Senses**

### 3.4.3 Representation Issues

Stimuli are not only are perceived by a particular sense, but also have a particular form. There are two main types of representations: analogical, in that their structure resembles that of the world, for instance pictures and diagrams; and word based representations, which do not have such a direct relationship, and are propositional. This seems to be reflected

internally in the cognitive system of the user, with Paivio's Dual-Coding theory (see the description from Eysenck & Keane, 1990) proposing two distinct symbolic systems, one based on linguistics, and the other non-verbal and based on images.

Modalities can make use of these properties of representation, since certain kinds of information are better suited to one form of representation than others. For instance, quantitative information is usually best represented visually, especially numerical values, whereas instructions are best represented in a lexical format. The representation is also easier to process if the stimulus resembles what it is depicting, for example using a symbol of a aeroplane to represent an airport on a map. However, if there is a difference between the symbol and what it represents, this can cause interference. This has been demonstrated by the Stroop effect (Preece et al, 1994), where it has been found that it is very difficult for someone to say a colour name if it is presented in a different colour, for example the word blue written in a green coloured font.

The HIPM explicitly makes use of these coding properties in order to minimise processing times. The ICS architecture also acknowledges these coding properties, with the range of different subsystems. Information is transformed through the various subsystems, and is processed differently depending on its form. The Morphonolexical subsystem deals with linguistic representations and the Object subsystem deals with visuo-spatial form. The Articulatory subsystem deals with spoken output, and the Limb subsystem deals with movement. Since there is only one subsystem dealing with a particular representation type, it follows that the subsystem can only deal with one particular instance of representation at a time, unless they can be blended. This explains why, for example, two spoken conversations cannot be attended to at the same time.

#### **3.4.4 Combination Issues**

Stimuli are rarely presented individually, but can vary in their load or combination, which raises issues of divided attention and interference. Divided attention is where attention is focused on two or more tasks at the same time. This can be problematic; for example, it has been established that if two messages are played at the same time, both using the same voice, it is very difficult for listeners to separate them out, because normally physical



differences between stimuli are used to distinguish streams of speech. If attention is focused on one message, the other message is not listened to. This is due to processing limitations of the sensory buffer. Although both stimuli arrive at the buffer at the same time, only one is filtered through. The other input remains in the buffer. There is therefore a bottleneck at the initial processing stage, which is why both inputs cannot be fully processed and thus understood. Thus messages which are similar and simultaneous cause interference and loss of processing, in that only one can be attended to. However, presenting identical information both visually and aurally at the same time can increase the likelihood of information being received.

When dealing with multiple visual outputs more time is needed to make judgements simply because of the time taken to look at them all. It seems that people tend to concentrate on one output, and possibly ignore the others. Thus multiple visual outputs will often only be attended to if attention is directed to them. One way of doing this is by cueing through the auditory channel, to direct attention visually. Not only does this reduce search time, it also reduces interference.

The HIPM states that tasks can cause interference if they share the same processing paths, or make use of the same memory codes. Thus for example two spoken conversations cannot be listened to at a time because they are demanding the same resources.

In the MHP, the perceptual and motor systems are configured so as to act in parallel, so that more than one thing can be processed at a time. However, the cognitive system is parallel only in the recognise phase, so that responses are serial in nature. This means that multi-tasking can only occur through skilled allocation of control actions. Cognitive resources are given to many things in turn very quickly.

With ICS, all processes can happen at once, but only one data stream can be handled by a subsystem at any one time. Integration occurs when information from different sources, for example speech and facial expression, arrives at the same subsystem, for example the Implicational subsystem. These can only be blended when they can be related. If the voice became out of synchrony with the face, as might occur with the soundtrack to the film, the two cannot be blended into one experience because the data flows are not timed so that they

can be integrated (Barnard, 1996). However, there is only one subsystem dealing with the thought processes governing movement. Unlike MHP and EPIC, there are no explicit constraints to the effectors detailed with the model. Over time, similar information flows are collected into abstracted records, which allow the development of expertise, by enabling one subsystem to produce an output representation for another subsystem directly, without the need for intermediary transformations. In this way experts have different cognitive requirements to novices.

In EPIC, the cognitive processor is configured to act in cycles of activity, and is able to set off numerous production rules at a time, limited only by the structural constraints of the sense organs and effectors. These however have a profound effect on the interaction, constraining the physical manifestation of the response. There are also many different factors which come into play when responding. For example, when making an aimed movement, the visual system is needed. This may however interfere with collecting the visual information needed for the next part of the task (Eberts, 1994). With the growth of expertise in a user, future movements can be anticipated and therefore stored ready to be generated, thus saving cognitive processor time and reducing overall time in execution. However, since EPIC is based on very low-level stimuli processing and response actions, it has to be configured for each individual doing a specific task. This makes it more difficult to derive generalisations of peoples' behaviour, expert or novice, over larger groupings of stimuli combinations, if, indeed, this is at all possible. **Temporal Issues**

Given that processing of information through the various cognitive stages takes time, it is an important property to examine. The four models have looked at this in differing ways. The HIPM does not specifically look at temporal issues overall, being mainly interested in reducing processing times by allocating the best response to the stimuli received. In the MHP time is seen as extremely important, in that it determines the response times, and also has bearings on the cycles of cognition in the cognitive processor. With ICS timing is regarded as important, in that there are timing considerations for blending of information, but there is no other work on the properties of the timing that need considering, for example the dynamic or repeated nature of a stimuli. The EPIC model is similar to the MHP in how



it regards time, with a cyclical central processor which deals with batches of stimuli and responses.

It is difficult to talk about the temporal properties of stimuli without being very precise about the particular stimuli. However, overall statements can be made about the temporal nature, and the difference this can cause the cognitive loading. Stimuli can take two forms: brief or ongoing. If brief, they will need only one cycle of processing. If ongoing, they can take two further forms. They can either be unchanging, or changing. If they are unchanging, they will need less processing after the initial processing, because there is nothing new to process. If they are changing, they will require cognitive processing to continue, in order to deal with the changing nature of the stimuli. There are therefore three different temporal forms that stimuli can take.

It was earlier established that there are limitations as to what can be processed at any one time. Therefore, if there are multiple stimuli, requiring processing, there will be interference.

#### **3.4.6 Derived Cognitive Issues**

This section has investigated cognitive issues affecting multi-modal usability by means of examining four cognitive models. Various issues are apparent from the above work, and should therefore be addressed by any theory of multi-modal usability. The issues are presented in summary below.

The senses are processed individually using separate processors, and only one thing can be properly processed at any one time. The visual sense has spatial constraints, in that in order to see something, the user must be looking at it. The auditory sense has temporal constraints, in that it takes time to hear something.

Stimuli may be represented spatially or lexically, and can be processed differently accordingly, with only one type being processed at a time. At a high level, the cognitive system will seek to resolve differences between stimuli where stimuli are reasonably expected to be similar.

There are effector bottlenecks which can affect response. Expertise can affect the response, since responses can be made more readily.

Stimuli can be brief or ongoing, and require different cognitive resources depending on the temporal form they take.

### **3.5 Summary**

This chapter has argued that there is a need for a clear definition of modality. To determine what form this definition should take, existing definitions based on sense, device, mode and information form, as well as definitions composed of multiple components, have been critically assessed. The examination established that current definitions are unsatisfactory for several reasons, and that a new definition should be based around the human sensory channels combined with other elements, possibly including information form.

This chapter continued by investigating usability properties of multi-modal systems by critically assessing five different approaches to the usability of multi-modal and multi-media systems, namely: Modality Theory (Bemsen, 1995a), Multi-Sensori-Motor (MSM) framework (Coutaz et al, 1993), Multimodal Multimedia Design Space (MMDS) (Dowell et al, 1994), CARE properties (Coutaz et al, 1995), Taxonomy of Representational Systems (TRS) (Purchase, 1999).

These were then used to derive critical properties of relevance to multi-modal usability. Four main areas of concern were identified and examined in detail: identification and granularity; user and system; combination; and time and ordering.

The importance of understanding the mental and physical capabilities of the user was considered next. Four cognitive models: the Human Information Processing Model (HIPM) (Wickens et al., 1983); the Model Human Processor (MHP) (Card et al, 1983); Interacting Cognitive Subsystems (ICS) (Barnard and Teasdale, 1991); and the Executive Process-Interactive Control (EPIC) (Kieras and Meyer, 1995) were examined with regards to the previously identified multi-modal properties in order to identify cognitive restrictions applicable to multi-modal usability.



Several important issues were identified by this study concerning: sense processing constraints; internal representation and processing of stimuli; effector bottlenecks; role of expertise; and cognitive resources for differing temporal stimuli.

The next chapter draws on the issues identified in this chapter to present a new definition of modality, a taxonomy for identifying instances of modality, the properties of modalities, and the cognitive restrictions on modalities.

# **4. MODALITY DEFINITION AND THEORY**

## **4.1 Introduction**

The previous chapter discussed the problems with existing definitions and theories of modality, and examined various issues relating to multi-modal usability. This chapter draws on this work to present a new definition of modality, an associated taxonomy, the properties of modalities, and the cognitive restrictions on modalities.

A new combined definition of modality, based on the elements of sense, information form and time, is proposed to overcome the problems identified with definitions in the previous chapter. An associated taxonomy is presented, comprising of three dimensions, each with three categories, for identifying particular modality instances.

From the four issues of: identification and granularity; user and system; combination; and time and ordering; detailed descriptions of modality properties are developed. These include user and system modalities, expressive and receptive modalities, and composite and atomic modalities. The cognitive restrictions identified previously are developed into five clashes which affect multi-modal usability.

## **4.2 A new definition of modality**

It was established in the previous chapter that defining modality in a way that has relevance for human-computer interaction is not a trivial task, and that there are several issues that any definition should address. Overall the definition must not be too general or too specific, otherwise it will suffer from being vague or context-dependent, neither of which will support an understanding of what is happening at the interface level. The definition should allow for comparisons and contrasts between different modalities, so that issues relating to groupings of modalities can be identified and addressed. Since the purpose of this definition is to define modality in a way that is relevant to the understanding of multi-modal interaction, and thus interface design, it should incorporate elements that are considered to be essential in this field.

In order that these differing requirements can be incorporated into one definition, a definition of modality based on a combination of elements is proposed. Three



components of this definition are examined below, followed by the presentation of the new definition based on these elements.

#### **4.2.1 Human Sense**

The human senses have direct relevance to multi-modal interaction and usability. Whereas computer technology can change, the human parameters are fixed, and although one or more senses may not be available to certain extraordinary users, on a practical basis it is useful to be able to distinguish between visual, auditory and haptic sensory experiences (ignoring smell and taste) when discussing what is occurring in a multi-modal interaction. There is an established cognitive base for this approach. The different senses rely on different cognitive processors, and have differing properties, as exemplified by the different models of cognition examined in the previous chapter.

This follows the approaches taken by Bernsen (1995b), Coutaz et al (1993), Dowell et al (1994) and Purchase (1999), who all explicitly relate their work to the human senses. Bernsen defines modality in terms of three potential media, and Purchase and Dowell with reference to vision and hearing. Coutaz leaves the sensory components relatively undefined, but talks about sensory channels of communication.

By defining modality in terms of a particular sense, the definition becomes user-dependent to some extent, since the computer modalities can be defined in any manner. It is the user for whom the interaction should be usable, so the user should be at the heart of any definition of modality. There are advantages to this, because the form of the user is relatively static in comparison to the form of the system, whose input and output possibilities are constantly changing and developing as new technologies are implemented. The use of senses within a definition of modality also allows for previous work on multi-modal usability to be more readily integrated.

#### **4.2.2 Information Form**

A definition of modality in terms of the sense used is not enough by itself, as this is still too general and does not allow leverage with regards to multi-modal usability, since it does not permit the identification of fine distinctions between various stimuli. The form that the information takes is important, since humans represent information cognitively in two ways, according to the dual-coding theory proposed by Paivio (Eberts, 1994). The ICS model (Barnard and Teasdale, 1991) has two subsystems for processing the

two types of stimuli: lexical and visuo-spatial, and two subsystems controlling response of an articulatory and movement form. Wickens has exploited the compatibility between stimulus and response in the Human Information Processing Model (Eberts, 1994), differentiating between verbal and spatial stimuli and responses. Dowell has included aspects of this in the Multimedia Multimodal Design Space (Dowell et al, 1994).

Three of the four ways that Bernsen distinguishes modalities in (Bernsen, 1995b) relate to the form of the information. The taxonomy differentiates between linguistic and non-linguistic, analogue and non-analogue, and arbitrary and non-arbitrary. Although these categories are somewhat confused, it is an attempt to define information form in terms of parameters of words, resemblance and familiarity of meaning. Purchase (1999) has examined information form from the perspective of established semiotic theory and a whole dimension of the taxonomy is concerned with information form, emphasising its importance in distinguishing between instances of text.

Therefore, being able to identify the information form of a modality has important consequences for multi-modal usability in terms of processing and response, and should be incorporated into a new definition.

#### **4.2.3 Temporal Form**

The issue of time is one that is crucial to multi-modal usability, because of the cognitive resources needed to process modalities; however, it is an element that has been rarely examined in other definitions of modality. Only Bernsen (1995b) explicitly refers to it, defining modality in terms of being static or dynamic. Within a wider context, Coutaz et al (1993) in the MSM framework discuss the temporal aspects of fusion.

An understanding of the temporal aspects of a modality allows for information about its potential duration, in the abstract sense, and the cognitive resources needed to process it, to be apparent. It is therefore included as a novel element within the new definition of modality proposed below.



#### 4.2.4 New Definition

Based on the three concepts examined above, a new definition of modality is proposed:

**A modality is a temporally based instance of information perceived by a particular sensory channel**

Modalities are here defined as being compound entities composed of the three elements of sensory channel, information form and temporal nature, and any potential modality can be distinguished by means of these elements. These elements have direct relevance and importance to multi-modal usability. By placing them together in one definition they allow modalities to be defined in a more powerful and practical manner. As such it encompasses aspects of previous definitions, where modalities were identified in terms of sense and information form.

By defining modality according to a combination of elements, the issues of over-generalisation and over-specificity can be addressed. The definition is precise enough to make fine-grained distinctions between modalities, yet general enough to allow groupings of modalities to be considered and evaluated. Combining elements is not a new way of addressing the definition of modality, and previous definitions have identified modalities in this manner. Of the more precise definitions, Bernsen (1995b) defines modality according to media, and four pairs of attributes, and Purchase defines texts according to three dimensions of signs, syntax and sense. Other, less precise, definitions have represented modalities as combinations of device and language (Nigay et al, 1993), (Duke & Harrison, 1995), or as combinations of sensory channel, device and language (Coutaz et al, 1995).

However, it is the inclusion of temporal nature as an element which makes this definition unique, allowing as it does an awareness of the attentional resources needed to process stimuli and the possible conflicts that may cause. This definition is also novel in that it allows for consideration of both input and output modalities since it is not explicitly tied to system or human capabilities but rather to shared issues, thus giving it a potentially wider usage across all aspects of multi-modal interaction instead of being limited to a particular side of the interaction.

This novel definition gives various advantages over previous definitions. By not including the idea of mode, but instead concentrating on the sensory channels, it is

possible to be more precise about the role of the definition and modalities. The addition of other elements to that of sense allows the definition to distinguish between different types of information that utilise the same sensory channel. For example the representation of written text and pictures both use the visual sensory channel, but are clearly different in composition. A similar example would be the difference between music or speech, or typing as opposed to mouse movement. By including information form as an element with other elements it is possible to make clear distinctions between forms using different senses, such as between written text and speech. The device is not included within this definition, which makes the definition less dependent upon a particular technology. A finer grain of distinction can be made between different information forms which are represented by the same device, for example the use of monitors to show written text and video. Although previous definitions have used combination of elements to provide greater preciseness and flexibility, they have not included the element of temporal duration, making it difficult to distinguish clearly between single instances and ongoing instances of modalities.

### **4.3 Modality Taxonomy**

A definition of what comprises a modality on its own is not enough. Previous definitions of modality, for example (Coutaz et al, 1993), have left the identification of particular modalities undefined, which hampers the identification of modalities by making the clear differentiation between particular instances problematic. A classification, or taxonomy, of the properties is necessary to aid the identification of individual modalities. By classifying modalities they can be described, and their complex nature reduced down to those elements considered essential to multi-modal usability. As (Bailey, 1994) observes, classification helps to achieve the research goals of explanation and prediction, and is a prerequisite for theorising, since “Theory cannot explain much if it is based on an inadequate system of classification.” Ideas can be managed and controlled in this manner (Marcella and Newton, 1994), for as (Gould, 1983) states: “...classifications both reflect and direct our thinking. The way we order represents the way we think.”

By proposing a taxonomy a structure is provided for modalities to be identified according to their elements. Without this it would not be possible for instances of modality to be compared and thus identified appropriately (Pankhurst, 1978). As it is



not straightforward to establish which distinct categories are essential to the three different dimensions of modality as defined above, the temptation is to make any classification as large and complex as possible in order to accommodate the many possibilities. This can result in a taxonomy that is difficult to apply and which overwhelms the user with the choice of categories. Therefore, since only a selection of categories may be necessary in order to make meaningful distinctions between modalities, the approach that this classification takes is to err on the side of brevity, on the basis that further refinements can be made if found to be necessary. The categories within these dimensions have been chosen to minimise within-group variance, whilst maximising between-group variance. This within-group homogeneity and between-group heterogeneity has resulted in categories which are as distinct as possible, in order to avoid overlapping (Bailey, 1994).

This taxonomy consists of three dimensions corresponding to the three elements of sense, information form and temporal nature, as embodied in the compound definition of modality proposed earlier. The three dimensions are each examined in turn, and their component categories identified and explained.

#### 4.3.1 Sensory Dimension

The first dimension of this taxonomy deals with the sensory element of the modality definition. Of the five senses the three senses of sight, hearing and touch are the main channels through which humans interact with computers. It is therefore sensible to have them as the components of this dimension. This dimension will obviously need expanding should any interface be developed that incorporates smell and taste; however, until that eventuality, the three categories are identified as visual, acoustic and haptic, as shown in the table below.

Visual	text on a computer screen
Acoustic	human speech
Haptic	typing

**Table 4-1: Examples of sensory modalities**

This follows previous work by Bernsen (1995b) and Purchase (1999). Bernsen defined modalities in terms of the three media of graphic, acoustic and haptic, and Purchase defined texts in terms of the visual and aural modalities. This category takes a novel approach in comparison to this previous work which concentrated only on output or representation, with this dimension designed to be equally applicable to humans and systems. This is of prime importance given the work on virtual environments (Mills & Noyes, 1999) and the development of robotic computers (Anzai, 1994), which blur the distinction between users and systems into a community of users.

#### **4.3.2 Information Form Dimension**

The dimension of information form concerns itself with the various ways in which information can be presented. Humans represent information internally in two distinct ways: word-based and image-based. The IHPM (Wickens et al., 1983) makes use of this in determining which response is best for a particular stimulus, and ICS (Barnard and Teasdale, 1991) assigns two subsystems for dealing with the two information forms.

Bernsen (1995b) looks at the information form in a slightly different manner, separating out the various possibilities. Modalities are described as linguistic or non-linguistic, in that they can have lexical properties. They can also be arbitrary or non-arbitrary, depending on the assigned value. For instance, a word might be lexical, but it may or may not have meaning, depending on the circumstances. The modalities are also described as analogue or non-analogue, depending on whether there is a direct relationship between what is being represented and its representation. This is slightly different to the idea of information form being either word or image based, since according to Bernsen's taxonomy it is possible to have linguistic analogue modalities, or a word representation which has a direct relationship to what is being represented. This is where the taxonomy becomes confused, and certain combinations of properties are removed because they are not feasible.

Purchase (1999) takes a different approach, basing the dimension of signs on established semiotic theory. There are three categories: concrete-iconic, where there is a perceptual relationship between the representation and the concept, abstract-iconic where the perceptual relationship is stylised, and symbolic where there is no perceptual relationship. This allows for a greater discrimination than that afforded by a distinction



between words and images, and allows for diagrams to be classified and represented more easily.

Lexical	speech or text
Concrete	photo
Symbolic	use of colour, a fire alarm

**Table 4-2: Examples of information form modalities**

Based on the work by Purchase, the dimension of information form is here given three categories, which have been re-named in order that their meaning might be more apparent: lexical, where the information is word-based; concrete, where the information is in the form of reproduction of a real life object; and symbolic, where the information is a representation of something rather than an actual reproduction of it. The table above gives examples in each category.

Separating out these three categories of information form is problematic. The distinction between lexical and other forms may seem straightforward, until one starts to consider exactly what the difference between concrete and symbolic, or symbolic and lexical, might be. One should think rather of these categories as divisions on a continuum, with at one end lexical representations, in the middle area symbolic representations, and at the other end concrete representations (Sasse, 1997). This makes it harder to be decisive regarding which category a modality fits into, since arguments might be made for something to be concrete to one person, and symbolic to another, or symbolic to one person and lexical to another. For example, a Chinese character to someone who reads Chinese would be lexical, but to someone who does not, would be symbolic. Deciding on which information form category to use is therefore dependent upon the individual circumstances, which may make it harder to apply this taxonomy to interactions. This will be examined in more detail in the case study reported in Chapter Six.

### **4.3.3 Temporal Dimension**

The temporal nature of the modality bears a direct relationship to the kind of cognitive processing needed to absorb it. In the examination of the cognitive models in the previous chapter it was established that the perception and processing of stimuli takes

time and that the temporal nature of the stimuli can be identified as falling into three different categories. This dimension is therefore given the three categories of discrete, continuous, and dynamic. A discrete occurrence of a modality is one that is unchanging within its occurrence, which is brief, and thus needs only preliminary processing, but however is transient and may therefore be missed or not attended to. A continuous occurrence of a modality is one that is repeated exactly the same more than once. It is distinguished from the discrete occurrence, because it allows for some degree of inspection, and has an element of repetition. A dynamic modality is one that changes in content within its occurrence, which may last for some time, and thus requires continuous processing. Examples are given in the table below.

Discrete	press of a button
Continuous	continuous display of an icon on the screen
Dynamic	video, music

**Table 4-3: Examples of temporal modalities**

These three categories are meant to be overall divisions of the element, and it is perfectly feasible that what might be considered to be dynamic at one level of experience might be considered to be discrete under others. For example, consider the movement necessary to change gear. Depending on the level of analysis, it might be better to regard this as a dynamic movement, or to take a high-level approach and consider it as one discrete movement. It could even, depending on the particular purpose of the analysis, be broken down into a set of discrete, individual movements.

#### **4.3.4 Analysis of Taxonomy**

This taxonomy is a taxonomy rather than a typology since, although it deals with concepts, which would more usually be described by a typology, it describes modalities which are empirical entities (Bailey, 1994). Its goal is to classify instances of modalities according to their similarity to observed variables. It is a faceted classification system (Marcella and Newton, 1994), taking its cue from Colon Classification (CC) developed by Ranganathan (1962). An analytico-synthetic approach has been taken (Marcella and Newton, 1994), in that the dimensions have first been analysed to determine their



individual components then these components synthesised to form a modality. Such a fully faceted classification does not need to list each modality individually, but rather provides a kit from which any modality can be specified, even if such a modality might never have been predicted at the time of compilation.

Each cell represents one taxon, indicating one particular type of modality. The modality is named after the intersection of these components, building the component names into a compound modality name following the analytico-synthetic approach described above. Examples of such modalities at the operational level (as defined by Bailey, 1994) are given in tables 4-4, 4-5 and 4-6.

VISUAL MODALITIES	LEXICAL	SYMBOLIC	CONCRETE
DISCRETE	A one word message appearing then disappearing on a screen	An icon appearing then disappearing on a screen	A photo appearing then disappearing on a screen
CONTINUOUS	A one word message appearing and remaining on a screen	An icon appearing and remaining on a screen	A photo appearing and remaining on a screen
DYNAMIC	Scrolling text on a screen	A cartoon	A film

**Table 4-4: Visual modalities**

ACOUSTIC MODALITIES	LEXICAL	SYMBOLIC	CONCRETE
DISCRETE	A spoken word	A bell rung once	A door slamming
CONTINUOUS	A spoken word repeated continuously	A telephone ringing	An engine running continuously
DYNAMIC	A conversation	A detector tracking something	A dog barking

**Table 4-5: Acoustic modalities**

This taxonomy has various advantages. The most obvious is that it is able to cover the wide range of information types used in interaction. Additionally, it has been done using simple and straightforward criteria, allowing for easy identification of the modalities, and making it less cumbersome to apply than classifications which use more complex elements. This is examined in more detail in chapter six, which details a case study of students using the taxonomy in the usability assessment of two multi-modal interfaces.

HAPTIC MODALITIES	LEXICAL	SYMBOLIC	CONCRETE
DISCRETE	A word typed at a keyboard	A brief movement of a mouse	The vibrations from a printer starting up
CONTINUOUS	A word repeatedly typed at a keyboard	Continuous movement of a mouse	The vibrations from a printer printing continuously
DYNAMIC	A large amount of text typed at a keyboard	Use of a mouse to do several different things	The vibrations from a machine travelling over different terrain

**Table 4-6: Haptic modalities**

Unlike other classifications of modality, this taxonomy can deal with issues relating to colour and font type, since they can be regarded as modalities in their own right, rather than as attributes of other modalities. Both are symbolic if meant to represent something; for example, using the colour red on a warning sign means that as well as the warning, an additional modality of visual-continuous-symbolic is being used, that of the colour. This is examined in more detail later on in this chapter in the section on clashes.

Time has not been explicitly defined in this context, and should be regarded as being changeable depending on the circumstances. Thus, one person's discrete modality might be another person's continuous modality, because people can process faster or slower depending on ability or circumstance. This is related to issues of expertise and is discussed in more detail in the section on clashes later in this chapter.

This shows the flexibility of the classification, in that it allows for modalities to be defined in relation to a particular user, allowing for usability problems specific to that user to be examined. The way of defining modality allows for user profiling, in that more experienced users will process modalities differently and regard them differently to novice users. It also provides a starting point to consider multi-modal usability, in that modalities can be explicitly defined and their properties considered in combination.

#### **4.3.5 Example Of Taxonomy Use**

With the definition of modality, and the associated taxonomy, any modality can now be identified, in terms of its sensory, information, and temporal, form. In order to demonstrate this identification a very simple interaction with a telephone, taking the form: the telephone ringing, the user picking up the receiver, and the caller speaking; is examined.



In this example, the interaction with the telephone can be considered in terms of three modalities: the bell ringing, the user lifting the receiver, and the caller speaking.

The bell ringing uses the acoustic sensory channel, is symbolic in its information form, in that the bell sound means “telephone”, and is continuous, in that it will keep making the same noise until someone answers the call. The modality can therefore be identified according to the taxonomy above as acoustic-symbolic-continuous (abbreviated to aco-sym-cont for convenience of writing).

The user lifting the receiver uses the haptic sensory channel, is symbolic in its information form, in that the user is lifting something and thereby signalling acceptance of the call, and is discrete, in that it is one action that is not prolonged or repeated. The modality can therefore be identified as haptic-symbolic-discrete (abbreviated to hap-sym-dis).

The caller speaking uses the acoustic sensory channel, is lexical in its information form, because the caller is using words, and is dynamic in nature, in that the words are part of a conversation and thus change and are not repeated. The modality can therefore be identified as acoustic-lexical-dynamic (abbreviated to aco-lex-dyn).

By breaking the interaction down into these component modalities, it is now possible to reason about the differing requirements, and to talk about their implications in depth. Examining the first modality, several issues become apparent. In order for a bell to be heard, it must be loud enough to be heard. This may sound obvious, but identifying the modality as acoustic in form makes this physical requirement explicit. It must also be meaningful. A bell ringing on its own does not necessarily imply that someone is on the end of a telephone and wishes to have a conversation. It is the symbolic association of the sound of the bell with the understanding that a call is waiting to be attended to that creates that implication. A person who does not understand that the bell signifies a telephone call would be unable to use the telephone easily and correctly. Lastly, a single ring might not be appropriate, since other auditory modalities may be in conflict and cause that single instance to be ignored. A continuous ringing sound would therefore give the modality a temporal duration and presence, enabling it to be attended to.

Therefore the identification of the component modalities in the interaction has allowed several critical issues to be examined and discussed. The best kind of sense to convey

the information, the form that the information takes, and the time that the information is present, are all laid open to debate, and the best alternatives can be explored.

This reasoning can be applied to the other modalities in the interaction. For example, if to answer the telephone the receiver must be picked up, then the user must be capable of such a haptic movement. For some users, that is not the case. A careful examination of the modalities in an interaction can allow constraints such as these to become apparent.

## **4.4 Modality Types and Properties**

The definition of modality and the related taxonomy do not by themselves address all the issues relating to multi-modal usability. There are other modality properties which add to the understanding of multi-modal interaction. In this section, the four areas of: identification and granularity; user and system; combination; and time and ordering; are translated into specific properties relating to modalities, and fully described.

### **4.4.1 User and System Modality Types**

In chapter three the issues of user and system were detailed. When examining interaction, it is helpful to be able to distinguish between modalities originating from the user and modalities originating from the system. Both have different processing requirements. For a system, it will be dependent on the configuration and power. User abilities will be determined by their cognitive resources, the knowledge they possess, and their level of expertise. By examining both sides of the interaction, a better balance can be achieved, as in the case of the CARE properties (Coutaz et al, 1995) which explicitly examine both user and system requirements.

Therefore, modalities should be explicitly typed according to whether they are user modalities or system modalities.

Figure 4-1 represents the user and the system, interacting within the context of the world. To the user, the system is the telephone, as well as the person on the other end of the telephone, since that person's speech is being mediated by the system. This whole interaction takes place within the context of the world.

Going back to the example interaction with the telephone, it can now be written as:

The bell ringing (system)

The user lifting the receiver (user)



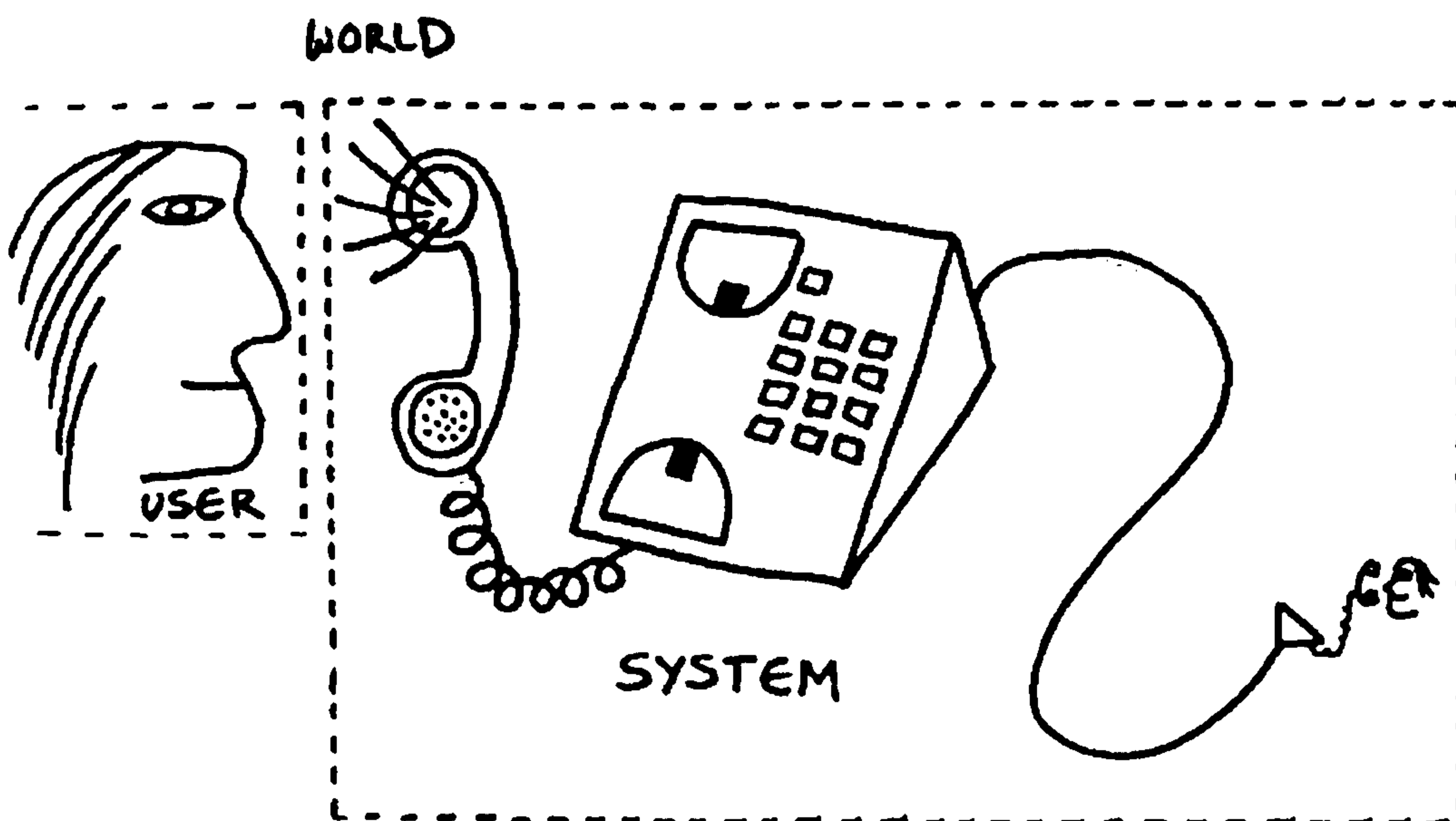


Figure 4-1: User and System interacting in context of World

#### 4.4.2 Expressive and Receptive Modality Types

Following on from separating out the system and the user modalities, it is also important to distinguish between modalities being expressed and modalities being received. For example, the system might move a lever, thus expressing a haptic modality, but the user would perceive this visually, and thus receive a visual modality. This allows modalities to be examined not only in terms of where they come from but also in terms of how they are received. This is a new way of looking at modalities which previous theories and definitions have not used.

Since computers do not sense stimuli in the same way as humans, referring to system receptive modalities as haptic, visual or acoustic is essentially meaningless. However, it does allow for consistency with the remainder of the interaction, so is still worth inclusion. At a later date the appropriate means of processing the information can be established, but at least certain properties of the information have already been established, in terms of its reception through optical channels, audio channels, or keypads and other motion sensors.

There is not always a direct pairing between expressive and receptive modalities, since what is expressed is not always what is received, and unintended information might come from modalities expressed. For example, shouting at a computer means that the user is expressing an acoustic-lexical-dynamic modality, but the computer will not receive this unless there is a microphone input device available. The modality is not received. An awareness of such pairings means that modalities can be tracked to determine whether they are received correctly. This is similar to the CARE (Coutaz et al, 1995) notion of compatibility between system and user properties.

Using this idea, the modalities used in the interaction with the telephone increases:

The bell ringing (expressed by the system)

The bell ringing (received by the user)

The user lifting the receiver (expressed by the user)

The user lifting the receiver (received by the system)

The caller speaking (expressed by the system)

The caller speaking (received by the user)

#### 4.4.3 Composite and Atomic Modalities

When examining interaction, it is helpful to be able to look at it at differing levels of detail. It was earlier established that modalities can be reasoned about at varying levels of abstraction, and Bensen (1995b), CARE (Coutaz et al, 1995) and MSM (Coutaz et al, 1993) regard modalities in this way, distinguishing between high-level modalities which can be made up of low-level component modalities. Here, building on the work done by the CARE properties, the overall modality is termed the **composite** modality, and the modalities at the lowest level of analysis are termed **atomic** modalities. An atomic modality may in turn be made up of other modalities if the analysis was taken down to a lower level of analysis. This atomic modality would become a composite modality, and those modalities of which it was comprised would be known as atomic modalities.

For example, imagine a telephone conversation where a person is talking. That modality would be a user expressive acoustic-lexical-dynamic modality. However, when a person is talking they can often add more information to what they are saying by varying their



tone, the volume, and even the accent in which they are speaking. These are all extra modalities, because they are different layers of information on the original modality of speech. By taking it down to a lower level of granularity, these additional atomic modalities can be identified. Obviously this depends on what level of analysis is appropriate or relevant: normally, when analysing a telephone conversation, the acoustic-lexical-dynamic modality would be sufficient.

The following example clarifies this. Imagine a Scotsman talking on the telephone. He is irritated, so he has raised his voice. He is in a hurry, so he is talking fast. The modalities would be as follows:

user expressive acoustic-lexical-dynamic: the Scotsman talking.

This would be the composite modality. This can then be broken down into atomic modalities at a lower level:

user expressive acoustic-lexical-dynamic: the words the Scotsman is using

user expressive acoustic-symbolic-dynamic: the volume of his words

user expressive acoustic-symbolic-dynamic: the speed of his words

user expressive acoustic-concrete-dynamic: his accent

The volume and speed are symbolic, because they represent something in the interaction: they show that the user is irritated and in a hurry. The accent is not important in this case, so it is defined as a concrete modality.

#### **4.4.4 Dependent Modalities**

Some modalities can only exist if there is a modality originally, and are therefore termed **dependent** modalities. For example, colour and size are dependent upon having an original modality. This is similar to the Purchase category of augmentation (Purchase, 1999) where texts can have additional attributes. However, here the attributes are considered as modalities in their own right. In the example above, the volume, speed and accent modalities are considered to be dependent, because they exist as derived properties of the acoustic-lexical-dynamic modality of the conversation.



#### 4.4.5 Mismatch

There is a difference between the expression of a modality and the reception of a modality, and simply because a modality is expressed does not guarantee that it is received, and additional modalities can be received on occasion in error, when none were intended. This property is termed **mismatch**, and can occur both when the expressed modality is incorrectly received, or if a modality is received that was not intended. This is similar, but not identical to, mode errors as described in Preece et al (1994, quoting Norman, 1981). It differs in that in a mode error the user believes themselves to be in one state when they are actually in another. With mismatch, the user or the system is interpreting information in one way when the information actually means something else.

An example of a potential mismatch of information concerns the Macintosh layout of menu options at the top of the screen display as shown in figure 4-2 below:

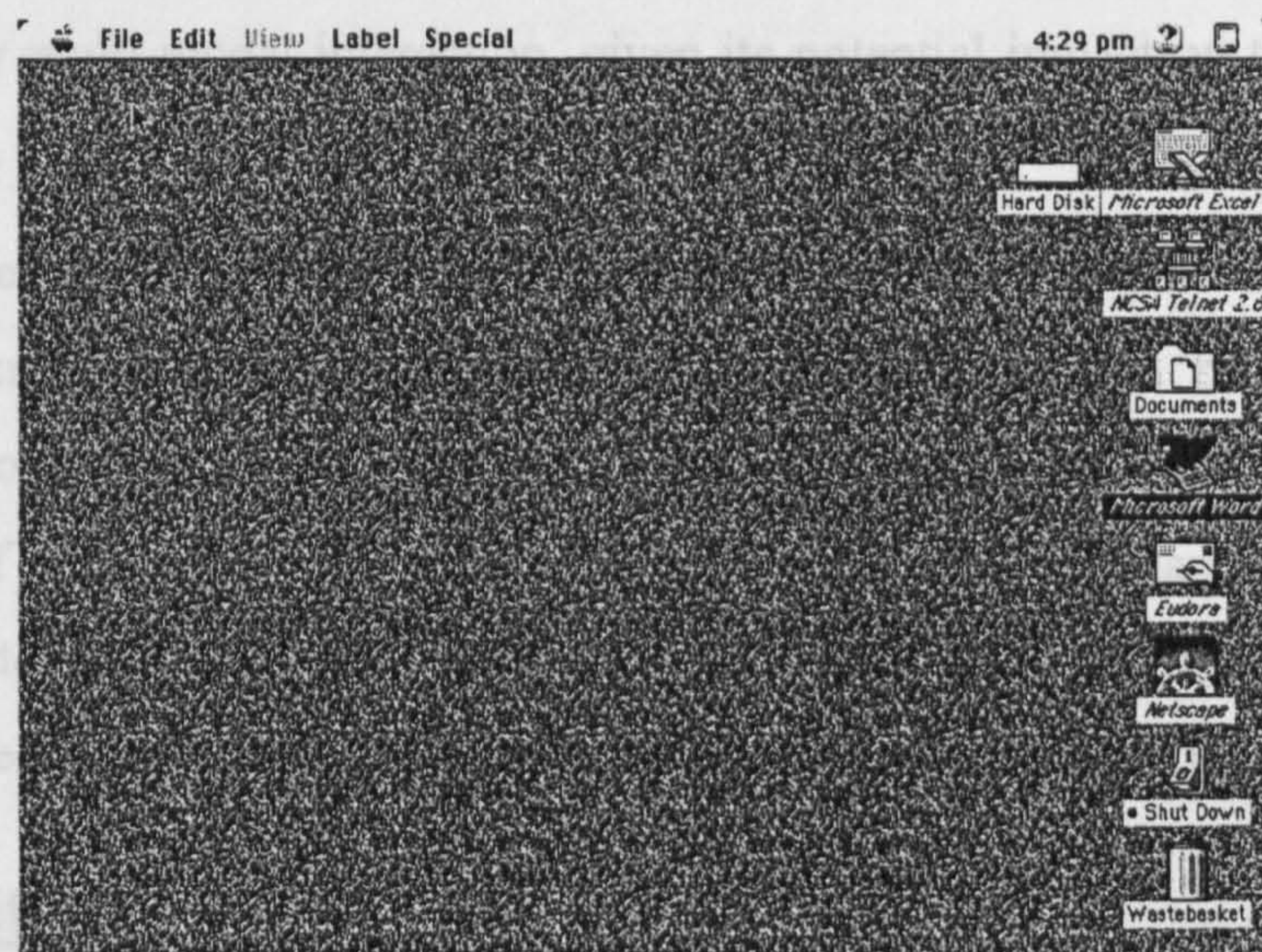


Figure 4-2: Menu options on Macintosh

There is a multi-coloured Apple logo, then a line of menu options. To a user used to a Windows platform, the Apple logo would be just that: an Apple logo. It would not have any inherent meaning other than that of decoration. However, in fact the Apple logo is not decorative but functional, in that it is an icon which when selected displays a list of options regarding the state of the computer. There is therefore a potential mismatch between what the user is receiving, i.e. that this is a decorative logo, and what the system is expressing, i.e. an icon.



Mismatches can therefore seriously affect the interaction by distorting the information flow by omission or addition. By examining the interaction from the perspectives of both the user and system (and additional users and systems if necessary) potential mismatches can be identified and rectified. This was established by Johnson & Nemetz (1998) in their analysis of a web site, which recorded the difficulty of the user discovering the actual function of certain animated features. Unless it is clear what is being expressed, mismatches in the reception will occur.

#### **4.4.6 Redundancy**

Redundancy has been identified as a key issue in multi-modal usability (Johnson & Nemetz, 1998), (Mills & Noyes, 1999) and occurs where exactly the same information is being expressed in two or more modalities. This can ensure that information is correctly conveyed, by conveying it in more than one manner at a time, but can on occasion be unnecessary, causing overload. As such it is an important property to identify in any multi-modal interaction, given its potential impact on the efficiency of the interaction.

As an example, refer back to the loquacious Scotsman and imagine that both volume and speed increased if he was speaking in a hurry. He would be expressing two acoustic-symbolic-dynamic modalities. However, the recipient would only need to identify one of those modalities in order to realise that the Scotsman was in a hurry. It would not matter whether it was volume or speed: either would do. Therefore one of the modalities is redundant, since it is not needed to convey the information.

### **4.5 Modality Clashes**

In the previous chapter various cognitive issues were examined as relevant to multi-modal usability, and it was established that there are processing constraints which can interfere with usability. This section examines various types of such **clashes**. It should be borne in mind that there can be more than one clash occurring at any given time. The clashes, with the exception of the first type, are user-dependent in the examples given. Since these clashes have not been defined in specific detail, they have to be identified by the analyst individually. This section establishes their existence, and prompts awareness of the issues involved.

### **4.5.1 Physical clashes**

As demonstrated by the cognitive models in the previous chapter, users have various physical constraints which limit both their perception and response to stimuli. Computers can be considered in a similar manner, since hardware and software limitations can affect their performance. These are termed **physical clashes**. For example, a computer may not be able to process two streams of voice input simultaneously. Similarly, a user cannot say two different things at the same time. It is not possible to detail exactly when system clashes will occur, because this will depend on the configuration of each particular system. It is possible to identify human physical clashes, although users with special needs will have different problems.

For example, a user cannot look in two different directions at the same time. Only if items are within a field of vision can they be visually processed. This is obviously important if a system is expressing two items of information, in two different spatial locations, where the user would not be able to see them both at once.

### **4.5.2 Lexical clashes**

The processing restrictions on representations, as exemplified by the morphonological subsystem in ICS (Barnard and Teasdale, 1991), result in constraints on what a user can and cannot process lexically, with a user unable to process two different lexical inputs or outputs at the same time. Examples of this type of **lexical clash** include reading a book and listening to a sports commentary, writing an address and reciting the alphabet, or writing a letter and listening to the news on the radio. Only if the lexical information is the same and is presented at the same time would a user be able to do this.

### **4.5.3 Temporal Clashes**

The temporal nature of the modalities can have an effect on how well they are processed. It is difficult to combine two dynamic modalities unless the modalities relate closely to each other, and can be thought of as a composite modality. An example of this is reading someone's lips and listening to what they say. If the two modalities (lip-movement and speech) are co-ordinated, with no time delay or synchronisation problems, then these two modalities are combined and do not interfere with each other.



However, if there is a time delay between the two, it becomes extremely difficult to process the modalities, and problems can occur (Barnard and Teasdale, 1991).

	DYNAMIC	CONTINUOUS	DISCRETE
DYNAMIC	possible interference	no interference	possible interference
CONTINUOUS	no interference	no interference	no interference
DISCRETE	possible interference	no interference	possible interference

**Table 4-7: Temporal clashes**

The same problem can occur with combinations of dynamic and discrete modalities, and discrete and discrete modalities. This is due to the user focusing on one of the modalities, and thus losing the information given in the other modality. For example, imagine that a person is listening to a radio commentary, and the doorbell rings. For a moment, attention will switch from the commentary to the doorbell, and the commentary is not processed for that brief instant. Table 4-7 clarifies this.

There is no problem when combining continuous modalities with other continuous modalities, or with discrete or dynamic modalities, since continuous modalities allow a certain degree of freedom of inspection.

#### **4.5.4 Semantic clashes**

Semantic clashes occur when information is provided in two modalities, which do not in themselves clash, but where there is a difference in the meaning of the two, causing interference because of the automatic cognitive processes involved. An example of such a **semantic clash** is the Stroop Effect as described by Preece et al (1994), where it is very difficult for someone to say a colour name if it is written using a different colour, because the automatic process of reading the word conflicts with the automatic process of perceiving the colour. Another example of semantic clash could be having the word 'butterfly' under the picture of a dog.

#### **4.5.5 Clash unless expert**

Clash unless expert occurs when the combination of modalities is such that only an experienced person would be able to either receive or express them properly. This **clash**

**unless expert** cannot be explicitly defined, and must be identified by the analyst using good judgement as to what the user may or may not be capable of. This clash is based on the issues examined in the cognitive models in the previous chapter, where it was established that expertise is due to experience in allocating resources and anticipating where resources are needed, in order to deal with multiple stimuli. Preece et al (1994) describes a three-stage model of learning where users move from the cognitive stage of learning, through the associative stage where the connections are strengthened, to the autonomous stage where the skill is more automated and rapid. This is linked to the way in which knowledge is stored in the memory, with experts having more refined “chunks” available to them than novices (Preece et al, 1994). An example of a clash unless expert is trying to change gears in a car whilst going round a bend and talking to a passenger. An expert driver would have no problem with this, since the relevant knowledge is stored in specific chunks and can be applied autonomously, whilst a learner driver might find that combination impossible, due to having the knowledge in less refined chunks and without the benefit of refinement through practice at the associative stage.

## **4.6 Summary**

This chapter has presented a theory of multi-modal usability based around a new definition of modality, a taxonomy for identifying modalities, the explicit description of modality properties, and the cognitive restrictions presented as possible clashes.

The new definition approaches the issue of modality in a novel manner by including temporal aspects within a compound definition encompassing sense and information form. This allows for greater flexibility of application, as well as precision of identification, and allows the implications of temporal attentional resources to be considered.

In order to identify modalities according to the new definition, a fully faceted classification system in the form of a taxonomy is presented, with three dimensions, sense, information form and temporal nature, corresponding to the three definition elements. Each dimension has three categories: sense with the categories of visual, audio and haptic; information form with the categories of lexical, symbolic and concrete; and temporal nature with the categories of discrete, continuous and dynamic. These were chosen to provide within-group homogeneity and between-group



heterogeneity and avoid overlapping. The resulting property space of twenty-seven cells allows for a large coverage of the interaction space and identification of modalities, while the classification is small enough to be easily applied. Each cell represents one taxon, indicating one particular instance of modality. The modality is named after the intersection of these categories, building the category names into a compound modality name following an analytico-synthetic approach.

The four areas of importance identified in chapter two, namely: identification and granularity; user and system; combination; and time and ordering; have been used to identify important modality types and properties. Modalities are identified as either User or System in type, and can be further typed as Expressive or Receptive. Their granularity is explored by defining them as composite, atomic or dependent. The possibility for error in the expression and reception of modalities is explored in the modality property of mismatch, and the use of more than one modality to convey identical information is explored in the modality property of redundancy.

The cognitive restrictions from chapter two are included in the form of five clearly explained modality clashes of the form physical, lexical, temporal, semantic, and clash unless expert.

In the next chapter this definition, taxonomy and theory is used as the basis for a methodology for supporting the identification of multi-modal usability issues.

# **5. DEVELOPMENT OF EMU, A MULTI-MODAL METHODOLOGY**

## **5.1 Introduction**

This chapter addresses the research issue of the lack of appropriate approaches to apply a theoretical understanding of multi-modality effectively by developing a methodology for Evaluating Multi-modal Usability (EMU). Previous chapters discussed issues of multi-modal usability with reference to existing research and cognitive constraints, culminating in a new definition, taxonomy and theory. Although this adds to an understanding of multi-modal usability, it is not enough to have these in isolation; they must be incorporated into a methodology in order to provide a context within which the usability or otherwise of an interface can be determined. As discussed in chapter two, existing methodologies, although able to consider some aspects, have not been developed explicitly to analyse the usability of multi-modal systems and can therefore have flaws and omissions. This research does not claim that any methodology will be able to deal with all the relevant properties of an interface: however, there is a case for developing an approach with the explicit aim of looking at multi-modal issues.

Inevitably, the development of a methodology is an iterative process, with various issues that must be considered: firstly, which properties and elements should be incorporated; secondly, what form of representation should be used within the methodology; and thirdly, the overall structure of the methodology, guiding its application and the kind of issues to be addressed. In this chapter, the issues that EMU should include and address are first examined. These are based on those identified in the previous chapters. Issues of representation are examined, by means of a case study of answering a telephone, resulting in a notation that can adequately represent the important aspects. Issues of structure are then examined, incorporating aspects of other approaches where appropriate, and a methodology of eight stages presented. This is considered to be a methodology rather than a method, according to the Concise Oxford Dictionary definition of methodology as a system of methods used in a particular field (Pearsall, 1999). The methodology is then used to assess the usability of the robotic arm from



chapter two, in order to examine how successful EMU is in identifying usability issues, and any problems from using it. Subsequent chapters will examine this methodology applied to the analysis of other multi-modal interactive devices.

## **5.2 Important Properties**

Various properties and issues were identified in previous chapters as being relevant to multi-modal usability. If a methodology is to support the usability analysis of multi-modal interaction, then it should include these elements.

### **5.2.1 Properties from Comparative Case Study**

The case study discussed in chapter two analysed five different modelling approaches (STN, CW, GOMS, PUM, Z) with reference to a portion of a multi-modal interface. From this study various issues of importance which should be incorporated into the development of a methodology were derived, and are summarised:

- The identification of issues is dependent upon the ability of the analyst; therefore, the approach should support the explicit identification of issues
- There is a need for well-defined procedures which can be applied in a straightforward and constrained manner
- The method, in terms of structure and notation, must be straightforward to learn and apply

### **5.2.2 Properties from existing theories of modality**

The examination of existing theories and approaches to multi-modality identified four significant issues, namely: identification; user and system compatibility; combination; and time and ordering.

A methodology should support the identification of a modality, in order that the basic elements of the interaction can be clearly examined. If the modalities cannot be identified, then the analysis of the interaction, and the information and constraints within the interaction, will suffer. Related to this is the issue of granularity and level of abstraction. Modalities often need to be considered at different levels of detail, so a methodology should support differing levels of granularity, depending on the level of detail needed in a particular analysis.

A methodology should be able to examine both the user and the system and the interaction and flow of information between the two. By representing both the system and the user, the potential usability issues can be examined from both perspectives, leading to a richer understanding of the process. The background profiles of both the system and the user, and their differing constraints, should be modelled.

Since modalities can be used both individually and in combination, and the use in these ways has different effects, the methodology should support the usability analysis of these circumstances. The methodology should allow for the identification of instances of modality combination, and of the different properties of combination and redundancy.

The methodology should be able to support the temporal and ordering aspects of the modalities, by detailing how modalities are ordered and the implications of the different temporal structures of the modalities on the usability of the interaction.

### **5.2.3 Cognitive Properties**

Various cognitive constraints relevant to multi-modal usability were identified as “clashes” in the previous chapter. These clashes: physical clashes; lexical clashes; temporal clashes; semantic clashes; and clash unless expert; need to be supported by the methodology, either implicitly by providing a means of identifying situations where these clashes may occur, or by explicitly supporting their identification.

### **5.2.4 Properties relating to the new definition**

The methodology should support the identification of modalities according to the new definition proposed, which takes a three-part form composed of sensory nature, information form and temporal structure. The modalities should be described as expressive and receptive, as initiated by the system or user, or received by the system or user. The identification of the mismatch of modalities should be supported, and it should be possible to represent modalities as atomic, composite and dependent, and redundant.



### 5.3 Robotic Arm Example

The definition of modality was applied to a portion of the example interface of the robotic arm to see what leverage being able to identify the modalities used gave to the analysis, and to provide a starting point for the development of the methodology.

A full listing of the modalities was written (see Appendix B). A portion of the listing, with an associated explanation in natural language, is shown in Figure 5-1, detailing the stage from when the robotic arm actually starts moving to when the user starts a new interaction cycle to move the arm in a different direction. There are two possible means of input by the user; using gesture when the flashing cursor is beneath the appropriate option on the screen, and using voice to articulate the desired option name.

This listing was straightforward to write, based as it is on natural language and the bare description of the modalities used at each stage in the interaction. The very act of writing the listing gave an opportunity for insight, due to having to write down what was happening in depth at every stage. Therefore, the process of examination was by itself of use, regardless of any further issues identified from the listing.

The listing allows the identification of system and user activity clearly, which makes it possible to see what is happening at any one time, and provides a starting point for examining issues of overload of any particular sense, and the choice of modalities, in order to ascertain if a particular modality is the most suitable to be used at this point.

By explicitly writing out both the expressive and receptive modalities, of both the system and the user, it is possible to check that they can be linked. This allows for the identification of occurrences where expressive modalities are not identified by the user or system, and thus not received, and ensures that each stage of the interaction is looked at in terms of both parties.

However, the listing does have several problems, mainly relating to its superficial nature and the fact that this is at present more a notation than a methodology with clearly defined procedures designed to identify usability issues. The lack of explicit consideration of the user and system characteristics and knowledge means that there is no way that the analyst can tell whether the interaction will fail due to a mismatch of modalities, or lack of resources, either technical or cognitive. Therefore, although it could be useful to an analyst with a high degree of craft skill, it does not allow for

usability issues to be quickly identified, and does not answer deeper questions as to what conflicts may arise except at the intuitive level. For example, vision is the sensory channel used most by the user in the interaction, and since users cannot look in two places at once, a gesture system for stop might have problems because the user might have to move their body to look at the arm, which might be interpreted by the system as a command. The listing does not support the identification of this kind of insight: it is left to the craft skill of the analyst.

---

Initially, the menu shows three choices:

**Go, Speed Level, End**

The system at this point is using two modalities to communicate, which are termed expressive modalities. The menu display of the three option choices is vis-lex-cont, meaning a visual display of lexical information that is unchanging. The cursor is moving continuously in discrete steps under the options and is therefore vis-sym-dyn, a visual occurrence of a symbolic form which is moving and changing.

The user is receiving information in two modalities, or has two receptive modalities. These are the menu display in the form of vis-lex-cont, and the cursor movement, in the form of vis-sym-dyn.

**System Expressive:**

Menu Display  
(vis-lex-cont)  
Cursor  
(vis-sym-dyn)

**User Receptive:**

Menu display  
(vis-lex-cont)  
Cursor  
(vis-sym-dyn)

The option **Go** is then chosen. To do this, the user can do one of two things: make an appropriate head movement when the cursor is under the correct option, or state the word "go". The head movement is an expressive modality of the form hap-sym-dis, meaning a single haptic movement with symbolic meaning. The voice selection is of the form aco-lex-dis, meaning a single acoustic utterance of a word.

**User Expressive:**

Head movement  
(hap-sym-dis)  
or name option  
(aco-lex-dis)

This results in the arm beginning to move. At this point, the user has two options. The user can wait for the arm to stop moving when it reaches its limit of movement, or the user can decide to stop the arm while in motion.

---

**Figure 5-1: Portion of initial modality listing of robotic arm**



The temporal aspect of the interaction is not fully conveyed, due to the way that the listing was divided into large discrete sections corresponding to the stages of the interaction, with the modalities written in a tabular form. Although not necessarily an issue for this example, it does not allow the interaction to be looked at a lower level of granularity, and for issues of modality ordering to be examined. The dynamic and changing nature of the interaction is not properly described by this representation and the interplay of the differing system and user modalities and their combinations can thus not be fully explored. Resulting from this, issues of redundancy and complementarity, although possible to identify at a high level of granularity, could not be examined in greater detail. Indeed, the granularity of the listing means that atomic and composite modalities are not considered.

Since this was merely a listing of the modalities and an accompanying description, there was no explicit consideration of the types of clashes proposed in chapter four. It would be possible to identify these clashes from the listing, but the listing as yet is descriptive as opposed to analytical.

There were also concerns about the use of natural language since a more precise way of describing the interaction and modalities may be appropriate.

There are therefore several reasons why this listing is not enough to support multi-modal usability analysis, mostly related to the fact that it was underdeveloped as a methodology and was concentrating more on the descriptive, notational aspects rather than an underlying theory or series of issues to address. From this process, various aspects were considered as essential to any further attempt at a methodology for multi-modal usability analysis, and were included in future iterations of the methodology. The next section examines the issues in turn, and describes how they were developed and finally incorporated.

## **5.4 Telephone Example**

After the initial application of the methodology to the robotic arm interface, the methodology was further developed, using a small case-study of answering a conventional telephone and having a conversation with the caller, taking the form:

A conventional telephone starts to ring. The user lifts the receiver, and speaks. The caller replies.

Although a small-scale example, it provided an opportunity to test various approaches, before scaling the methodology up. If the approaches had been unable to adequately represent such a small example, then it would be unlikely that they could be applied to a larger study.

There were two main strands to the development of the methodology: the form of representation used, and the structure of the methodology in terms of the component parts included.

## **5.5 Representation**

The issue of representation, in terms of ease, style and time taken, was of great importance in developing this approach. Many techniques are difficult and time consuming to apply and therefore are seen as not worth the effort, since the information can often be too general to usefully guide the design process. Formalisation is often a barrier to straightforward application (Blandford et al, 1998) and can lead to craft skill becoming more important. There can also be problems in determining which particular technique to use, since they can place different demands on the analyst in terms of the process to follow and the skills required (Blandford et al, 1998).

The form of representation used in an analysis can have an effect on what can be represented. For example, as Blandford & Young (1996) observe, it is easier to represent functionality in Z than timing constraints and inter-object communication, whereas process algebras and dialogue grammars have the opposite problems. Accordingly, several different types of representation were tried. It was important that the representation should be able to identify and represent both the individual modalities, and their temporal and ordering constraints, as well as being straightforward to draw. These were issues originally highlighted by the comparative study of the robotic arm as discussed in chapter two.

### **5.5.1 Tabular Representation**

The modalities were first represented in the description of the robotic arm interface by means of natural language and a list of the modalities used at each stage. An example of this representation of the modalities is shown in figure 5-2.



System Expressive:	User Receptive:
Menu Display (vis-lex-cont)	Menu display (vis-lex-cont)
Cursor (vis-sym-dyn)	Cursor (vis-sym-dyn)

**Figure 5-2: Natural language representation of modalities**

action	System	User
Phone rings	E aco-sym-cont	R aco-sym-cont
User picks up handset	E aco-sym-cont terminates	E hap-con-dyn R aco-sym-cont terminates
User speaks into handset	E aco-con-cont (line is quiet, showing that link is made) R aco-lex-dyn (user talking)	R aco-con-cont (user knows that line is live) E aco-lex-dyn E hap-con-cont (handset in hand)
Caller responds	E aco-lex-dyn	R aco-lex-dyn E hap-con-cont

**Table 5-1: Tabular representation of modalities**

This did not, however, show clearly what action was related to these modalities. Therefore the first iteration of the methodology, based around the analysis of a ringing telephone, included the actions, using a tabular approach, an example of which is given in table 5-1. In this example the action of the user picking up the handset is shown as being dynamic. However, as per the discussion in Chapter Four, it could have been broken down into a series of smaller discrete actions, such as user reaching for handset, grasping handset, and picking up the handset. For the purposes of this analysis, breaking it down to a lower level was not necessary, so it was kept as one dynamic action.

The tabular representation is similar in form to User Action Notation (UAN) (Hartson et al, 1990). This representation categorises user interface events, in terms of either the

user or the computer, and has a simple syntax and easy structure, with columns allowing the reader to keep track of what is happening (Johnson, 1996). UAN is a notation able to represent many different issues, including sequence, iteration and concurrency, and thus could potentially be very useful for examining ways of representing modalities.

However, despite its advantages, UAN was not comprehensive enough for modelling certain essential aspects of multi-modal interaction. The table above makes it clear which modalities are associated with an action. However, the tabular form, with the action separated into discrete blocks, meant that the interaction is still being written at a high level of detail, and is unable to represent issues of dependency. For example, in the above table, when the user picks up the handset the phone stops ringing immediately. Yet there is no way of representing this dependency, that the first modality causes the change in the second. This is an important part of multi-modal interaction, as discussed below, and is why UAN was not considered further as a suitable representation.

### **5.5.2 Diagrammatic Representations**

Since the tabular approach was unable to deal with particular issues of ordering at a low level, it was decided to see if diagrammatic representations could offer a clearer view of the interaction between the different modalities and the user and system. Therefore, before the second iteration of the methodology, various diagrammatic options were considered, as discussed below.

Critical Path Mapping (CPM) is the representation format used by CPM-GOMS (John & Kieras, 1996) and was attractive because of its clear way of showing what was happening in the interaction. However, it was quickly established that there were too many modalities to be effectively represented in this fashion.

It was too confusing to represent modalities in any more detail than that of the sensory dimension, thus obscuring their temporal and informational aspects. In addition, there was no way of efficiently representing both expressive and receptive modalities without creating an overly complex and confusing diagram, which would negate the advantages of a graphical approach.

Since Petri Nets (Johnson, 1995) are a form of directed graph which can be used to represent and reason about timing properties of system events and can support the engineering of concurrent systems at various levels of granularity depending on the



hierarchy used, they were felt to provide a possible way of representing multi-modal interaction.

However, it was found that Petri Nets are not salient enough, and do not fulfil the requirements for a representation as set out in the earlier sections. It was difficult to model the user and system separately without going to a very low, and consequently very complicated, level of granularity. Since the user and system were not explicitly separated the flow of interaction between the two is not clear as the Petri Nets diagram concentrates on the overall flow of events and conditions rather than the interplay between the two agents. As such the Petri Nets were unable to convey that modalities can be expressed as well as received. There was also potential difficulty in representing more complex interactions with multiple events triggering several actions.

Diagrams are suitable for some small and simple examples, but they are difficult to scale up to represent more complex interactions. One of the main problems is in the representation of time. For example, picking up the telephone handset changes to holding the handset at a point undefined in relation to the other modalities currently happening, e.g. line quiet, etc. All that can be said is that the modality of holding the handset occurs after the handset has been picked up. Nothing can be stated about when the person starts talking in relation to holding the handset, because these modalities may or may not overlap. Therefore, there are “hard” and “soft” timelines, which results in a timeline that is not “real” but has more to do with ordering. This can be very difficult to represent diagrammatically.

Another issue is the representation of individual modalities. Neither approach could adequately represent the modalities in enough detail, not only in terms of their component parts, but also in terms of expressive and receptive.

### **5.5.3 Notational Representations**

Since tabular and diagrammatic representations could not deal adequately with the issues of time, ordering and dependency, as well as clearly identifying modalities, it was decided to try a notational representation. Accordingly, the second iteration of the methodology used a different way of representing the flow of the interaction, where the modalities were listed in order, clearly labelled, and the ordering indicated by the symbol  $\rightarrow$ . An example section of this is given below in figure 5-3.

---

[S E aco-sym-cont] \*phone rings\*  
-> [U R aco-sym-cont] \*user hears phone\*  
-> [U E hap-con-dis] \*user picks up phone and places next to ear\*  
-> [S R hap-sym-dis] \*system notes user has removed handset\*  
-> [S E aco-sym-cont terminates] \*phone stops ringing\*  
-> [U R aco-sym-cont terminates] \*user hears phone has stopped ringing\*, [U E hap-con-cont] \*user is holding handset to ear\*  
-> [S E aco-con-cont] \*line is quiet\*, [U E hap-con-cont] \*user is holding handset to ear\*

---

### Figure 5-3: Initial notational representation of modalities

This representation uses symbols to represent the ordering of the modalities, and a stylised convention to describe the modalities, as well as accompanying notes in natural language. It is simple, but effective, in that the flow of the interaction is easily seen. However, the use of the symbol -> is flawed because it confuses the issues of consequence and consecution, which may cause inaccuracies. For example, the user picking up the telephone and placing it next to the ear comes after the modality where the user hears the telephone, but is not actually dependent upon it. The user may have decided to pick up the telephone anyway, in order to make a call themselves. Therefore, using this way of representing ordering, issues of dependency are not made transparent, and modalities may be misconstrued as being dependent when in fact they are merely sequential in time.

In order that the modalities might be clearly and unambiguously identified, descriptions in natural language were placed after the notational descriptions of the modalities enclosed by \*. This mix of natural language and notation allows the reader to track what is happening more easily.

#### 5.5.3.1 Preconditions

In order to clearly separate the issues of ordering and dependency, the idea of preconditions was incorporated from PUM (Blandford & Young, 1996) in order to show what has to be true for the next thing to be true. This allowed for ordering to be



represented by order, and dependency to be represented by preconditions. An example section of the interaction, with comments, is shown in figure 5-4.

---

[SE aco-sym-cont] \*phone rings\*

The phone rings.

[UR aco-sym-cont] \*user hears phone\*

preconditions: [SE aco-sym-cont] \*phone rings\*

[SE aco-sym-cont] \*phone ringing\*

The user hears the phone, which has to be ringing for the user to hear it, and the phone continues ringing.

[UE hap-con-dis] \*user picks up phone and places next to ear\*

precondition: [UR aco-sym-cont] \*user hears phone\*

The user picks up the phone, which the user has to have heard. Since the phone has been picked up, the phone is no longer ringing and the user is no longer hearing the phone ring.

---

**Figure 5-4: Notation with preconditions and comments**

This way of representing the interaction not only allows preconditions to be represented, but also allows other ongoing modalities to be represented at the same time, for example, while the user hears the phone ringing, the phone is ringing.

### **5.5.3.2 Numbering**

With the use of preconditions, the ordering and dependencies of the modalities in the interaction are clearer. However, the interaction is written as one mass, whereas separating it out into discrete stages of interaction would give an insight into the underlying structure of the interaction and make it clearer to the analyst. Therefore, the third iteration numbered stages of the interaction. The numberings are done according to the level of granularity considered appropriate at the time. Stages can be expanded or contracted according to the needs of the analyst. The listing in figure 5-5 gives an example.

Modalities which occur at around the same time can be represented as happening at the same stage, with preconditions giving more explicit information about orderings.

---

1.	[SE aco-sym-cont] *phone ringing*	[UR aco-sym-cont] *user hears phone ringing* precon: [SE aco-sym-cont] *phone ringing*	
2.	[SE aco-sym-cont] *phone ringing*	[UR aco-sym-cont] *user hears phone ringing* precon: [SE aco-sym-cont] *phone ringing*	[UE hap-con-dis] *user picks up phone* precon: [UR aco-sym-cont] *user hears phone ringing*
3.	[UE hap-con-cont] *user holds phone* precon:[UE hap-con- dis] *user picks up phone*	[SE aco-sym-cont] *line quiet* precon:[UE hap-con-dis] *user picks up phone*	[UR aco-sym-cont] *user hears line quiet* precon:[UE hap-con-dis] *user picks up phone*

---

**Figure 5-5: Listing showing use of numbering**

### 5.5.3.3 Choice

This way of representing the interaction makes the stages of the interaction clearer, but does not allow choices to be represented. It was decided to use AND and OR operators to represent occasions where modalities happened simultaneously, or where there was a choice between modalities. This was incorporated into the fourth iteration, based around the previous example of a telephone, this time with the addition of a light that flashed when the phone rang. A section of the listing is given in figure 5-6.

The telephone is both ringing and flashing a light. Since both of these are happening at the same time, the logical AND is used to connect them. The user may only react to one of these modalities. Since this might be hearing the ringing or seeing the flashing, the logical OR is used between these modalities, to indicate that only one of them might happen.



However, AND/OR is used only to indicate where modalities are happening together or where there is choice. For example, in Figure 5-6, the SE modalities are linked by the use of AND to show that they are happening at the same time. Below, there are two UR modalities linked by an OR to show that only one or the other is being used. If AND and OR are used together in a single statement then normal left to right precedence is assumed unless brackets are used to indicate otherwise.

---

```
2. [SE aco-sym-cont] AND [SE vis-sym-cont]
    *phone ringing*      *phone flashing*

[UR aco-sym-cont] OR [UR vis-sym-cont]
*user hears phone ringing* *user sees phone flashing*
precon: [SE aco-sym-cont] precon:[SE vis-sym-cont]
*phone ringing*          *phone flashing

[UE hap-con-dis]
*user picks up phone*
precon:[UR aco-sym-cont] OR [UR vis-sym-cont]
    *user hears phone*   *user sees phone flashing*
```

---

**Figure 5-6: Listing showing use of AND/OR to denote choice**

#### **5.5.3.4 Large Interfaces**

The example of answering a telephone given above is simple and straightforward, and does not adequately reflect the complex nature of many interactions using many more modalities. It is necessary to be able to represent large numbers of modalities in an effective manner. To achieve this, the importance of granularity of representation is emphasised. This is explained in more detail using the example of a task using the word-processing package Word 6 for a Mac to change a paragraph of an open document from Times font to Courier font.

---

1.  
[SE vis-sym-cont]  
\*scroll bars\*  
AND  
[SE vis-sym-cont]  
\*menu options\*  
AND  
[SE vis-sym-cont]  
\*icons at top of text space\*  
AND  
[SE vis-sym-cont]  
\*information strip at bottom\*  
AND  
[SE vis-sym-dyn]  
\*mouse cursor moving\*  
precon: [UE hap-con-dyn]  
    \*user moving mouse\*  
AND  
[SE vis-lex-cont]  
\*text\*  
  
[UE hap-con-dyn]  
\*moves mouse\*  
  
[UR vis-lex-cont]  
\*scanning text\*  
OR  
[UR vis-sym-dyn]  
\*watching mouse cursor\*  
precon: [UE hap-con-dyn]  
    \*user moving mouse\*

---

**Figure 5-7: Portion of a listing for a large interface**

The Word 6 word-processing package for the Mac has a lot of information on the screen at a time. There is the text space, where text is entered; there are the scroll bars which



border this space; there is the menu at the top of the window giving options such as File, Edit, View; and there is a row of icons and text information at both the top and bottom of the text space. There are thus many visual modalities being expressed by the computer at once, which can make it quite complicated to write out the interaction modality listing.

Effectively what will happen is that the user will only be looking at certain of the modalities displayed, but the system will be expressing all of them. Therefore, the ones the user is not looking at can be expressed at a very high level of abstraction, in order to show that they exist, but without unnecessary detail.

This is illustrated by a portion of the interaction sequence in figure 5-7.

By only writing the composite modalities rather than all the atomic modalities the list of modalities is still somewhat lengthy but is still short enough to allow the analyst to focus in on what is happening.

As the interaction progresses and the focus of the interaction switches, the high-level modalities can be described at a lower level of detail as in figure 5-8.

Therefore large numbers of modalities can be accommodated by the notation, through the use of composite modalities and granularity of representation.

The system is displaying information on the screen. This information is represented at a very high level of abstraction. The user is only looking at one of two of these modalities, and is moving the mouse.

---

the section of the screen giving the menu options [SE vis-sym-cont] would be written:

[SE vis-sym-cont] \*area containing menu options\*

[SE vis-lex-cont] \*File\*

[SE vis-lex-cont] \*Edit\*

[SE vis-lex-cont] \*View\*

[SE vis-lex-cont] \*Insert\*

[SE vis-lex-cont] \*Format\*

[SE vis-lex-cont] \*Tools\*

[SE vis-lex-cont] \*Table\*

[SE vis-lex-cont] \*Window\*

[SE vis-lex-cont] \*time display\*

[SE vis-sym-cont] \*system option\*

[SE vis-sym-cont] \*help option\*

[SE vis-sym-cont] \*current packages option\*

and the rest of the screen could be described as:

[SE vis-sym-cont] \*scroll bars\*

[SE vis-sym-cont] \*icons at top of text space\*

[SE vis-sym-cont] \*information strip at bottom\*

[SE vis-sym-cont] \*mouse cursor\*

[SE vis-lex-cont] \*text\*

---

**Figure 5-8: Listing showing more detail of large interface**

#### 5.5.4 Analysis

This notation in its final form relates back directly to the original aims expressed at the beginning of the chapter. Modalities identified in the interaction can be clearly represented, both in terms of the modality definition which allows their components to be reasoned about, and also using natural language which provides an easier level of description for the analyst, which was not possible with the five approaches examined in chapter two. The differences between system and user modalities, and their expressive and receptive natures, are explicitly noted, and the temporal and ordering aspects are separated and represented appropriately.



Differing levels of granularity can be accommodated within this style of notation, using composite and atomic modalities to represent those aspects of the interaction which are most relevant at any given point. The clarity of this way of representing modalities supports the analyst in identifying issues of mismatch, redundancy and complementarity, as well as the particular clashes that might arise.

Overall, the notation has been kept as simple and straightforward as possible, so that it takes less time to learn and apply. The style is not as easy as STN and CW in representation, but the notation is less formal than that used by PUM and Z. Since this notation uses a minimum of formalism, it does not have to be particularly precise. The representation of modalities in terms of their component parts and in natural language allows the analyst to keep track of the interaction, and the identification of modalities as system or user, and expressive or receptive, is done through only one letter each, minimising effort. The use of AND/OR operators follows basic Boolean logic, and preconditions allow the consequences of modalities to be explored. This notation will be examined in use in chapter six.

## **5.6 Structure**

In the previous section issues relating to the representation of the interaction in terms of modalities, their ordering and dependencies, were examined. Although representation is of importance in analysing the usability of an interaction, it needs to be placed within the context of a structure or procedure of application and analysis. As discussed earlier, the first stage in the development of a methodology was to apply the basic taxonomy to the interface of the robotic arm. This highlighted several issues, notably that the act of writing gave an opportunity for insight, and that a taxonomic approach by itself was too superficial, in that the usability issues had to be brought out by the skill of the analyst as they were implicit in what was written rather than being explicitly uncovered by the process.

The original comparative study of the robotic arm highlighted the difference between representation and methods, in that the methods examined were able to identify more usability issues than the representations, due to their directed nature which gives a structure to the analysis. For this reason, a purely representational notation was not enough. An overlying method was needed.

---

### **Stage 1. Define the task that is to be analysed**

The selection of this particular task as opposed to any other task has to be justified and explicitly stated. Is this a standard task for this device? Is the task complicated enough to show a wide selection of potential problems?

### **Stage 2. Modality lists**

At this stage all the modalities used by the user and the system should be listed, both descriptively, i.e. "phone ringing", and notationally, i.e. "SE aco-sym-cont". The level of granularity should be discussed at this point, in terms of the appropriate level of detail.

### **Stage 3. Define the user, system and environment profiles**

At this point, the analyst should explicitly state the characteristics of the user, system and environment that can reasonably be expected to have a bearing on the usability.

### **Stage 4. Profiles compared to modality listings**

The profiles of the user, system and environment are compared against the lists of modalities in order to check for obvious discrepancies.

### **Stage 5. Interaction sequence listing**

At this point the interaction should be written out in sequence and in terms of the modalities used. Check that for every expressive modality, there is a corresponding receptive modality. There will not always be a pairing, but in most cases there will be.

Number each part of the interaction sequence.

Write the modalities both notationally, with square brackets round them, in order that their constituent parts can be examined, and descriptively, within the star symbols (\*) in order that the analyst can keep track of which modalities are being discussed.

If modalities happen at the same time they should be connected with the logical **and**.

If there is a choice between modalities, they should be connected with the logical **or**.

Take into account any pre-conditions relating to ordering or the use of the visual sense.

### **Stage 6. Add in clashes, etc.**

Now that the modalities have been established, and the sequence and pre-conditions explicitly stated, look for the properties of the modalities and the potential clashes.

### **Stage 7. Assess the use of modalities**

The analyst should assess the modalities used at this point in order to see if any are under-utilised, and if any of the properties flag issues that should be addressed.

### **Stage 8. Final report**

Finally, write up into a report, assessing the overall usability of the system as a result of analysing the interaction in this way. Include conclusions and recommendations.

---

**Figure 5-9: The eight stage methodology**



Using aspects of the five approaches (STN, CW, GOMS, PUM, Z) studied in chapter two, and by iteration around the case study of the telephone, an eight stage methodology was developed. This is first presented in summary in figure 5-9, then each stage is examined in detail, with an indication of which usability issues are expected to be found at each stage.

### **5.6.1 Stage 1. Define the task that is to be analysed**

There were a number of initial decisions as to how the methodology should be structured. One primary decision was that the methodology should be task-based and procedural. This is in common with other modelling techniques, such as GOMS, PUM, and CW, which are all task-based to some extent: GOMS in terms of the goals and operations needed; CW in terms of the success or failure stories at each stage of the interaction; and PUM in terms of the knowledge needed. By taking a task-based approach, the analysis is constrained, in that it concentrates only on the tasks to be performed. The analysis is given a ready structure, which expedites the analysis, and thus speeds up the overall process. The flow of interaction between the user and the system can be examined in depth on a staged basis, and issues of load and combination, and time and ordering, also considered.

The explicit use of a task-based approach necessitates the definition of the task at the start of the analysis. This determines what is to be covered, and identifies the scope of the resulting analysis. The appropriateness of the task is also justified. If the task is not appropriate for analysis, due to it being infrequent, or less important than other potential tasks, the act of justifying this task might help make those issues transparent, and thus save conducting an inappropriate analysis. This follows the approach taken by CW, where the task is examined in terms of the participants and the routine nature of the task. This concentrates attention on the utility of the analysis from the start, thus helping to address issues of unwieldiness and inability to uncover relevant information. If a task can be justified as one that is useful to analyse, then it is perhaps more likely to uncover relevant usability issues than a task which concentrates on those aspects of the interaction which are less critical.

### **5.6.2 Stage 2. Modality lists**

One of the problems highlighted by the initial robotic arm analysis was the lack of explicit identification of the modalities, so that it was impossible for them to be assessed in terms of conflict and other issues. Therefore this stage of the methodology explicitly lists all the modalities used in the interaction, both descriptively, in terms of the information, and notationally, ready for the interaction analysis, as well as in terms of their expressive or receptive natures.

By listing which modalities are used the analyst can begin to investigate the sequence of modalities being used. This can be done iteratively, with the modalities first written in natural language and then transformed into the notation. Writing it at this stage of the methodology separates the modalities from the interaction sequence, thus minimising confusion. It also helps to concentrate the analyst on what level of granularity is actually important or appropriate, or how to model those modalities that have been identified. The analyst is therefore constrained into being explicit about what is being modelled and why, and what the analysis is trying to show. It can also serve to identify areas of confusion surrounding the modalities being used, acting as pointers to further analysis or the need for clarification. This relates back to the original aim of identifying individual modalities. Receptive and expressive modalities can be checked against each other for omissions.

### **5.6.3 Stage 3. Define the user, system and environment profiles**

In order that both the system and the user, and the environment in which they interact, can be examined in detail, system, user and environmental profiles were included in the methodology. This is based on the approach taken by CW and PUM. In CW, the user of the system has to be explicitly described, which promotes the user perspective. In this case, it is not constrained to CW's novice user: expert users can also be accommodated. Explicit consideration of the user enables issues affecting the user to be addressed. Usability issues are often user-specific, and different user groups will have different needs and characteristics.

The use of profiles was extended to include the system and the environment, in order that the constraints acting on all three can be considered such as, for example, that the system can only receive spoken input at certain audio levels, and that the user can see and hear with no difficulties. The other approaches, with the possible exception of CW,



do not take into consideration the environment in which the interaction occurs. However, this can potentially have a big impact on the conditions the user will be operating under, and hence affect the interaction to a large extent. By thinking about the environment explicitly, the analyst is not longer working in a vacuum but can consider such issues immediately.

#### **5.6.4 Stage 4. Profiles compared to modality listings**

Once the lists of modalities have been identified, and the profiles established, they can be compared to check for obvious mismatches. For example a system which relies on visual output is of little benefit to a blind user. Similarly, a system which relies on typed input is going to be difficult for a mobility impaired user. Therefore before the interaction sequence itself is analysed, there have been many opportunities to identify usability issues of concern. Instead of concentrating only on the interaction sequence, a more holistic approach to uncovering usability issues is taken, priming the analyst to uncover more issues when writing the interaction sequence.

#### **5.6.5 Stage 5. Interaction sequence listing**

This interaction-sequence listing relates back to the approaches taken by CW and GOMS, and is similar to PUM in that preconditions are made transparent. By following the task through the interaction on a step-by-step basis, usability issues can be addressed at every stage of the interaction with regards to the identification of modalities and their load and combination. This temporal sequencing allows for the overall structure of the interaction to be apparent, as well as issues surrounding individual modalities and combinations.

#### **5.6.6 Stage 6. Add in clashes, etc.**

Once the initial interaction listing has been completed, the cognitive constraints identified as clashes in chapter four, and the properties of modalities, can be added in. By going back over the analysis, the analyst has a greater opportunity to find issues which may originally have been missed. The explicit guidance for the analyst to look for these issues makes them more likely to become apparent. This is similar to the questions used in CW at each stage.

### **5.6.7 Stage 7. Assess the use of modalities**

So that the information contained in the interaction listing is used to reason about the usability of the interaction, this stage guides the analyst towards a critical assessment of the modalities used. This results in alternatives being considered, and a wider view of the modalities taken. Again, constraining the analyst to reconsider the information helps to draw out further usability issues.

### **5.6.8 Stage 8. Final report**

Finally, the analyst is expected to write the results of the usability analysis into a report. This helps to prevent the analysis from being merely an interaction sequence with little or no critical assessment. Everything is drawn together in this stage, and those issues that the analyst has found implicit in the methodology are explicitly detailed.

### **5.6.9 Analysis**

This methodology supports the analyst, not by bypassing any inherent skills, but by providing opportunities for insight. The stages of the methodology directly concentrate on those issues of importance to multi-modal usability. As such it overcomes some of the limitations of STN and Z, which have no guidelines, and CMN-GOMS, CW and PUM, which look for different kinds of issues.

By drawing attention to the task at the start, this methodology follows CW and PUM, in that issues relating to the task and domain/device issues can be investigated. The task-based structure gives a clearly identifiable starting point, unlike analyses using PUM and Z, where it can be more difficult to decide where to start. The interaction listing can be written on an iterative basis, and certain sections could be missed out if not felt to be of interest. The level of granularity can be adjusted so as to allow certain sections to be examined in more detail if required.

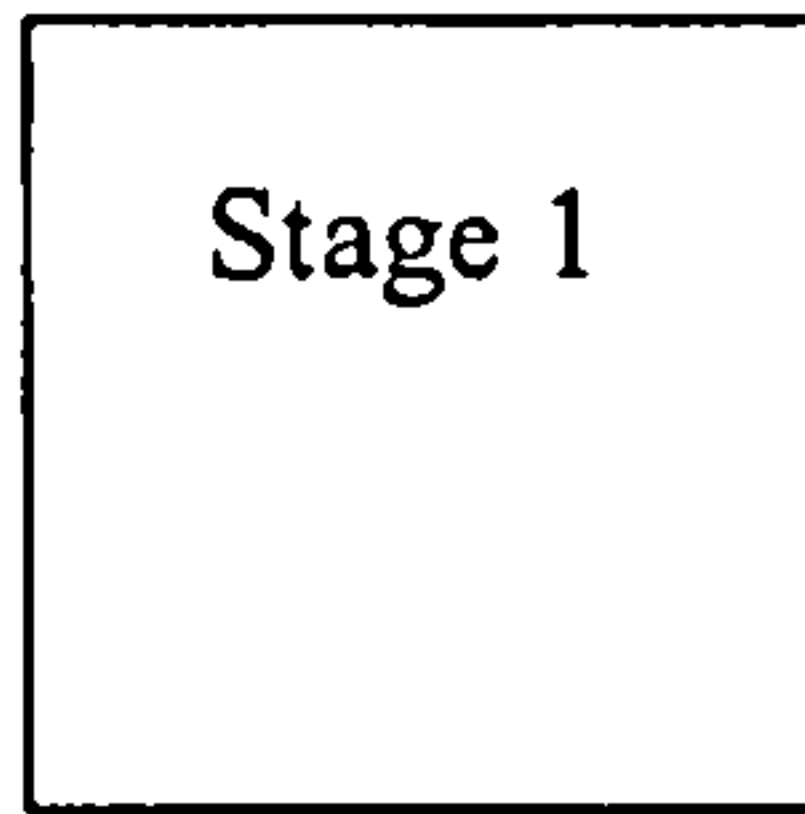
The different stages of the methodology encourage the analyst to re-examine the interaction listing as understanding and insight into the interaction grows. The ordering of the stages is not from 1 to 8, with every stage having to be done in that order. Instead, some of the stages can be completed in parallel, as Figure 5-10 makes clear. Stage 1 has to be done first, to define the task and scope. Then either Stage 2, or Stage 3, or Stage 5,



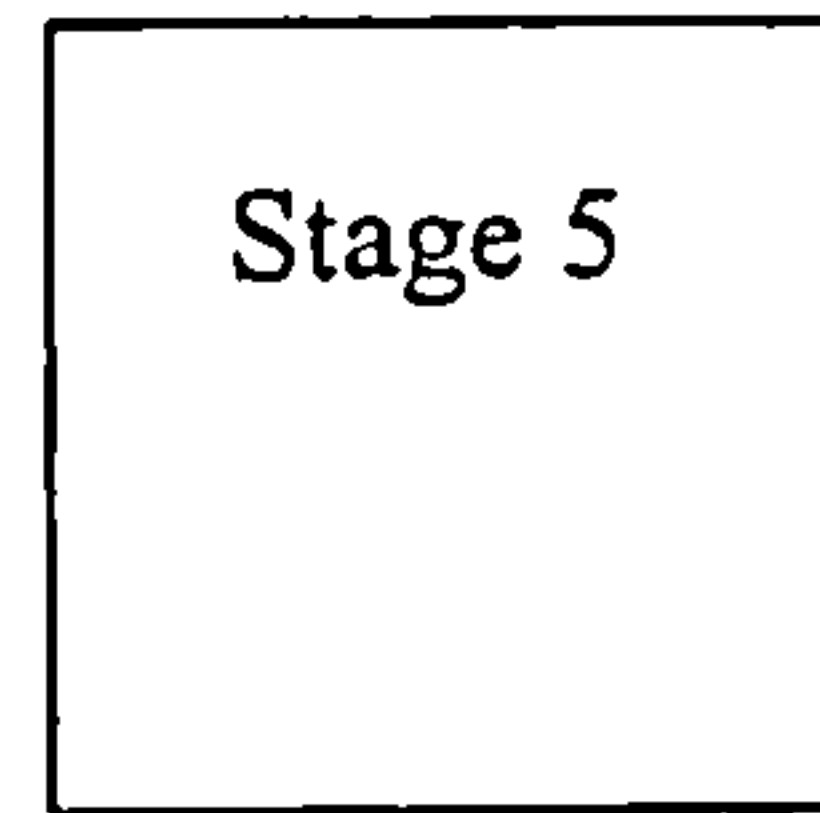
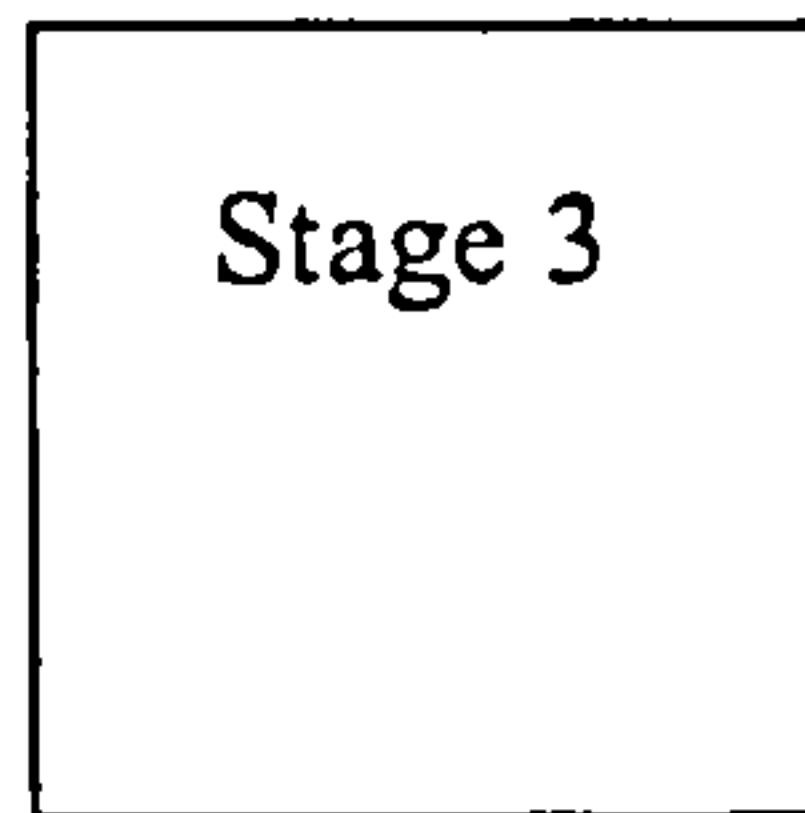
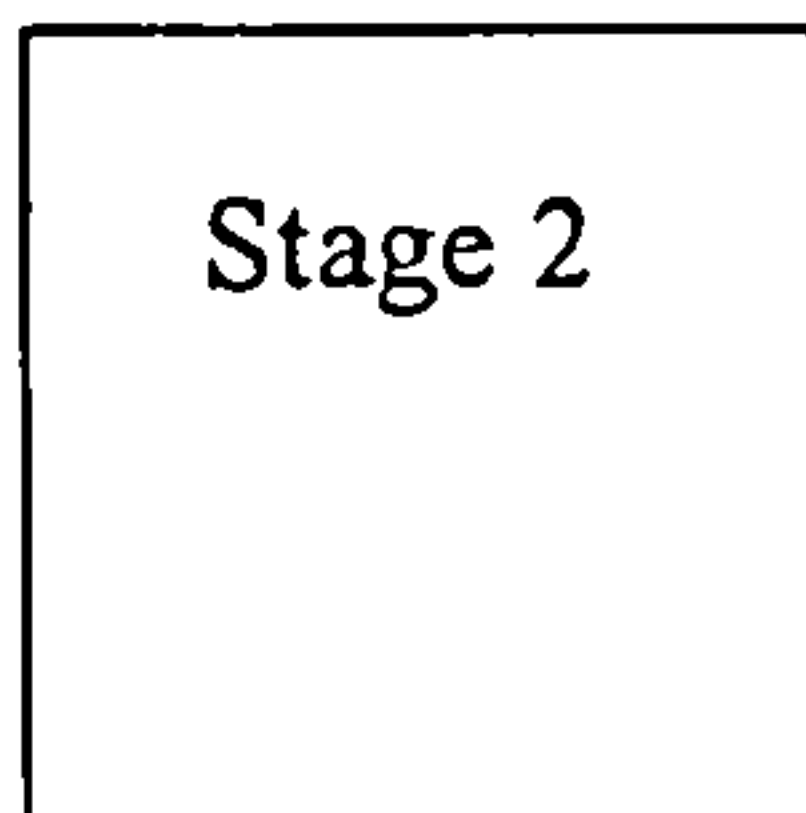
or all, can be completed. Stages 2 and 3 have to be completed before Stage 4, and Stages 6 and 7 can only be done once Stage 5 has been completed. Stage 8 is done last.

---

Stage 1 of EMU  
must be done  
first.

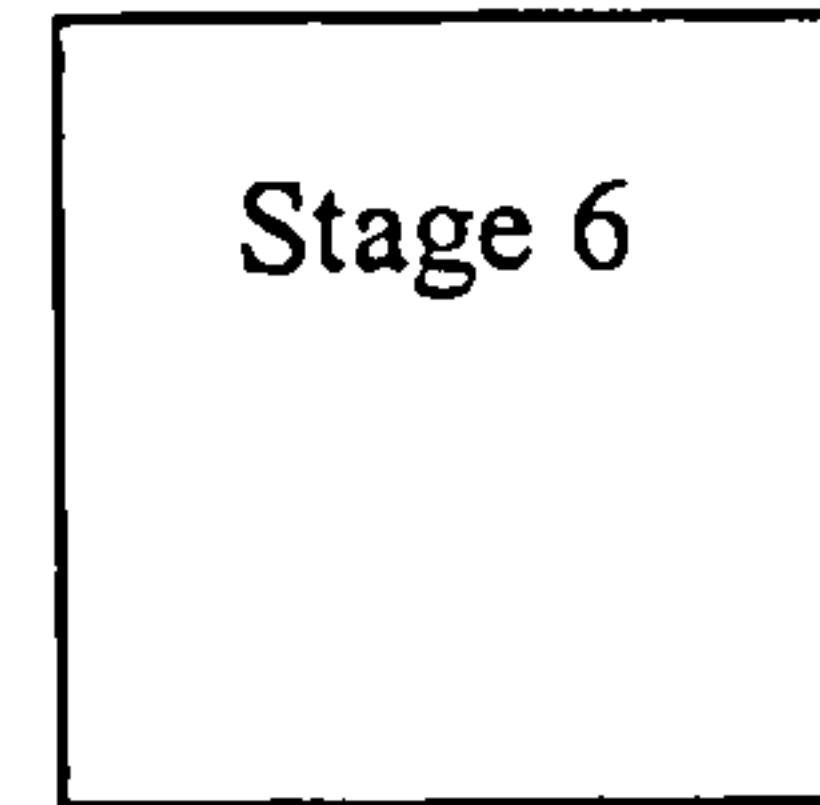
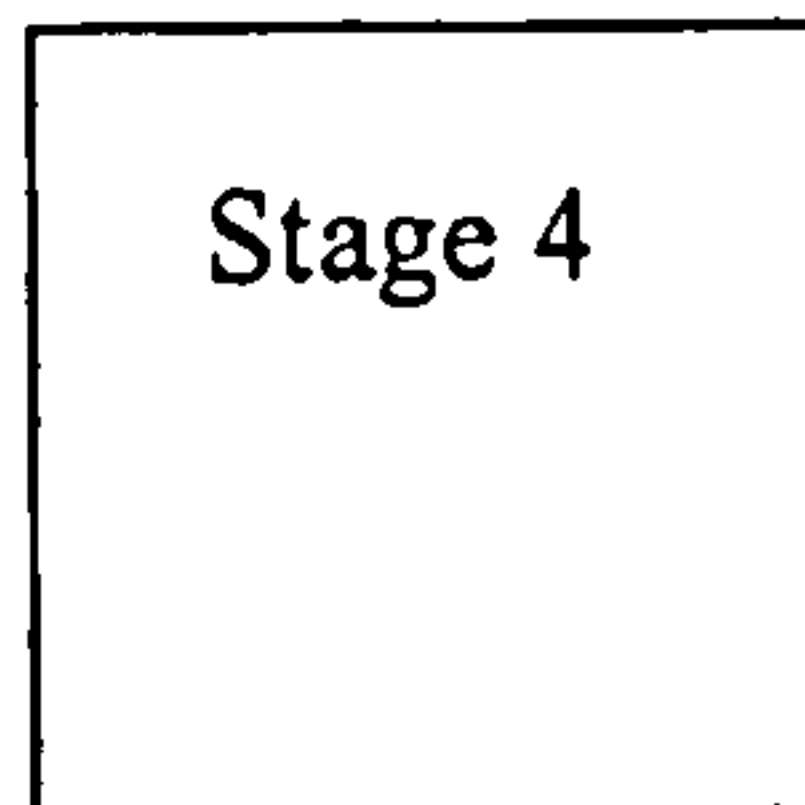


Stages 2, 3 and 5 can be done in any order after Stage 1.

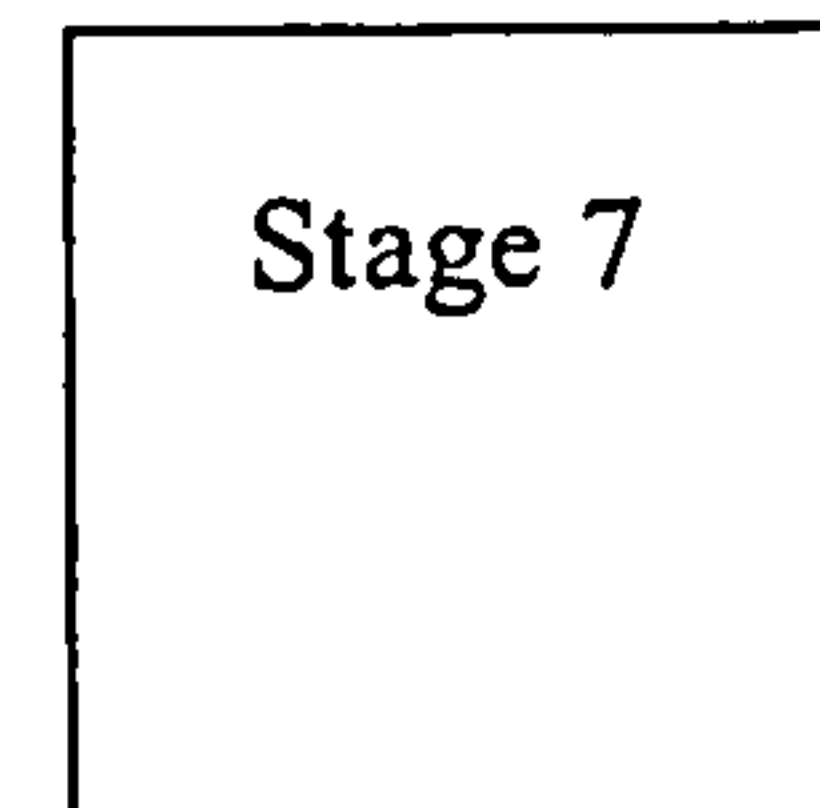


Stage 4 can only be done after  
Stages 2 and 3 are done.

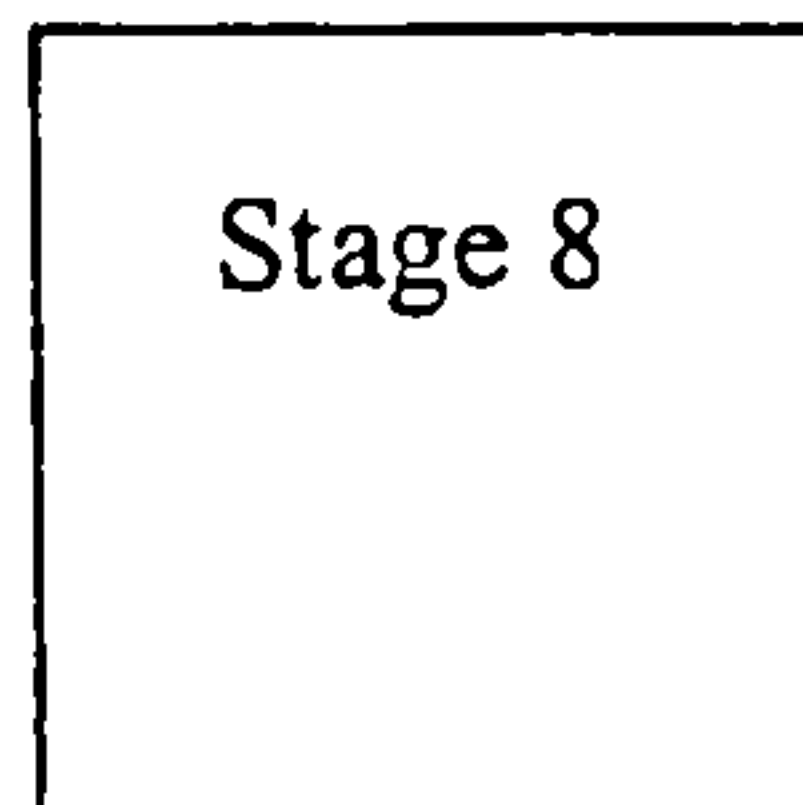
Stage 6 can only be  
done after Stage 5.



Stage 7 can only be  
done after Stage 6.



Stage 8 is done  
after all the other  
stages are  
completed.



---

**Figure 5-10: Ordering of EMU stages**

The notation is designed to be quick to use, with few items represented at any time: modalities, ordering, preconditions and choice. This allows iterations to be quickly completed and an overview of the analysis rapidly achieved, by first using natural language. The final representation is not as free as that used by STN and CW, but is less formal and constrained than that used by PUM and Z.

Both expert and novice users can be represented by this methodology, since the user profiles can be changed accordingly. Changes would affect the interaction sequence in terms of the potential for mismatch of modalities, confusion between what is expressed by the system and what is received by the user.

It is not claimed that this methodology is complete; indeed, there are certain issues which this methodology does not address, for example redundant actions and illegal states. The coverage and use of EMU will be examined later in this chapter, and in chapters six and seven, which report on its practical application. However, overall the methodology supports the explicit identification of issues, has well-defined procedures which can be applied in a straightforward and constrained manner, and is straightforward to learn and apply. It is able to concentrate attention on those issues of multi-modal usability which were identified as being of importance, namely those relating to: identification, user and system compatibility, load and combination, and time and ordering; and provides a basis for examining the cognitive restrictions on the interaction.

## **5.7 EMU Analysis of Robotic Arm**

In order to assess how effective EMU was in identifying usability issues, the methodology was applied to the portion of the robotic arm interface previously examined by other approaches in chapter two.

Ten usability issues were identified by the EMU analysis. Six of these had also been found by the previous analyses of the robotic arm, and four were new issues, as summarised in table 5-2. The analysis also allowed an opportunity for insight into problems regarding the application of EMU. The complete EMU analysis is available in Appendix C.



Key

5.7.1 Usability Issues

issue identified



CMN-GOMS and CPM-GOMS are included under the heading of GOMS

	PROBLEM	STN	CW	GOMS	PUM	Z	EM U
1	Long sequence of operators to move arm						
2	Inability to backtrack						
3	Difficulty of choosing between Move Arm or Move						
4	Lack of short cuts						
5	Continue versus Go: Continue seen as redundant						
6	Confusion over joint called Arm						
7	Gesture input with twice as many operations as voice because dependent on cursor movement						
8	Problem if head moved to look at arm while gesture system operational may be interpreted as a command						
9	If user pauses in middle of saying "Move arm"...						
10	If user engaged in conversation...						
11	Lack of feedback after Move Arm selected, no indication that whole arm is to be moved						
12	Problems of determining left and right, especially when arm contorted						
13	User cannot check direction choice until arm starts to move						
14	Time taken to interact with system to stop arm						
15	Similarity between moving joint and moving whole arm						
16	Illegal options						
17	Mismatch between way that arm works and way that user would move arm						
18	Not clear that End returns user to main menu						
19	End having two meanings						
20	Lighting conditions						
21	Difficulty for user to move field of vision						
22	User looking one way, menu options in other direction						
23	Difficulty of judging arm movements						

Table 5-2: Revised summary table of usability problems



### 5.7.1 Usability issues

The analysis of the task defined in stage one found that switching on the light switch involves planning and careful thought in order to establish how best to move the arm and joints, pointing to a possible conceptual problem for the user between their idea of the task, and the actual constraints limiting the way in which the task may be accomplished, which is the same issue previously identified as (17). Issue 8 (problem if head moved to look at arm while gesture system operational: may be interpreted as a command) was identified in stage four of EMU, when the system profile was compared with the user modalities. When writing out the interaction sequence for stage five, the long and repetitive nature of the listing suggested that the structure of the interaction had too many steps, as previously noted by issue 1 (long sequence of operators to move arm). The length of time taken to stop the arm (issue 14) was also identified at this point, with the examination of the modality sequence. In stage six of EMU, semantic clashes relating to the terms **Move/MoveArm**, and **Arm** were noted, which correspond to issues 3 (difficulty of choosing between Move Arm or Move) and 6 (confusion over joint called Arm).

There were four issues identified by the analysis which had not been identified by the previous analyses. In stage three, where the environment profile was defined, attention was drawn to the level of lighting, as well as what lighting might be available from the display, and its possible effects on the interaction (issue 20). When the profiles of the user, system and environment were compared against the lists of modalities in stage four to check for obvious discrepancies, the comparison of user profile with system modalities suggested that it might be difficult for the user, due to disabilities of movement, to move their field of vision from the display to the arm and back again (issue 21). The comparison of the user and system modalities with the environment profile again highlighted issue 20, with the user maybe not having enough light to see the menus. There was one potential physical clash identified in stage six, which was frequently repeated, resulting from the user having to potentially look in one direction to observe the arm, and in another to observe the menu options (issue 22). Another clash of the type clash-unless-expert was found regarding the difficulty of the user in judging the arm movements (issue 23).

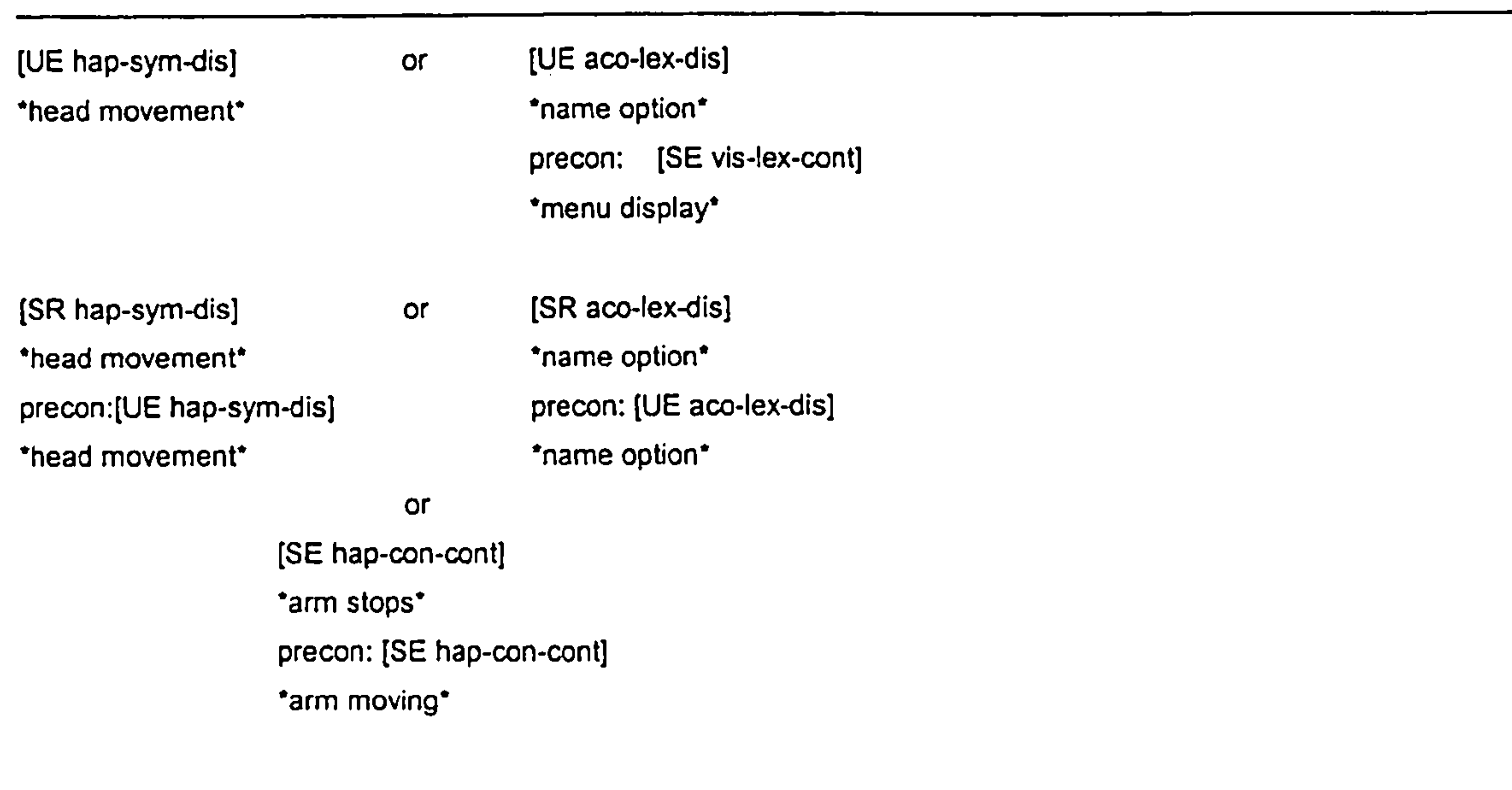


Therefore, different usability issues are discovered at different stages in EMU, as shown by the issues identified in this study. Stage one identified an issue relating to the mismatch between the domain and the device (issue 17). Stage three identified an issue relating to the effect of the environment (issue 20). This issue was also identified in stage four, as were other issues relating to comparing the user, system and environment (issues 8 and 21). In stage five issues relating to sequence and time were identified (issues 1 and 14). Various clashes were identified in stage 6 (issues 3, 6, 22 and 23). This relates back to the description of the different stages of the methodology (sections 5.6.1 through to 5.6.8), with what was expected to be found at each stage, and shows that this does indeed happen.

### 5.7.2 Application issues

The use of EMU in the analysis of the robotic arm interface allowed an opportunity for insight into the methodology's limitations in terms of coverage and application.

Although the analyst was able to model the effects of the user and the system on each other, there was no way of modelling the effect of autonomous system changes such as the arm stopping automatically when it reached the limit of movement. It was unclear using the notation in the interaction listing sequence, as shown in figure 5-10, that the arm could either stop due to user action, (the SR hap-sym-dis and the SR aco-lex-dis modalities) or due to the limit of motion being reached (the SE hap-con-cont modality).



**Figure 5-11: Section of listing for stop command**

The notation at present does not properly support the issue of transition, or when modalities actually start and stop. In the section of the interaction sequence listing detailing the stopping of the arm, it was necessary to include a modality for the arm stopping, or the arm's absence of movement. The need to mark the start and end points of modalities had not previously been considered.

When applying EMU it became apparent that the methodology would benefit from some form of automated support. The hand-checking of the interaction sequence suggests that tool support would be helpful in finding clashes, and in checking that all sections of the methodology had been completed. For example, when writing the modality listing of the modalities used in the interaction, it was noticed that there were no SR modalities written. These had not been included in the interaction sequence listing by accident. One positive aspect of this is that it demonstrates that the way the methodology is structured lends itself to internal checking and helps to avoid omissions made due to outside pressures. This omission was corrected, and the full sequence finished. However, the automation of this would have brought it to the analyst's attention earlier.

The use of automated support is not without problems. When writing the interaction sequence listing it was tempting, because of the similarity of the interaction stages, to make use of the cut-and-paste facility to speed up the writing of the notation, due to the time-consuming nature of writing. However, use of cut-and-paste may have a detrimental effect on the quality of the observation of each stage, since issues local to a particular stage might be missed.

### **5.7.3 Discussion**

That the EMU analysis was able to successfully identify six previously identified usability issues and four new issues demonstrates the ability of EMU to identify usability issues.

The issues identified can be divided into four groups. By focusing on the task and establishing why it might be important, the mismatch between the way that the arm works and how the user might use the arm (issue 17) could be established. The long sequence of operators needed to use the arm (issue 1) and the time taken to interact with the system to stop the arm (issue 14) were identified through the detailed examination of the modalities in the interaction sequence listing. Issues 3, 6, 22 and 23 concerned clashes. Issues 8, 20 and 21 were identified through the comparison of user, system and



environmental profiles with the modality listings. EMU explicitly supports the identification of clashes, and the identification of issues arising from incompatibilities of profiles and listings.

It must be noted that the insights gained were derived quite strongly from the characteristics of this particular example, and that other examples would yield different, but not inconsistent, insights.

However, the use of EMU demonstrated that there is further work to be done regarding the adequate representation of non-user effects on the system, and the representation of transitions. By applying EMU it was apparent that tool support would be useful in providing internal consistency checking, such as checking that all modality types had been included, checking for particular clashes, and in by-passing the time-consuming and repetitive action of writing the interaction sequence. However, tool-support, although facilitating the application of the methodology, might be detrimental, in that issues relevant to a particular section of the interaction might not be noticed, and invalid assumptions made.

## **5.8 Summary**

This chapter has specifically addressed the issue of the lack of appropriate approaches for applying theoretical understanding of multi-modality. This has been achieved through the development of an eight-stage methodology.

The results of the research examined and presented in the previous chapters were used to identify those matters which the methodology should address, and the definition of modality was applied to the robotic arm interface examined in chapter two to provide a starting point for development. The robotic arm study identified the need for well-defined procedures and a more meaningful notation.

Various representation formats were tried, including tabular representations such as User Action Notation, and diagrammatic representations such as CPM and Petri Nets. It was found that these could not represent the complexity of the modalities and their properties. A notational representation was developed which could show the temporal flow of the interaction, as a whole and in smaller sections, as well as make explicit the preconditions applying to the modalities and the choice between them. It was also able to handle the complex representation of large interfaces. The final representation can

clearly show modalities in terms of individual modalities, preconditions, choice, as well as at different levels of granularity.

The structure of the methodology was developed by iterating around the example of the telephone, and resulted in a methodology of eight stages to support the analyst in identifying usability issues, both of a general nature and those specifically relating to multi-modal usability. The approaches examined in chapter two, namely STN, CW, GOMS, PUM and Z, were used extensively both as sources of inspiration and as examples of what not to do. The methodology stages encourage the analyst to iterate round the interaction, and specifically encourage the investigation of multi-modal usability issues. The methodology overall can handle issues of task, profiles, and clashes, as well as the listing of the interaction sequence.

The EMU analysis was able to identify several existing usability issues as well as new issues relating to a known interface, supporting the use of EMU as a usability evaluation method. Various application problems were identified, including the adequate representation of non-user effects on the system, and the representation of transitions. Tool support was suggested as one way of supporting the analyst in avoiding errors.



# **6. APPLICATION OF DEFINITION AND**

## **METHODOLOGY**

### **6.1 Introduction**

In the previous chapters a new definition of modality, with an associated taxonomy and methodology, was proposed. The goal of this chapter is to establish the usability and usefulness of this definition and methodology by means of a case study using computer science students to assess the usability issues associated with two tasks performed on two multi-modal interfaces.

This study grounds the claims that modalities can be identified using the definition of modality and taxonomy, and an understanding gained of usability issues by using the methodology. The usefulness and usability of the definition and taxonomy can be examined by assessing how well subjects were able to identify modalities. The extent to which the definition and methodology allow insight into the interaction is assessed by examining how subjects modelled the interaction. The opportunities for insight into usability issues provided by the definition and methodology, and the coverage of the methodology, are determined by assessing the subjects' identification of usability issues.

The transfer of a technique from theory to practice is hard. As the developers of other methods such as GOMS, CW and PUM have found, there are issues involving application that have to be resolved if a technique is to have value in practice (Blandford et al, 1998a). It is only through the use of an approach by non-originators that the power of a technique can be established, and by careful analysis separate out the power of the technique from the craft skills of those using it. As Blandford et al (1998) note, "Novel approaches to designing or analysing systems only become useful when they are usable by practitioners in the field, and not just by their originators." Only through the large-scale application of an approach can its usefulness be examined, since, as noted by Blandford & Young (1996), small case-studies are not always amenable to scaling up.

This chapter first describes the structure of the case study, then goes on to present and discuss results relating to the identification of modalities, understanding of the interaction, and the usability issues identified. A final general discussion brings together the issues highlighted by this study.

## **6.2 Method**

The structure of the case study, based on subjects using the methodology to assess the usability of two multi-modal interactive interfaces, is described in detail below.

### **6.2.1 Domain**

Two interactive multi-modal interfaces were assessed by the subjects; a tube ticket machine, and a web-based heating control panel. The two devices were selected in order to provide a large number of potential usability issues and to give a wide range of modalities to be identified and analysed. The tube-ticket machine is a good example of a real-world device used by a large user group. The web-based heating control panel was chosen to provide the subjects with a computer-based device for analysis, since this is the kind of system traditionally analysed in HCI.

### **6.2.2 Subjects**

The subjects were a mixed gender group of twenty-eight third-year computer science undergraduates at Middlesex University, taking an advanced course in HCI. All had some previous experience of HCI, having taken one module in the subject previously.

### **6.2.3 Procedure**

The subjects were set two tasks, based on the two different device interfaces being analysed, both of which were to be assessed using the methodology and written up into a report. The tasks were as follows:

- buy an adult return ticket to St Pancras at 10 am on a week-day from Bounds Green tube station, using a ten pound note. You may assume that there is sufficient change in the machine.



- using a web-based simulation of a heating control panel set the heating controls for the whole week (Monday through to Sunday), with the heat coming on at 7am each day, and going off at 10pm each day.

Both of the tasks chosen for analysis were tasks that it would be reasonable to expect the interfaces to support, but were deliberately made as difficult as possible. This was to ensure that the subjects had enough scope to assess the usability of the devices, and to show that even well-developed systems, such as the tube-ticket machine, can have usability problems. Indeed, the tube-ticket machine task was actually impossible to do in the form asked for, since adult return tickets are not available after 10am on a weekday, and a passenger has to buy a four-zone railcard instead. Setting difficult tasks also meant that the use of the modality theory and methodology was less straightforward, and would thus give more interesting results in terms of their usability. For pictures of the web site heating control panel please refer to Appendix D. A summary of the task steps for the tube ticket machine is included in Appendix E.

For the purpose of doing the reports, the subjects were randomly paired with other subjects within the group. Their instructions were that they should work together on the analysis, but submit separate written reports. This was so that they could support each other with the mechanics of the methodology, but that any final insights would remain their own.

The tasks were set in the form of written reports which were used to see how well the subjects had used the methodology to analyse the usability of the two devices in relation to the set tasks. The subjects were asked to write a report for each task, detailing each stage of the methodology and summing up the final usability considerations. These reports were assessed in terms of how correctly they had been written in the component stages, and the usability issues identified.

The subjects were provided with an information pack containing details of the modality theory used and explaining the methodology (see Appendix F) and given a one and a half hour lecture on the methodology and modality theory. They then had a one hour laboratory session where they attempted to identify modalities on a computer package, followed by a one and a half hour seminar where questions were answered and further simple examples, for instance the modalities used to interact with an ATM, worked through. There were two

more laboratory sessions over the next two weeks while the subjects were analysing the tasks, and two seminars run on a drop-in basis to answer questions. The reports were submitted ten days after the last seminar, and were marked as coursework, the result counting towards the subjects' final module mark.

The results were assessed in three ways: a questionnaire was distributed after the initial lecture; the written usability reports were examined; and a final questionnaire conducted after the reports had been collected.

The first questionnaire gathered preliminary data about the subjects, in terms of their background, experience, and how they felt about using the methodology to do the set tasks. This questionnaire was distributed to those subjects who had attended the lecture, during the first laboratory session, and the responses collected at the end of the session. Only those who had attended the lecture were given a questionnaire, and it was emphasised that filling it in was not obligatory and would have no bearing on their final report mark. Twenty-five subjects had attended the lecture, and they all filled in the questionnaire. The second questionnaire gained feedback from the subjects about the methodology having had a chance to use it. This questionnaire was distributed during a laboratory session ten days after the reports had been handed in, and collected at the end of the session. Again, it was emphasised that filling it in was not obligatory, and would have no bearing on their final mark. Nineteen questionnaires were completed and returned. The questions required making a mark on a Likert scale. A 100mm line was used in order to facilitate analysis of the responses given, with the rating recorded as a value between 0 and 100, measured to the nearest mm. A full listing of the questions asked in both questionnaires can be found in Appendix G.

The purpose of the written report was to see how well the subjects had used the methodology to analyse the usability of the two devices in relation to the set tasks. Twenty-eight of the subjects handed in reports, which were numbered from 1 to 28, and are referred to in the results by these numbers. These reports were assessed in terms of how correctly they had been written in the component stages, and the usability issues raised. Because of the nature of the analyses, which allow for various interpretations of the interaction, it was not possible to provide a model answer for each task.



### **6.3 Identification of Modalities**

The most basic test to see whether the definition of modality and the associated taxonomy and theory can be applied is to see whether the subjects can identify the modalities used in the interaction, and the subjects' perceptions of using the definition and taxonomy. This is examined here in four ways: through an in-depth assessment of one part of the interaction sequence for the tube ticket machine task; by examining the listings stages of the reports for both the tube ticket machine task and the heating control panel task; by examining the sequence stages of the reports for both tasks; and by examining subject responses from the two questionnaires.

Not only should the subjects be able to identify the modalities in natural language general terms, for instance "display changed", "button pressed", but also in terms of their component parts, whether they are system or user modalities, and whether the modalities are expressive or receptive. This would demonstrate both that the subjects can use the definition and taxonomy, and also that they are aware of what was happening in the interaction and can describe it accurately and clearly in terms of both the system and user, and in terms of who or what was expressing or receiving information at any time.

The questionnaire data was used to assess subject perceptions of using the definition and taxonomy before and after the analyses. If the subjects were confident about using the definitions, it would imply that the definitions were understandable at a theoretical level, and hence potentially usable. This is closely related to the clarity of the initial lecture where the concepts were first explained. An unclear lecture would cloud the concepts and reduce student levels of confidence as to their use. Final confidence levels were assessed by means of the second questionnaire. If the students were confident that they had correctly identified the modalities once the reports were completed this would imply that they had been able to use the definitions, and hence give an indication as to the usability of the definition and taxonomy.

## 6.3.1 Results

### 6.3.1.1 *In-depth Examination*

A small portion of the written interaction listings detailing interaction with the tube ticket machine is examined here in depth. By examining a small “snippet” it is possible to see in microcosm the issues surrounding the use of the definition and taxonomy.

The portion of interaction chosen is the point before the user inserts the ten pound note into the tube-ticket machine. This particular section was chosen because it is multi-modal in a very obvious manner as various things are happening: the system is displaying a message detailing the cost of the travelcard, and the machine is using lights and noise to prompt the user to insert the money in the correct location.

The reports were examined to see whether the subjects could correctly identify the modalities used, in terms of whether the modalities were expressive or receptive, and system or user. This would show that they were able to distinguish the agents in the interaction, and the flow of information in the interaction. The correct use of the components of the modality was also looked at in order to see how accurately subjects were able to describe the modalities and distinguish between similar modalities.

#### 6.3.1.1.1 Modalities used

Twenty-two of the subject reports detailed this stage of the interaction. Of those not examined, reports {17,24} were too badly written for this stage to be identified, in that the sequence did not allow for identification of this stage of the interaction. Reports {3,5,12,18} had terminated the task before this point, since the machine had stated that there was no ticket available.

At this stage in the interaction various modalities could be identified, such as the system display changing, the green light by the coin slot showing and the note slot making a ticking noise. It depended upon the level of granularity taken by the analyst as to whether they were included in this section of the interaction listing, hence the wide variation in the number of modalities included. As a minimum, however, the system display prompting the user for the



	Display	Green light/ Ticking Noise	Other
1	[SE vis-lex-cont] *Display information '1 day Travelcard - zone 1234 - Please pay £3.80' [UR vis-lex-cont] *Read system display*	[SE vis-sym-cont] *Displays green light* [UR vis-sym-cont] *See light flashing* [SE aco-sym-cont] *Make ticking noise* [UR aco-con-cont] *Hear ticket machine ticking*	
2	SE vis-lex-cont * system displays message* UR vis-lex-dis * user reads message *	SE vis-sym-cont *has green light* UR vis-sym-dis *see machine, light*	SE vis-con-cont * money slot * SE vis-sym-cont * has icons (arrows) * UR vis-sym-dis *see machine, icons* UR vis-lex-dis * user sees screen *
4	[SE vis-lex-cont] *System displays "-Adult Day Travelcard Zones 1234/Please Pay: £3.80-" [UR vis-lex-cont] *View system display*	[SE vis-sym-cont] *green light comes on* [UR vis-sym-cont] *See green light on note feeder* [SE aco-sym-cont] *ticking sound occurs* [UR aco-sym-cont] *Hear ticking sound*	[SE vis-sym-cont] *coin slot opens* [UR vis-sym-dis] *Notice coin slot opens*
6	[SE vis-sym-cont] *display amount*	[SE vis-sym-cont] *flashing at the money slot* [UR vis-sym-cont] *sees flashing* [SE aco-sym-cont] *give sounds* [UR aco-sym-cont] *hear sound*	
7	[SE vis-sym-cont] *system displays amount £3.80*	[SE vis-sym-cont] *Machine flashing* [UR vis-sym-cont] *User sees light flashing* [SE aco-sym-cont] *Machine makes sound* [UR aco-sym-cont] *User hears machine makes sound*	
8	[SE vis-lex-cont] *displays 'please pay: £3.80'* [UR vis-lex-cont] *sees 'please pay: £3.80'*		[UR vis-sym-cont] *see note slot*
9	[SE vis-lex-cont] *Display message on Screen* [UR vis-lex-cont] *reads message on Screen*		[UR hap-con-cont] *waits for response of machine* [UR vis-sym-cont] *looks at display*
10	[SE vis-lex-dyn] *display shows price [UR vis-lex-dyn] *user reads display		
11	[SE vis-lex-cont] *shows text* [UR vis-lex-cont] *reads text*		
13	[SE vis-con-lex] Adult one day travel card zones 1234 please pay £3.80		
14	[SE vis-lex-cont] *System displays "-Adult Day Travelcard Zones 1234/Please Pay: £3.80-" [UR vis-lex-cont] *View system display*	[SE vis-sym-cont] *green light comes on* [UR vis-sym-cont] *See green light on note feeder* [SE aco-sym-cont] *ticking sound occurs* [UR aco-sym-cont] *Hear ticking sound*	[SE vis-sym-cont] *System response, coin slot opens* [UR vis-sym-dis] *See coin slot open*
15	[SE vis-lex-cont] *system displays amount to pay and indicates to insert by green light*	(SE)	
16	[SE vis-lex-cont] *Display travelcard price.* [UR vis-lex-cont] *Read displayed price *		
19	[SE vis-lex-cont] *Price Display*		
20	[SE vis-lex-cont] *Machine displayed insert money and price of the ticket and type* [UR vis-lex-cont] *Sees displayed insert money and price of the ticket and type*		
21	[SE vis-sym-cont] *Display £3.80* [UR vis-sym-cont] *Display £3.80*		
22	[SE vis-lex-cont] *system displays "Adult One Day Travel Card Zone 1-4 £3.80" [UR vis-lex-cont] *user reads screen display*		
23	[SE vis-lex-cont] *screen displays pay 3.80* [UR vis-lex-cont] *user sees information on screen]		
25	[SE VIS-CON-LEX] *insert £3.80p.*	[SR vis-dyn-sym] *Notes slot light flashes*	[SR aco-hap-con] *Coin slot opens.* [SR vis-con-sym] *LED sign flashes "Change Given"*
26	[SE vis-lex-cont] *system display shows fee* [UR vis-lex-cont] *system display shows fee*	[SE vis-sym-cont] *system shows green light* [UR vis-sym-cont] *system shows green light*	[SE aco-con-dis] *sound of systems coin slot opening* [SE vis-sym-cont] *system shows arrows and money icon* [SE vis-con-cont] *system shows money slot* [UR aco-con-dis] *sound of system slot opening* [UR vis-sym-cont] *system shows arrows and money icon* [UR vis-con-cont] *system shows money slot*
27	[SE vis-lex-dis] *Screen displayed a message of 'Please pay amount of £3.80'		
28	[SE vis-lex-con] *Screen displays the amount to be paid*		

**Table 6-1: Modalities identified by subjects in in-depth portion of interaction**







	SENSORY FORM			INFORMATION FORM			TEMPORAL FORM		
	VIS	ACO	HAP	LEX	SYM	CON	DIS	CONT	DYN
1									
2				■					
4			■						
6									
7									
8									
9									
10									■
11									
13									
14			■						
15				■					
16									
19									
20				■					
21									
22									
23									
25		■	■						
26									
27				■					
28									

Key

Wrongly written modality ■

Visual component VIS

Acoustic component ACO

Haptic component HAP

Lexical component LEX

Symbolic component SYM

Concrete component CON

Discrete component DIS

Continuous component CONT

Dynamic component DYN

**Table 6-3: Errors made with component parts of modalities in section of interaction examined**

### 6.3.1.1.3 Modality Components

The actual components of the modalities, in terms of the sense, information form and temporal nature, as is clear from table 6-3, were mostly written correctly, with only eight reports making mistakes, and seven of these being minor, in that the mistake was a single error, typically confusing one modality component with another. Three reports made mistakes regarding the sensory component: reports {4,14} writing a haptic modality as a visual modality, and report {25} apparently trying to merge two different modalities into one representation, resulting in a modality written as “aco-hap-con”, when it should have been split into one acoustic and one haptic modality. Four reports {2,15,20,27} had errors of information form, writing “lex” instead of “sym” or “con”. One report {10} made a mistake regarding the temporal form of a modality, writing the system displaying the price as a dynamic rather than continuous modality.

### **6.3.1.2 Modality Listing Results**

In order to determine the use of the definition and modality on a wider scale, the modality listing stages of the reports for both the tube ticket machine and the web site heating control panel tasks were examined. This stage of the methodology requires the listing of all the user and system modalities, in order that any obvious discrepancies might be noted. The results for the modality components are first examined, followed by the results for the use of types.

Five reports {1,2,11,12,26} made such minor mistakes regarding modality form or types in either the description of the tube ticket machine modalities or the web site heating controller modalities that they were considered to be error free. Minor errors were where a significant error, such as the using the wrong type or modality, was made only once, indicating a momentary misunderstanding or lapse of concentration. Major errors were where persistent mistakes were made in the reports, indicating either a lack of understanding of the concepts or major lapse in concentration.

#### **6.3.1.2.1 Modality Types**

Nineteen reports made mistakes in the use of types, as shown in table 6-4, either confusing the expressive and receptive, or system and user, aspects of the modalities, or not including them where appropriate. Seven of these reports made minor slips, typically confusing an expressive and a receptive type. For example, {15} in the tube ticket machine listing wrote “looking at machine” as “UE vis-con-cont” instead of “UR vis-con-cont”. The other twelve reports made more serious errors, with repeated confusions and omissions of type. For example, in report {16}, no button pushes (SR modalities) were recorded as being received by either the tube ticket machine or the web site heating controller. In report {25}, expressive and receptive modalities were repeatedly confused, as in the tube ticket machine listing where the modality “user selects ticket type” was wrongly written as “UR hap-dis-lex”. Thus identification of the type of modality caused problems for almost half the subjects, both in terms of recognising that the modality needed to be included, and in assigning it the correct type.



	MODALITY COMPONENTS			
	TYPES	SENSE	INFO	TIME
1				
2				
3				
4				
5				
6				
7				
8				
9				
10				
11				
12				
13				
14				
15				
16				
17				
18				
19				
20				
21				
22				
23				
24				
25				
26				
27				
28				

Key

No errors made



Minor errors made



Major errors made



**Table 6-4: Mistakes made in the use of types and modality components in the modality listing stages of the reports**

### 6.3.1.2.2 Modality Components

Out of the eleven reports which made mistakes regarding the sense part of the modality, nine of these were one-off mistakes, typically confusing the “vis” component with “hap” (see table 6-4). For example, {6} wrote the modality “Looked for St.Pancras/Kings Cross” as “UE hap-sym-cont” when it should have been a visual modality of type UR. Of the two remaining reports, {17} made a repeated mistake of selecting a button as “SR hap-vis-lex” rather than “SR hap-dis-lex” which suggests an error in notation (writing “vis” instead of “dis”) rather than in understanding. Report {24} had major errors with regards to the sense of the modalities listed, writing more than one sensory component, and often using the wrong sensory component. For example, in the tube ticket machine listing, the modality



“press the buttons” is written as “UE aco-vis-hap”. This is wrong, since all three varieties of sense are written instead of just one, and no information form or temporal nature components are given. This kind of confusion occurs repeatedly throughout the two listings and shows that the subject was very unsure as to the correct identification and use of the sensory component of the modalities. This was the exception since most subjects had either no difficulty or only minor problems in establishing the correct sense used by particular modalities.

Fifteen reports made errors with the information form of the modalities. Seven of these were minor slips made only once, mostly writing “lex” instead of “sym” or “con” when describing the action of pressing buttons. For example, {10} wrote the modality “press St.Pancras” as “UE hap-lex-dis”. “Lex” should not have been used, and the modality should have been written as “UE hap-sym-dis”, since it is a discrete instance of a haptic action, which conveys symbolic information. Eight reports had problems of a more serious nature. Reports {3,5,17,18,27} confused the use of “lex” with “sym” and “con” more than once, writing pressing a button as a lexical rather than a symbolic or concrete modality, possibly since they were confused because the buttons had writing on them. In reports {7,21} modalities for messages were written as “sym” when they should have been written as “lex”. For example, in report {7} the tube ticket machine listing modality “display select destination” is written as “SE vis-sym-cont”, and this kind of mistake is repeated throughout the report. Report {24} often omitted the information form of the modality, as in the tube ticket machine listing modality “take the ticket and the changes” which was written as “UR vis-aco-hap”, thereby indicating great confusion as to the correct use of that aspect of modality.

Twenty-two subjects wrote the correct temporal form for the modalities listed. Of the six reports which made errors, five reports made minor one-off slips, typically writing “cont” instead of “dis” or “dyn”, or vice versa. For example, in report {9}, the web site heating controller modality “continually click plus button to Monday” was written as “UE hap-sym-dyn” when it should more appropriately have been written as “UE hap-sym-cont”. Again, report {24} had major problems with this component of modality, apparently confusing the



use of the concrete and continuous, as well as on occasion missing out the temporal component completely.

### ***6.3.1.3 Modality Interaction Sequence Results***

The modality interaction listing stages of the reports for both the tube ticket machine and the web site heating control panel tasks were examined here to assess use of modalities in terms of both components and types. This section of the methodology requires the writing out in full of the interaction, with the appropriate modalities used in sequence.

#### **6.3.1.3.1 Modality Types**

Seven reports made no omissions or confusions of the types of the modalities in the interaction listing section of the reports. However, as table 6-5 demonstrates, many subjects had difficulty using types at this stage.

Sixteen reports made mistakes relating to the use of UR types. Of these, three reports made minor errors, writing the modality of the user reading instructions as type UE rather than UR. The thirteen remaining reports were confused as to the role of UR modalities, and either missed out several UR modalities, or confused them with other types, or both. For example, in {21}, the tube ticket machine modality \*user looking at device\* was written as [UE vis-sym-cont] instead of [UR vis-sym-cont].

Twelve reports made errors with regards to the use of UE types. Five reports made an occasional, minor error in the use of UE types, typically confusing UE and UR to describe modalities where the user was looking at the system. Four reports {5,13,18,19} repeatedly confused UE and UR types in the interaction sequence. Three reports {17,25,28} were very poorly written with many different errors. Report {28} had no UE modalities used in the tube ticket machine or web site heating controller interaction sequences at all. Report {25} omitted many UE types, and confused others with type UR. Report {17} had an interaction sequence for the tube ticket machine of only three stages, thus omitting various modalities. There were similar problems with the web site heating controller interaction sequence.

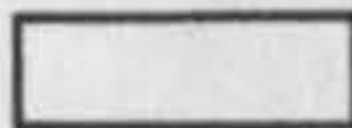
Subjects clearly had great difficulty with the use of SR types, with twenty reports making errors. Two reports {21,23} made minor errors, {23} missing out an SR type apparently by





accident since everything else was fine, and {21} using a wrong SR type once. Ten reports {5,7,9,11,12,13,22,24,27,28} had no SR types listed at all for one or both interaction listings. Eight reports {5,10,13,14,18,19,24,25} confused SR types with other types, typically SE types, and four reports {2,7,11,15} omitted SR types. Report {17} was very poorly written with many errors regarding this type.

	UR	UE	SR	SE
1				
2				
3				
4				
5				
6				
7				
8				
9				
10				
11				
12				
13				
14				
15				
16				
17				
18				
19				
20				
21				
22				
23				
24				
25				
26				
27				
28				

Key

No errors made 

Minor errors made 

Major errors made 

**Table 6-5: Table showing use of types in modality interaction listing stage of the reports**

Twelve reports made errors regarding the use of SE types. All of these were major errors, typically omitting many SE types, and also confusing them with SR types. For example, {12} confused SE and UR types in the tube ticket machine interaction sequence, and omitted several other SE types from both interaction sequences.

### 6.3.1.3.2 Modality Components

Six reports out of the twenty-eight examined wrongly wrote a modality in the interaction sequence, for example, writing a temporal aspect as discrete when it should have been



written as dynamic. The main mistakes were slips regarding the use of “lex” as opposed to “sym” or “con” {10,17,18}, and the correct temporal form {7,16}. One report {24} had a serious problem with modalities, with the modalities written with four or two components rather than three. Table 6-6 summarises these results.

	SENSE	FORM	TIME
1			
2			
3			
4			
5			
6			
7			Minor errors made
8			
9			
10		Minor errors made	
11			
12			
13			
14			
15			
16			Minor errors made
17		Minor errors made	
18		Minor errors made	
19			
20			
21			
22			
23			
24	Major errors made	Major errors made	Major errors made
25			
26			
27			
28			

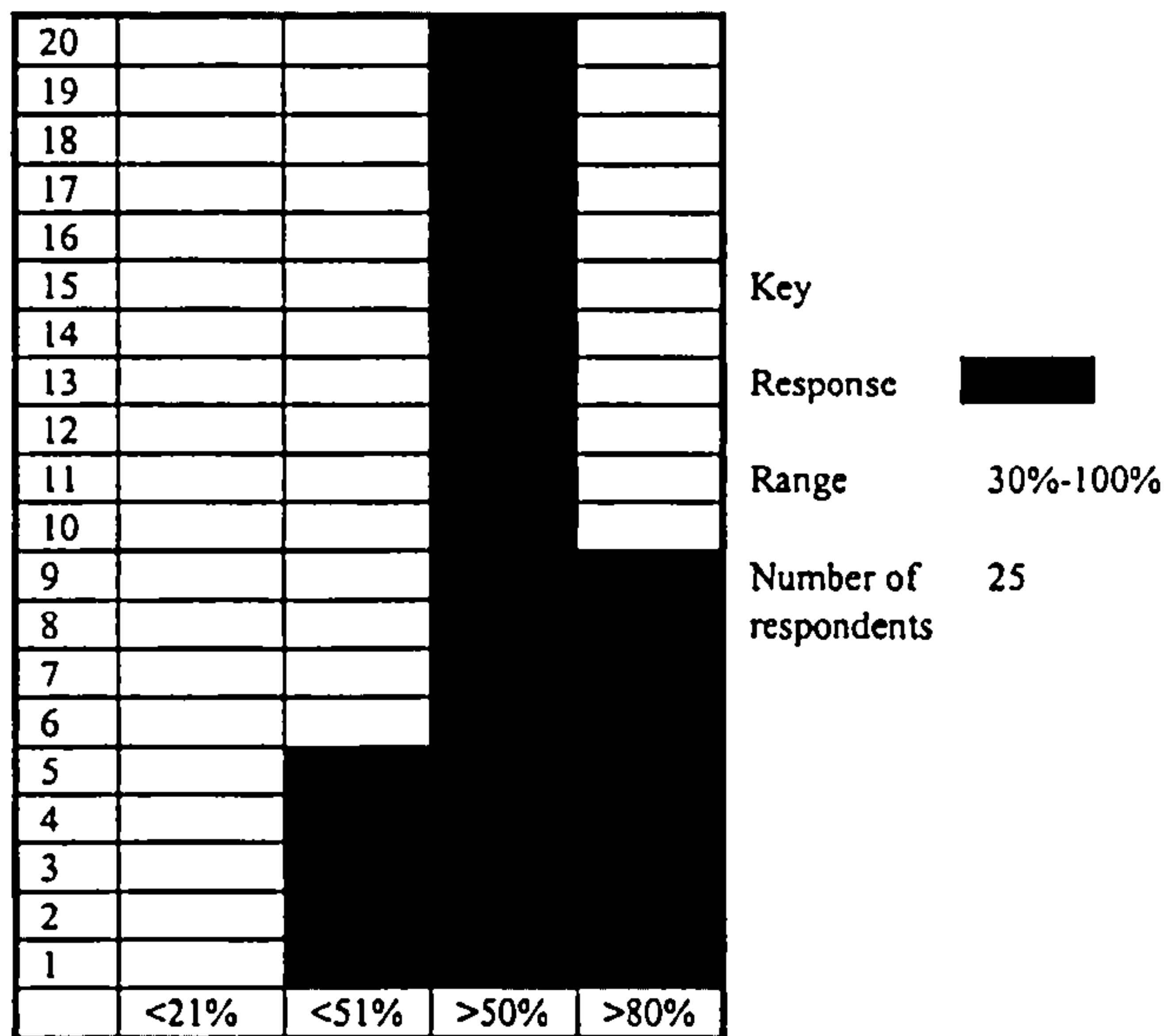
**Table 6-6: Use of modality components in interaction sequence listings stages of reports**

#### 6.3.1.4 Questionnaire Results

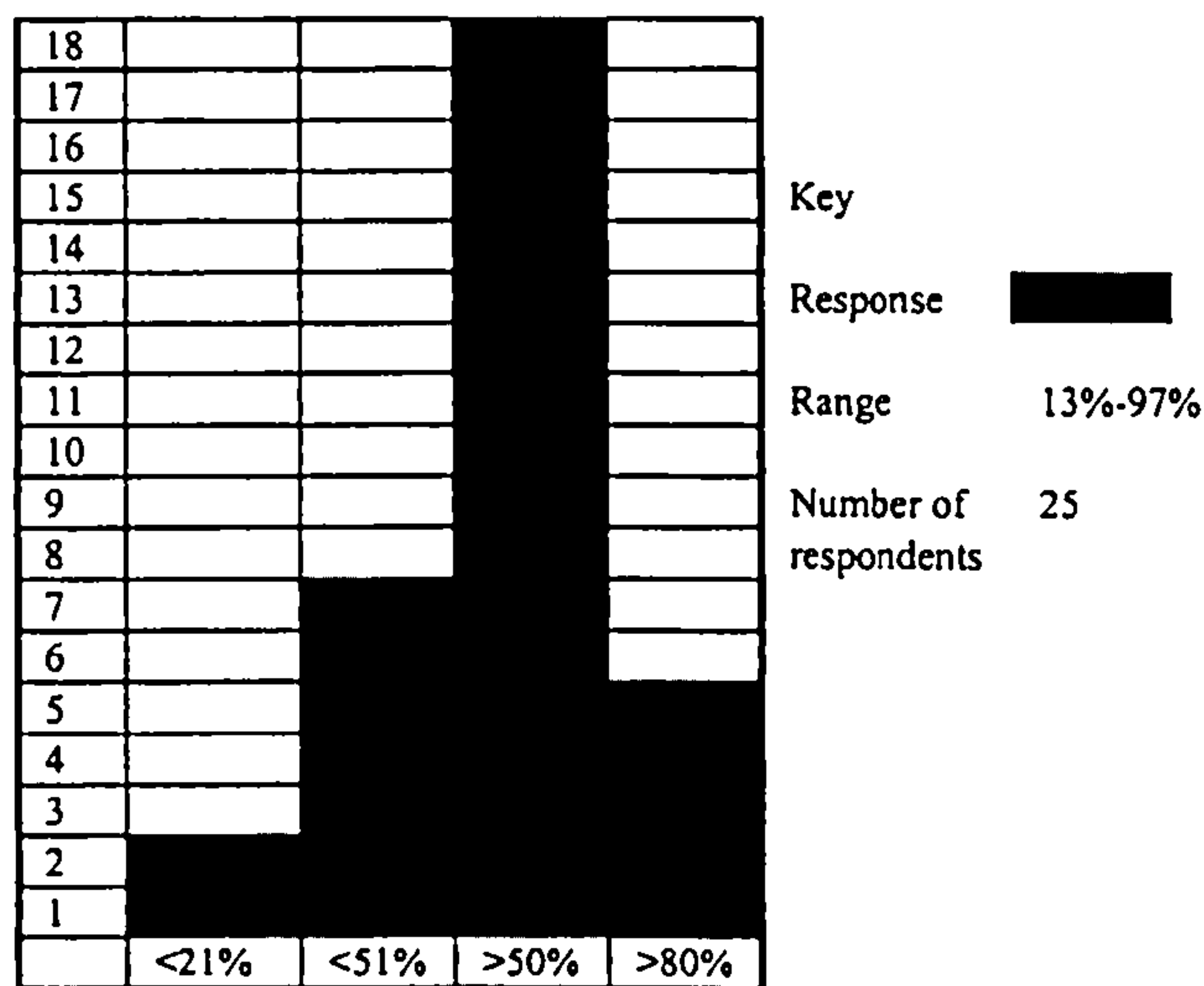
In order to examine subjects’ perceptions of the definition and taxonomy before and after the analyses, some of the responses to the two questionnaires are examined here. The first questionnaire was completed by the subjects after the initial introductory lecture on the definition and associated methodology, and before the opportunity to do any practical work using it. The second questionnaire was completed by nineteen of the twenty-eight subjects after the reports had been submitted. The results are presented in tables which show the number of responses above and below 50%, with additional columns giving responses also



20% or below, and above 80%, in order to provide a further breakdown of the results. Thus a response of 87% would be given in two columns; the above 50% column and the above 80% column.



**Table 6-7: Responses to Question 1: How clear did you find the lecture?**

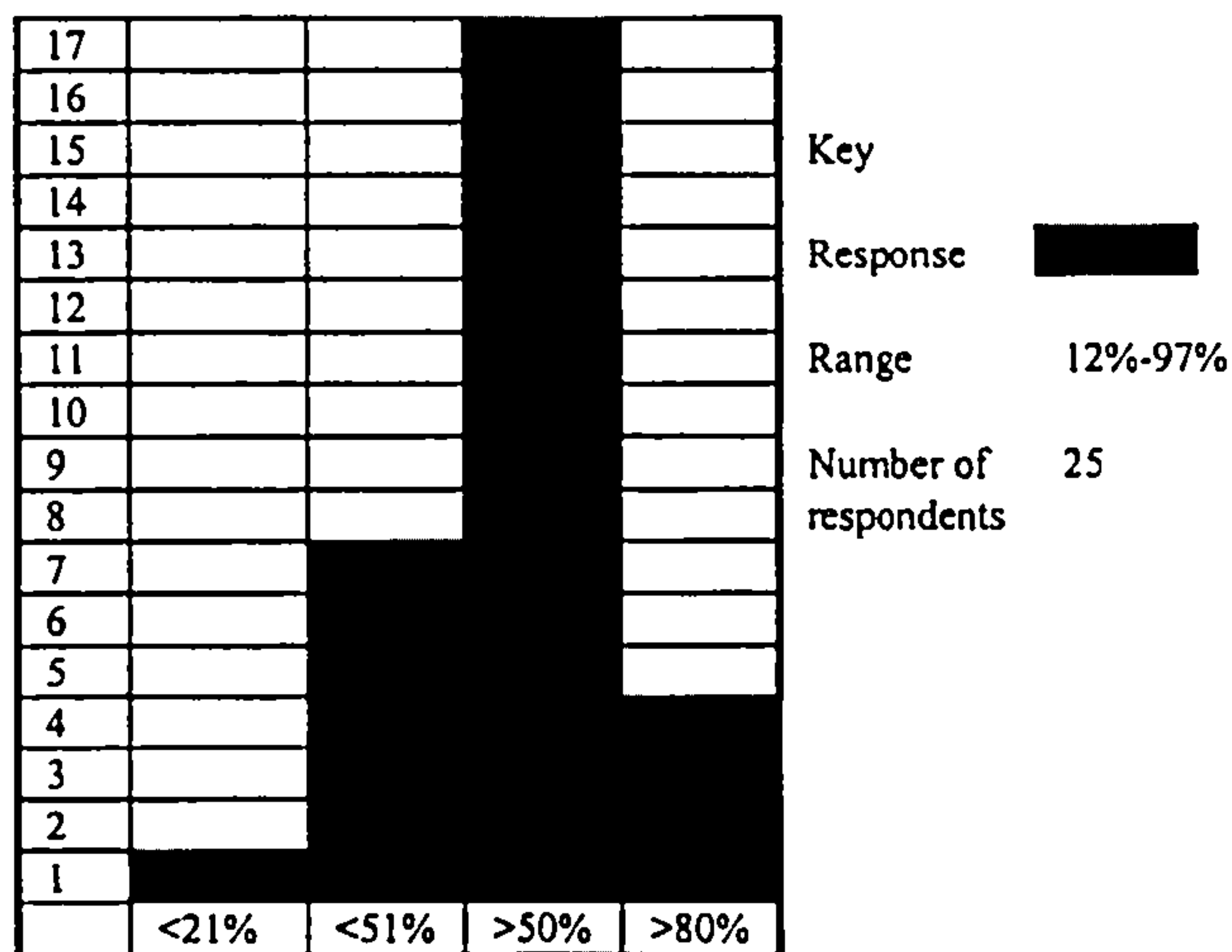


**Table 6-8: Responses to Question 2: How confident are you that you would be able to identify the modalities used in an interaction?**



#### 6.3.1.4.1 First Questionnaire Results

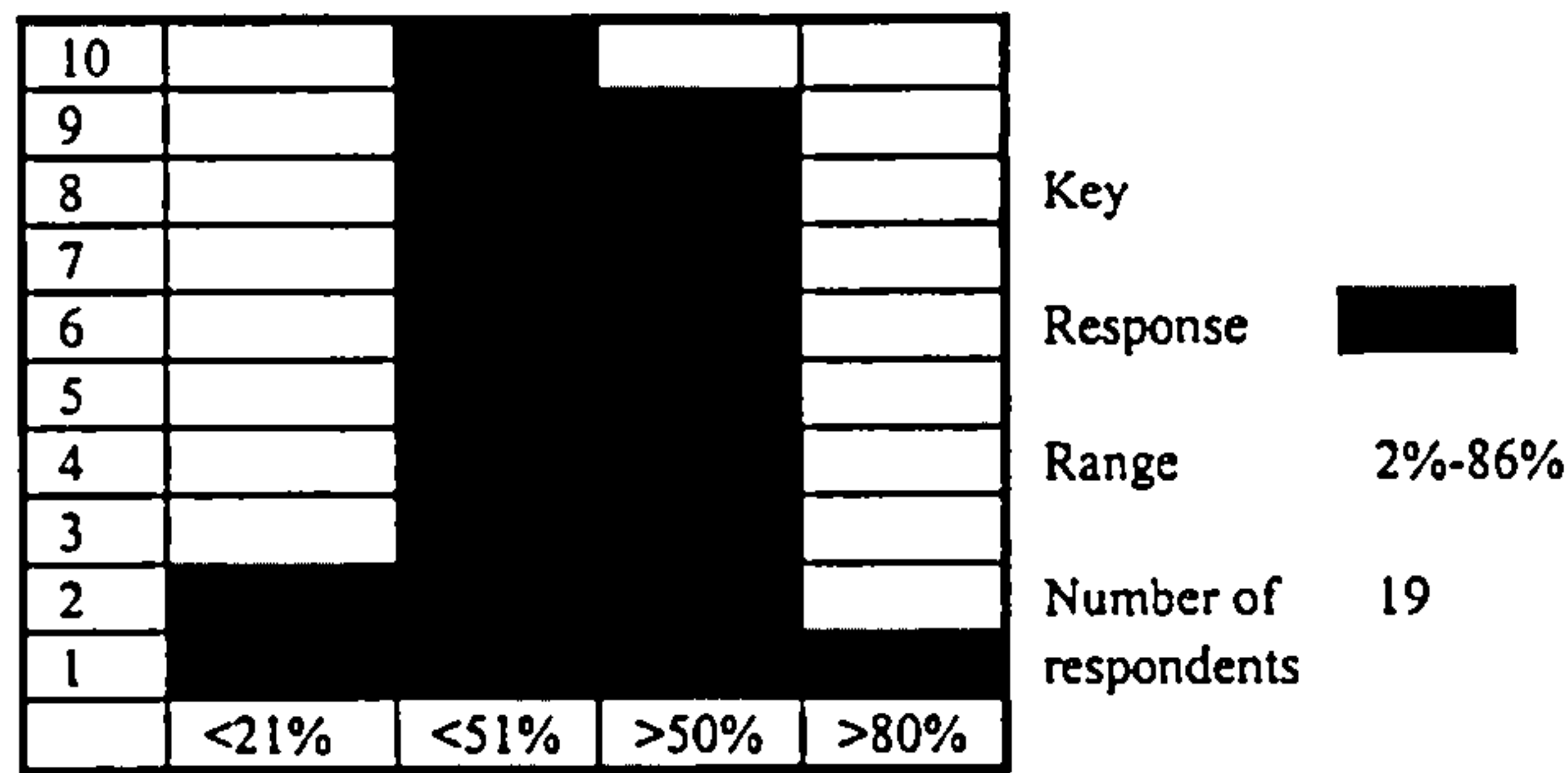
The question responses presented here relate to the clarity of the lecture, initial confidence in identifying modalities, and initial confidence in identifying types. As tables 6-7, 6-8 and 6-9 show, most subjects found the lecture detailing the modality theory and methodology to be very clear. Over two-thirds of the subjects (eighteen respondents) had a reasonable level of confidence in identifying modalities, with five appearing very confident. Seventeen subjects were confident about identifying types, with four subjects indicating very high levels of confidence.



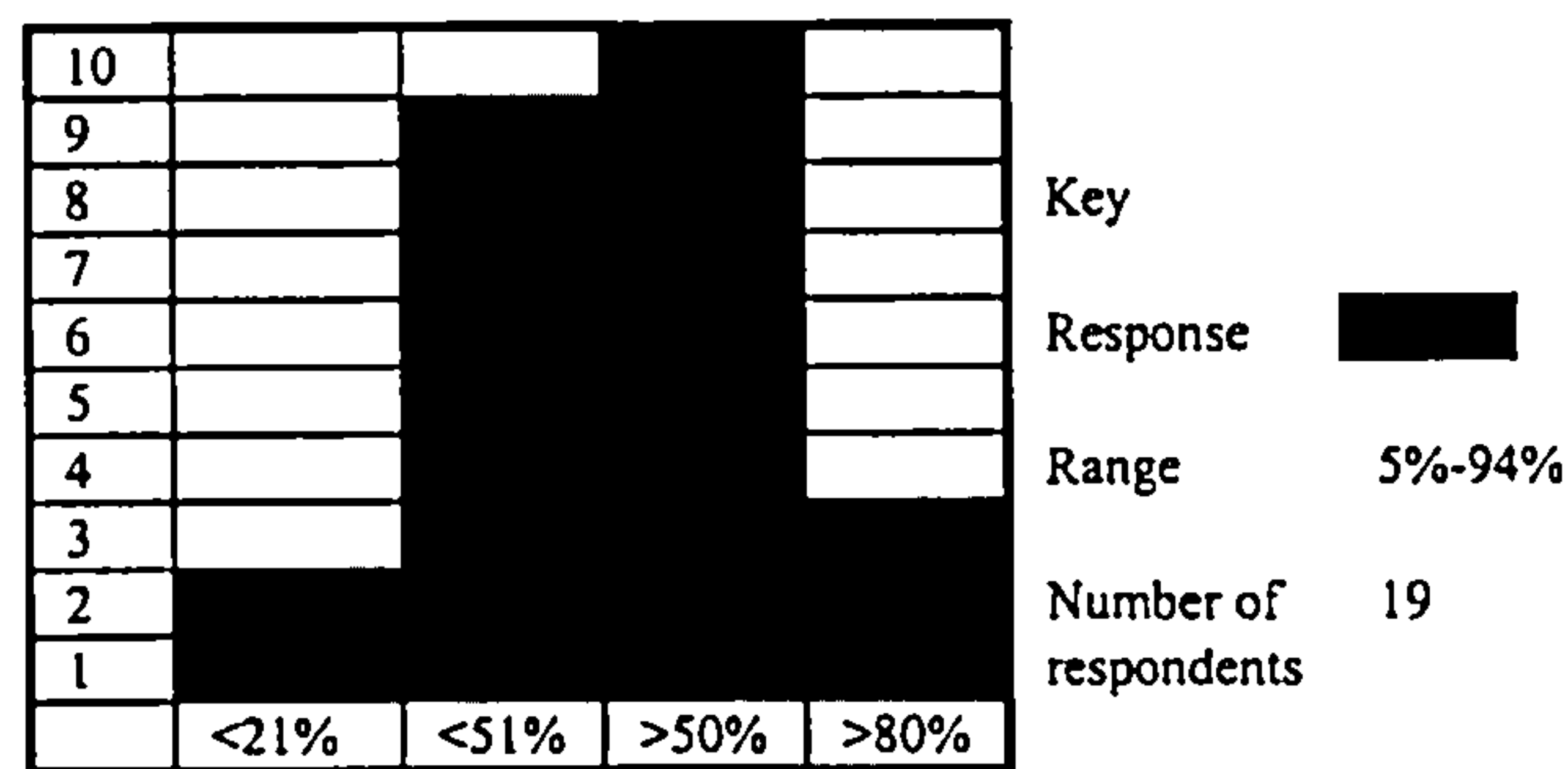
**Table 6-9: Responses to Question 4: How confident are you that you can identify modality types?**

#### 6.3.1.4.2 Second Questionnaire Results

The question responses presented in tables 6-10 and 6-11 relate to subjects' confidence in correctly identifying the modalities and types used in the interactions. The subjects were not confident that they had correctly identified the modalities used, with ten subjects giving low levels of confidence. They were slightly more confident regarding the correct use of types, with ten student giving responses of above 50%.



**Table 6-10: Responses to Question 2: How confident are you that you correctly identified the modalities used in the two interactions?**



**Table 6-11: Responses to Question 4: How confident are you that you correctly identified the modality types?**

### 6.3.2 Discussion

The results are discussed here in terms of the identification of modalities, component parts, and types. From this various points are made relating to the overall aims of assessing the definition and taxonomy and methodology in terms of usability and usefulness.

#### 6.3.2.1.1 Modality Identification

The correct identification of modalities is central to the definition and methodology examined in this case study. Therefore the subjects' initial confidence in identifying modalities was reassuring, given the new and complex ideas that been presented in the lecture. This indicates that it had been presented in an easily understandable way, and that the subjects found the ideas to be potentially usable, at least at a theoretical level before any practical application.



Subjects were able to apply the overall concept of modalities to successfully isolate an example of one or more modalities, as demonstrated by the in-depth examination of the portion of the tube ticket machine task, where all twenty-two subjects examined were able to both identify the critical modality for that section of the interaction, and correctly identify it as an SE modality.

#### 6.3.2.1.2 Modality Components

Modalities are identified using the definition given in terms of their component parts, and the successful identification of modalities in this way is dependent upon a complete understanding of their function in the interaction at any time. An inability to identify modalities using this definition in terms of these three parts would suggest that the definition is not usable and that the taxonomic categories are not easy and straightforward to use and apply.

The lack of errors overall in the portion of the tube ticket machine task examined in-depth, with seven reports making minor errors and only one report with substantial errors, demonstrates that the users were generally able to apply the definition and taxonomy to identify the component parts of all the modalities used at this point, and to correctly distinguish between the different alternative possibilities.

The use of sense and temporal form in the reports overall was generally correct, with subjects finding greater difficulty with the information form of the modalities. Most of the errors made regarding the sensory form were occasional slips in distinguishing visual and haptic modalities from the system's perspective. This is possibly due to the subjects natural bias towards the user's perspective, which resulted in them identifying modalities from the user's point of view even when it should have been done from the system's perspective.

With the information form, subjects had most difficulty with the modalities of button presses. They appeared to confuse the fact that the buttons had writing on them with the action of pressing the buttons, thereby writing that a lexical modality was happening when the modality was actually symbolic or concrete. It seems that they were assuming that the lexical information on the button would be conveyed, when actually it would be the symbolic representation of the button being pressed. However, many of these were single

errors, which were not repeated, suggesting a lack of concentration rather than a more serious lack of understanding.

Although subjects were initially confident about identifying modalities, in their responses to the second questionnaire subjects were less confident about their use, which suggests that they found using the definitions difficult in practice.

#### 6.3.2.1.3 Modality Types

Associated with the definition of modality is the idea of modality type. In order to correctly identify a modality its source, either user or system, must be identified. Whether or not it is being expressed or received is also critical to an understanding of the interaction. Problems with distinguishing between user and system modalities, and expressive and receptive modalities, would imply that these distinctions cannot be made easily.

Subjects were generally confident of being able to identify modality types after the lecture, which implies that the concept of modality types was usable at a theoretical level.

With the in-depth examination of the portion of the tube ticket machine task, all twenty-two of the subjects were able to identify it as an SE modality, and fourteen were able to identify the relevant UR modality, thereby showing that most subjects could distinguish between the two and were aware of the relationships between expressive and receptive modalities.

Most of the mistakes made with the use of types in this section of the tube ticket machine task were those of omission, missing out either receptive or expressive modalities, with only one report actually making a mistake with regards to the correct type of modality, writing them as receptive when they should have been expressive. The subjects were generally able to correctly identify whether a modality was expressive or receptive, and in most cases were able to provide the relevant associated modalities, for instance a user receptive modality to match a system expressive modality, showing that many of the subjects had a clear understanding of the flow of information in the interaction at this stage.

However, overall there were many problems for subjects with the use of types in the listings and interaction sequence stages of the reports. That so few reports made errors confusing system and user modality types shows that subjects were generally able to distinguish between the agents in the interaction. The main problems were in distinguishing between



expressive and receptive modalities, and in omitting receptive modalities. This indicates that separating out the expression of a modality and the reception of a modality caused problems, and that subjects did not always understand that the expression of a modality does not automatically guarantee its reception. This is an important concept for interaction relies upon the successful communication of information through modalities. Subjects were unable at times to demonstrate that flow.

This is reinforced by the wide range of responses to the second questionnaire question, which indicates that subjects had mixed feelings about their correct use of the definition of types. That identifying modality types was found to be so difficult by almost half the subjects casts doubts on the usability of the definition in practice.

## **6.4 Understanding of Interaction**

In order for the definition and methodology to be considered useful, they must allow insight into the interaction. This is examined in this section using the in-depth assessment of one part of the interaction sequence for the tube ticket machine task, and the interaction listing stage of the reports. The use of the notation, preconditions and AND/OR, is assessed to determine how far the notation and methodology supported subjects in gaining a better understanding of the structure of the interaction, thereby giving them a greater opportunity for insights into relevant usability issues.

### **6.4.1 Results**

#### ***6.4.1.1 In-depth Examination***

The small portion of the written interaction listings detailing the stage of interaction with the tube ticket machine just before the user inserts a ten pound note are examined here. The structure of the portion of interaction listed is examined to see what insight the twenty-two subjects gained into the interaction by using the notation. This section was also examined to determine the use of preconditions and AND/OR to see how well subjects were able to represent the structure of the interaction using them.

#### 6.4.1.1.1 Interaction Cycles

By examining how subjects structured this particular portion of the interaction sequence, it is possible to gain an insight into their understanding of the interaction. The cycle of interaction proposed by (Norman, 1988) envisages interaction in terms of a user acting with the system accepting that action and providing feedback which the user then evaluates. This is similar to the way nine of the twenty-two reports examined here {9,11,13,15,16,19,20,21,26} structured the interaction, writing that section of interaction as part of a sequence of modalities, with the user acting and the system responding. They most commonly took the form UE, SR, **SE**, **UR**, the modalities in bold being the ones examined in detail. An example taken from report {16} is shown in figure 6-1. The listing is numbered 5 at that stage, since previous parts of the interaction have already been detailed. At this stage of interaction, the user expressive action, the system receptive response, the system expressive action and the user receptive response are clearly shown.

```
5.  
  
[UE hap-sym-dis]  
*Select zones 1-4.*  
precon: [vis-lex-cont]  
*Read display*  
  
[SR hap-sym-dis]  
*Zones 1-4 selected*  
precon: [UE hap-sym-dis]  
*User selects zones 1-4.*  
  
[SE vis-lex-cont]  
*Display travelcard price.*  
precon: [SR hap-sym-dis]  
*Zones 1-4 selected.*  
  
[UR vis-lex-cont]  
*Read displayed price.*  
precon: [SE vis-lex-cont]  
*Display travelcard price.*
```

Figure 6-1: Section of interaction from report {16}



6.

**[SE vis-lex-cont]**  
**\*displays 'please pay: £3.80'**  
**Precon: user looking at display**

**[UR vis-lex-cont]**  
**\*sees 'please pay: £3.80'**  
**precon: user looking at display**

**[UR vis-sym-cont]**  
**\*see note slot\***  
**precon: user scans device**

**[UE hap-con-dis]**  
**\*inserts note in slot slightly\***  
**precon: [UR vis-sym-cont]**  
**\*see note slot\***

**[SR hap-sym-dis]**  
**\*note inserted slightly in slot\***  
**precon: [UE hap-sym-dis]**  
**\*see note slot\***

**Figure 6-2: Section of interaction from report {8}**

Seven reports {8,10,14,23,25,27,28} took a different perspective, writing the section of the interaction in terms of the system acting first, thus not having any preliminary user action and system reaction in that part of the listing. An example from report {8} is shown in figure 6-2.

The above reports wrote the section examined as a complete stage of interaction in its own right, or as part of a larger stage of interaction. However, five reports wrote this section of the interaction as split over two discrete stages. Report {1} wrote the system modalities bridging two stages of interaction, with UR, UE, SR, SE modalities in the first stage, followed by UR, UE, SR, SE modalities in the second, with the modalities examined in detail highlighted in bold. This is a reasonable split, and simply has a different version of the interaction cycle, starting with the user receiving modalities and then acting, as opposed to the two cycles examined above.

The remaining four reports split the interaction sequence wrongly, believing that the system modalities of the coin slot and the light did not happen at the same time as the system display changing. They therefore wrote the modalities in two discrete stages, of SE modalities and UR modalities, followed by SE modalities and UR modalities, rather than as one complete stage. These sequences were therefore incorrect, and the stages do not follow a consistent pattern of sequence of modalities.

### 6.4.1.1.2 Preconditions

Four reports {11,13,25,28} did not use any preconditions at all for this section of the interaction sequence. Nine reports used the wrong types for the preconditions. Ten reports omitted modalities, with six reports {10,11,13,21,25,28} giving no preconditions for the individual modalities, although {10,21} gave general preconditions in natural language. The remaining four reports {6,7,15,22} missed out preconditions for SE modalities. Table 6-12 summarises these results.

### 6.4.1.1.3 Use of AND/OR

The use of AND/OR was not relevant for the majority of the reports at this stage since many reports did not have multiple modalities and did not need to show that modalities were concurrent or that a choice between modalities was necessary. Out of the nine reports which used AND/OR, three {6,7,9} used them incorrectly. Report {6} mistakenly used OR instead of AND between two system expressive modalities, and placed an unnecessary OR

	MIS S	NAT LAN	WR	UN	UE FOR SR	UE FOR SE	UR FOR SE	UR FOR SR	SE FOR SR	SR FOR SE	NONE USED	
1												Key
2												Errors made
4												
6												Missing modalities
7												MIS
8												
9												
10												Natural language
11												NAT LAN
13												
14												Wrong modality
15												WR
16												
19												Unnecessary Modality
20												UN
21												
22												UE type used instead of SR
23												UE FOR SR
25												
26												No preconds used
27												NONE USED
28												

**Table 6-12: Use of preconditions in section of interaction listings examined in-depth**

between two preconditions. Report {7} mistakenly used OR instead of AND between two user receptive modalities, and {9} used AND when it was not necessary.



Out of the ten reports where AND/OR could conceivably have been used, three reports {2,8,14} omitted AND/OR where they should have been included. Report {2} needed ANDs to link various system expressive modalities. Report {8} needed to place OR between user receptive modalities. Report {14} had a missing AND between system expressive modalities, and needed an AND or an OR between user receptive modalities.

#### ***6.4.1.2 Report Results***

In order to determine the use of preconditions and AND/OR on a wider scale, the modality interaction listing stages of all twenty-eight reports were examined here. This section of the methodology requires the writing out in full of the interaction for both tasks, with the appropriate modalities used in sequence. The use of preconditions by the subjects is first examined, followed by the use of AND/OR.

##### **6.4.1.2.1 Preconditions**

There were various problems with the use of preconditions in the modality interaction sequence listing stage of the reports, with twenty-two of the twenty-eight reports examined omitting at least one precondition (see table 6-13). Fourteen reports had wrong preconditions, in that the precondition used was incorrect, and seventeen reports had wrongly written preconditions, in that the wrong type was used (e.g. UE instead of UR). Ten reports had unnecessary preconditions, with the precondition used not needed at this point. Three of the reports {21,13,4} had some or all of their preconditions not explicitly linked to a modality, but “hovering” after an interaction stage.

	PRECONS OMITTED	UNNECESS PRECONS	WRONG PRECONS	WRONG TYPE
1				
2				
3				
4				
5				
6				
7				
8				
9				
10				
11				
12				
13				
14				
15				
16				
17				
18				
19				
20				
21				
22				
23				
24				
25				
26				
27				
28				

Key

Errors made



**Table 6-13: Table showing use of preconditions in modality interaction sequence listings**

#### 6.4.1.2.2 Use of AND/OR

As table 6-14 demonstrates, fifteen subjects used AND/OR in the tube ticket machine interaction listings, and sixteen used AND/OR in the web site heating control panel interaction listings, with nine subjects not using AND/OR anywhere in their reports. Five subjects used AND/OR wrongly in their reports, the main type of mistake being with the use of AND. Two of the reports {5,27} used an AND between two modalities, implying that they happened simultaneously, when actually one of the modalities was causally dependent upon the other. Other reports {1,6} mixed the use of AND and OR, and the last report {22} mistakenly listed a precondition with an AND value, when the modalities were not written with one earlier in the listing.



1	Correctly used
2	Correctly used
3	Did not use
4	Correctly used
5	Errors in use
6	Errors in use
7	Correctly used
8	Correctly used
9	Correctly used
10	Did not use
11	Did not use
12	Did not use
13	Did not use
14	Correctly used
15	Correctly used
16	Did not use
17	Did not use
18	Did not use
19	Did not use
20	Correctly used
21	Correctly used
22	Errors in use
23	Correctly used
24	Correctly used
25	Correctly used
26	Correctly used
27	Errors in use
28	Correctly used

**Table 6-14: Use of AND/OR in modality interaction listings stages of reports**

### 6.4.2 Discussion

The results are discussed to determine how well the identification of modalities and the writing of the interaction sequence in the notation using AND/OR and preconditions gave the subjects an understanding of the structure of the interaction, and provided opportunities for insight into associated usability issues. From this various points are made relating to the overall aims of assessing the definition and taxonomy and methodology in terms of usability and usefulness.

#### 6.4.2.1.1 Interaction Cycles

The in-depth examination of one portion of the tube ticket machine task interaction sequence showed that most of the subjects segmented the interaction into discrete cycles, emphasising the flow of the interaction. Seventeen of the twenty-two subjects examined here were able to do so correctly, implying that they had a clear understanding of the interaction. Nine subjects wrote it in terms of the flow of information from the user to the system, similar to the cycle of interaction proposed by (Norman, 1988), and seven in terms



of the flow of information from the system to the user. The notation supports this segmentation, thus directing the subjects towards the notion of cycles of interaction events. Although some kind of structure is therefore presupposed, no one cycle form is imposed; subjects are free to structure the cycle according to their needs and understanding of the interaction, and are not constrained. That so many subjects were able to regard the interaction in this light, whether or not they were conscious of doing so, suggests that the notation afforded the subjects an insight into the structure of the interaction.

#### 6.4.2.1.2 Preconditions

The examination of the use of preconditions demonstrates subjects' understanding of the flow of the interaction, and how well they could identify which modalities were necessary before other modalities could occur. This is a difficult concept, since it requires a very clear understanding of the causality of the interaction.

Although the subjects were generally able to determine the causal effect of one modality on another, and were able to state that modalities had preconditions, four reports in the portion of the tube ticket machine interaction sequence examined in-depth did not use any preconditions at all, and ten out of the twenty-eight reports examined in the interaction sequences had unnecessary preconditions, showing that some subjects were confused as to where and how to use preconditions. This suggests that the subjects may have been unhappy in the use of the preconditions, or were perhaps unclear as to the causal relationships between the modalities at that point, implying a lack of understanding regarding the flow of information in the interaction sequences.

Nine of the reports examined in-depth and seventeen reports overall used the wrong types for the preconditions. This demonstrates that many subjects were confused as to which modalities were the preconditions for existing modalities, implying that they were unsure as to the flow of the interaction and which modalities were the preconditions for what. The omission of many preconditions by the subjects reinforces this uncertainty.

Therefore it seems that the subjects were unsure about the use of preconditions, finding it difficult to decide where they should be used, and confusion as to what type they should be when they were used.



#### 6.4.2.1.3 Use of AND/OR

The use of AND/OR in the interaction sequences highlights concurrency and choice between modalities. Therefore in order to use AND/OR the subjects needed to be able to assess the structure of the interaction in these terms and represent it accordingly.

Many subjects did not use AND/OR, despite there being several occasions in both tasks where it could be considered appropriate to show that modalities were concurrent or that a choice between modalities was necessary. This may be due to subjects being unsure about how to use AND/OR and thus writing the sequences so as to avoid its use altogether, or subjects being unable to determine possibilities for such use within the interaction sequences.

The few errors made by those subjects who did use AND/OR in their interaction sequences shows that these subjects were generally able to correctly identify where modalities were concurrent and where there were choices between particular modalities. This implies that these subjects were able to assess those aspects of the structure of the interaction and represent them accordingly. Therefore, although some subjects had reservations about using AND/OR, when they did use it they mostly used it correctly.

### **6.5 Identification of Usability Issues**

The identification of usability issues by the subjects is assessed in this section in order to determine how the definition and methodology provided opportunities for insight. This is assessed in three ways: by examining the number of clashes and properties identified by the subjects; the usability issues raised during and at the end of the reports; and subjects' perceptions of using the methodology and its component stages from the two questionnaires.

#### **6.5.1 Results**

##### ***6.5.1.1 Properties and Clashes***

The properties and clashes identified by the twenty-eight subjects in stage six of the reports for both tasks are examined here and presented in table 6-15. Twenty-four of the reports

mentioned at least one property or clash. Six reports did not mention any in connection with the tube ticket machine, and ten reports did not mention any in connection with the heating control panel web site.

Ten reports mentioned physical clashes of some kind, but only five reports {2,5,8,13,22} mentioned the physical constraint that a user has to be looking at something in order to see it, which was explicitly mentioned in the worked example in the student pack, and would therefore be expected to be included as a minimum. The other physical clashes noted included the impossibility of entering two data streams at once {2,9}, issues involving the field of vision of the user {13,15,22,25,26}, and that the mouse has to be visible before it can be used {22}.

No report mentioned any lexical or semantic clashes, and although three subjects {1,15,20} mentioned temporal clashes of some sort, one of these {20} was confused as to what constituted a temporal clash.



Eight reports explicitly indicated a clash of the type clash unless expert. However, many of the subjects had explicitly done one or both tasks as an expert user, thus making the identification of this type of clash unnecessary.

Of the eight reports which mentioned redundancy, five of these {6,7,9,15,24} mentioned the redundancy of the tube ticket machine using both light and sound when money to be inserted. Report {14} talked about the importance of redundancy in helping the user by providing reinforcing information.

Atomic properties and mismatches were largely ignored, with only three reports {2,20,23} mentioning them. Report {23} mentioned atomic properties in passing, {2} went into detail about composite and atomic modalities, and {20} talked about frequent mismatches in the heating control panel task.

Eight subjects made mistakes in identifying clashes and properties. The mistakes were of three types: two reports misinterpreted redundancy {6,7}; five reports wrongly assigned the label temporal to clashes, or else wrongly stated that there was a clash at this point {1,2,15,17,28}; and one report misused a "clash-unless-expert" {11}.



	PHYS	LEX	SEM	TEMP	EXP	RDN	ATMC	MISM	
1									Key
2									
3									Properties and clashes mentioned 
4									
5									
6									Properties and clashes not mentioned 
7									
8									
9									Physical clashes PHYS
10									
11									Lexical clashes LEX
12									
13									Semantic clashes SEM
14									
15									Temporal clashes TEMP
16									
17									Clash unless expert EXP
18									
19									Redundancy RDN
20									
21									Atomic modalities ATMC
22									
23									Mismatches MISM
24									
25									
26									
27									
28									

**Table 6-15: Table showing use of properties and clashes in reports**

### 6.5.1.2 Usability Issues

The subjects collectively identified sixty-three separate types of issue relating to the tube ticket machine within the context of the task specified. These covered a wide spectrum, including issues related to the buttons, use of sound, the instructions, the display, layout, and availability of the ticket (see table 6-20 for the full listing of issues raised). Table 6-16 shows the number of issues that subjects identified, with subjects typically identifying four or fewer issues (nineteen of the twenty-eight reports). Thirteen of the reports made no mention of the fact that the purchase of the ticket was not possible. Three reports on the tube ticket machine interaction made no recommendations for change or identified any usability issues of concern, claiming that the system was usable.

Seventy-four separate types of issue relating to the heating control panel were identified by the subjects in stage eight of their reports, covering issues such as time settings, feedback and error recovery (see table 6-21). As shown in table 6-17, all the reports raised at least one issue, with twenty of the reports identifying four issues or less.







	No of issues raised in Stage 8	Issues in Stage 8 also mentioned earlier	Issues raised in Stage 8 and not mentioned earlier	Uncertain as to mentioned earlier	Mentioned earlier but not included
1	1	1	0	0	1
2	0	0	0	0	3
3	4	1	3	0	0
4	3	2	1	0	3
5	2	2	0	0	1
6	1	1	0	0	1
7	1	1	0	0	0
8	1	0	1	0	1
9	4	3	1	0	4
10	4	1	3	0	1
11	6	2	4	0	1
12	3	0	3	0	3
13	4	0	4	0	3
14	4	2	2	0	3
15	5	1	4	0	2
16	1	0	1	0	1
17	12	0	12	0	2
18	2	1	1	0	0
19	0	0	0	0	1
20	3	2	1	0	1
21	3	2	1	0	1
22	3	3	0	0	0
23	4	1	3	0	2
24	6	0	6	0	1
25	3	1	2	0	3
26	9	3	6	0	1
27	2	2	0	0	1
28	0	0	0	0	3

**Table 6-18: Usability issues mentioned during tube ticket machine task reports**

	No of issues raised in Stage 8	Issues in Stage 8 also mentioned earlier	Issues raised in Stage 8 and not mentioned earlier	Uncertain as to mentioned earlier	Mentioned earlier but not included
1	2	0	1	1	1
2	2	0	1	1	6
3	2	1	0	1	2
4	5	1	4	0	1
5	3	1	1	1	1
6	1	1	0	0	4
7	3	3	0	0	2
8	5	0	5	0	1
9	3	0	3	0	1
10	9	1	8	0	0
11	4	2	2	0	2
12	3	0	3	0	1
13	3	1	2	0	0
14	7	0	7	0	1
15	4	1	3	0	0
16	4	1	2	1	0
17	4	0	4	0	4
18	2	1	1	0	0
19	3	1	1	1	4
20	9	0	6	3	5
21	4	1	3	0	1
22	7	2	5	0	0
23	6	0	6	0	0
24	3	0	3	0	0
25	5	1	3	1	2
26	4	2	2	0	1
27	3	0	3	0	0
28	4	1	3	0	1

**Table 6-19: Usability issues mentioned during heating control panel web site reports**

<b>No problems found</b>	
no usability issues raised, or no recommendations for change	(2,19,28)
<b>Buttons</b>	
buttons should light up when pressed	(11)
travelcard button should flash to show alternative	(11)
too many buttons	(17,24)
buttons difficult to read	(17,24)
no sculptured buttons	(24)
cancel button should go back a stage rather than right back to beginning	(23)
buttons should be bigger	(12)
buttons should use bigger text	(17)
buttons should use brighter text	(17)
<b>Sound</b>	
needs auditory feedback for the visually impaired	(26)
needs repositioning or a plastic shell to allow auditory feedback to be heard more clearly	(26)
needs better use of auditory feedback	(8,10,12,17,21)
no sound feedback when buttons pushed	(24)
use of sound redundant	(1,6,7)
feedback sounds should be higher	(23)
doesn't use sound	(3,17)
should use voice for special needs customers	(18)
use of sound not useful due to noisy environment	(9)
<b>Help/Instructions</b>	
there should be a guide showing how to use the system	(10,17)
instructions should be short and precise	(17)
instructions for new users do not grab attention	(20)
instructions visibility could be improved through use of colour	(20)
help option should be available	(20)
help option should be located away from machine	(20)
<b>Display</b>	
display area is small and badly lit	(10,24)
display area should be made bigger	(10,11,12,23)
font-colour of screen not attractive	(24)
flashing of "Not Available" display unnecessary	(9)
"change given" display irrelevant to new user	(20)
"change given" display should be changed to "ready"	(20)
<b>Layout</b>	
layout can cause confusion	(26)
clash with note-slot and screen message	(26)
note-slot should be nearer coin slot	(25,26)
cancel key is difficult to locate	(26)
green light in money slot redundant	(15)
change/ticket slot should be labelled	(22)
screen should be closer to coin and note slots	(25)
change given sign should be scaled down	(25)
note-slot should be wider and easier to use	(17)
<b>Ticket</b>	
system does not allow return ticket to be bought	(5,9,11,15,27)
system does not show travelcard as alternative to return ticket	(11)
system should explain that travelcard is available and why	(11,15,20,21,23,27)
there should be better feedback about tickets not available at certain times	(5,14,22)
system should allow a return ticket to be bought after 10am	(15,18)
"Not Available" message confusing	(3,4,15,20,21,22)
<b>Other</b>	
system only in English	(3,17,26)
interface should use icons	(26)
system status not always obvious	(26)
system resets after "Not Available" display, which may confuse users	(14)
users may be confused by multiple visual information	(4,14)
users may not notice all of the visual information	(4,14)
general lighting should be improved to make environment more relaxing	(16)
system difficult for wheelchair users to use	(17)
system doesn't always accept money	(13)
child options hidden beneath flap	(13)
child options access should be improved	(13)
visa/switch payment methods should be incorporated	(13)
system looks big and intimidating	(9,17)
change given in coins only	(20)
change should be given in notes	(20)
system feedback is poor	(3,20)
novice users would not know what stations are in what zones	(20)
machine should have zone identification facility	(20)

**Table 6-20: Types of issues raised in tube ticket machine reports**



<b>Instructions</b>	
instruction guide not clear	(14,19,20,22)
instructions do not adequately explain copy function	(2)
there should be a sign saying "instructions below"	(22)
system is difficult to use without reading instructions	(28)
instructions are hidden	(13,20,22,28)
<b>Time Settings</b>	
system should allow users to type in days and times	(10,11,15,20,27)
timing settings can only be incremented by ten minutes at a time	(8)
alternative divisions of time for incrementing timings should be used, e.g. half an hour or one hour	(3,8)
setting of times is long-winded	(11,16,20,27)
possible to set OFF time earlier than ON time	(16)
users cannot type times into boxes	(14)
there should be a way of changing the times using less button clicks	(4,18)
<b>Time Periods and Days</b>	
setting three different ON/OFF times a day confusing	(11)
ON/OFF displays should show times for all three periods	(11)
user has to click through second and third periods even when the user doesn't want to use them	(14)
system displays Tuesday as first day	(15,22)
system should display Monday as first day	(15,22)
user should only have to set one period a day	(15)
problems with time periods	(12)
use of three time periods difficult since ones not needed cannot be set to 0	(1)
system does not allow days to be missed out when copying	(23)
there needs to be a way of allowing days to be copied out of order	(14,23)
the settings for the first day should be the default for the rest of the week	(16,18)
the system should allow flexibility of moving between period settings in any one day	(4,23)
time periods confusing	(25)
time periods should be removed	(12)
<b>Buttons</b>	
advance button functions should be separated out	(20)
advance button should be relabelled "next"	(24)
copy button doesn't always work	(5)
advance button has too many functions	(20)
labelling of buttons is confusing	(10)
buttons need clearer labels	(5)
button functions not known by looking at them	(13)
sometimes buttons have to be clicked twice before anything happens	(13)
copy button is confusing	(25)
advance button difficult to understand without reading instructions	(24)
several clicks on advance button can confuse novice users	(7)
<b>Feedback</b>	
no feedback given for success of copy function	(10,22,26,27)
feedback necessary for copy function	(25,26)
lack of auditory modalities	(17,26)
no feedback when task completed	(10,19,20)
there should be dialogic boxes to give feedback	(10)
copied settings should appear to provide feedback before system moves on to next screen	(22)
auditory feedback should be used	(8,10,25)
lack of effective feedback	(8,19,20,21,28)
system needs better feedback	(4,14)
system needs to provide feedback for times entered	(23)
system needs to provide feedback that value entered is correct	(23)
<b>Error Recovery</b>	
no error recovery	(20,21)
no back button for corrections	(28)
system needs a "reset" button	(23)
system needs a "cancel" button	(7)
system needs a "backspace" button	(12)
<b>Font and Colour</b>	
background colours initially distracting	(16)
font-size should be varied to support handicapped users	(17)
cursor and colour sometimes redundant	(7)
<b>Other</b>	
touch-screen may be a better input device than a mouse	(26)
the two boxes make for poor mappings	(10)
system should give drop-down lists of options	(10)
support of typing would speed system up	(10)
system does not support novice users	(17)
too many things happening on screen at one time	(8)
use of pull-down menus would speed up process	(14)
needs simple help-system	(14)
harsh on fingers	(2)
language different to ordinary heating control panels	(24)
system very complicated	(17,21)
system is repetitive	(9)
repetition can lead to errors	(9)
system needs meaningful and self-explanatory icons	(21)
repetition of user information unnecessary	(6)
very long	(1,3,4,5,9)
reduce number of steps and button clicks	(4)
tedious and laborious	(25)

Table 6-21: Types of issues raised in heating control panel reports

Subjects did not necessarily include issues in stage eight of the methodology that were raised in earlier sections (see tables 6-18 and 6-19). The issues referred to here are those raised separately by the subjects, rather than after they had been grouped together by category and identical issues combined, as in the previous tables showing the issues raised. Twenty-three reports mentioned issues relating to the usability of the tube ticket machine interaction but did not include them at the end, a total of forty separate issues, and nineteen reports mentioned issues relating to the usability of the heating control panel interaction but did not include them at the end, a total of forty-one issues.

Nineteen of the tube ticket machine reports mentioned issues raised in stage eight earlier in the report, a total of thirty-three issues, and nineteen reports on the same device mentioned issues in stage eight which were not explicitly mentioned earlier in the reports, a total of sixty-seven issues. Seventeen of the heating control panel reports mentioned issues raised in stage eight earlier in the report, a total of twenty-two issues, and twenty-six reports on the same device mentioned issues in stage eight which were not explicitly mentioned earlier in the reports, a total of ninety-two issues.

### ***6.5.1.3 Questionnaire Results***

In order to examine subjects' experience of identifying clashes and properties, and the methodology, some of the responses to the two questionnaires are examined. The first questionnaire was completed by the subjects after the initial introductory lecture on the definition and associated methodology, and before the opportunity to do any practical work using it. The second questionnaire was completed by nineteen of the twenty-eight subjects after the reports had been submitted. Please refer to Appendix G for the results in full.

#### **6.5.1.3.1 First Questionnaire Results**

The responses examined here relate to confidence in applying the methodology and identifying clashes and properties, and previous experience in HCI and analysis. The seven questions relate to: confidence in using the methodology; confidence in identifying modality properties; confidence in identifying modality clashes; experience in HCI; experience in structured HCI reports; previous experience of HCI analysis; and experience in applying HCI and software engineering methodologies.



The subjects were generally very confident about using the methodology, with fifteen out of the twenty-five subjects who answered this question giving positive responses. Eighteen subjects initially felt confident about identifying modality properties, with four of these very confident. Two-thirds of the respondents (eighteen subjects) gave answers indicating that they were confident about identifying modality clashes, with four of these very confident.

Eighteen subjects considered themselves to be experienced in HCI and gave responses of over 50%, and seven subjects did not consider themselves to be experienced, giving responses of 50% or under. Fourteen subjects did not feel experienced in writing structured HCI reports and gave responses of 50% and under, with eleven subjects feeling experienced, giving responses of over 50%. Almost half the subjects (twelve) felt that they did not have much previous experience in HCI analysis and gave a response of 50% or less. Of the thirteen subjects who felt experienced in HCI analysis, four of them gave responses of over 80%, indicating that they felt very experienced in this field. Sixteen of the subjects did not consider themselves experienced in applying HCI and software engineering methodologies, and gave responses of 50% or under. Nine subjects considered themselves experienced and gave responses of over 51%, with one of these giving a response of over 80% indicating high experience.

#### 6.5.1.3.2 Second Questionnaire Results

The responses examined here relate to subjects' confidence in identifying modality properties and clashes, and their use of the methodology as a whole, and with regards to the individual stages.

The responses to seven questions are presented here, relating to: subjects' confidence as to the correct identification of modality properties; subjects' confidence as to the correct identification of modality clashes; how straightforward the subjects found the methodology to apply; subjects' confidence in using the methodology again; how easy the subjects found each of the stages of the methodology to apply; and how useful in uncovering usability issues the subjects found each of the stages of the methodology.

There was a very wide range of responses to the question concerning the correct identification of modality properties, from 3% to 93%, showing that subjects varied greatly

in their confidence levels. Despite nine subjects expressing confidence, with one response of 93%, the majority of the respondents (ten replies) were not confident, one giving a response of as low as 3%. The overall response to the correct identification of modality clashes was negative, with fourteen subjects indicating a lack of confidence, and eight of these subjects giving responses of 20% or under, showing very low confidence levels. That this was not the experience of all the subjects is evidenced by two of the five confident subjects giving responses of over 80%, however generally subjects were not confident regarding this issue.

Over half the subjects (ten responses) did not find the methodology to be straightforward, and gave responses of 50% and under. A majority of ten subjects gave responses indicating a lack of confidence in using this methodology in the future, four of these indicating very low levels of confidence. Overall, subjects found the methodology easy to apply, with thirteen subjects giving responses of over 50%, and one giving a response of over 80%.

Overall, the subjects found stages one, two and three of the methodology to be easy. Stages four, five, seven and eight they found more problematic, with opinion divided, and stage six, the identification of properties and clashes was found to be very difficult.

The subjects found stages one, two and three of the methodology to be useful. The subjects were divided as to the usefulness of stages four, five, seven and eight, and did not consider stage six, relating to the identification of clashes and properties, to be useful at all.

## **6.5.2 Discussion**

These results are now discussed to determine how well the methodology supported the identification of usability issues. From this various points are made relating to the overall aims of assessing the definition and taxonomy and methodology in terms of usability and usefulness.

### **6.5.2.1.1 Properties and Clashes**

Subjects initially felt confident about identifying modality clashes and properties, suggesting that these issues were usable on a theoretical basis. However, despite this initial confidence, and properties and clashes being explicitly described in the student pack, the



subjects identified very few occurrences, with eight reports not mentioning any. This was despite there being ample opportunity to identify various physical clashes or instances of redundancy at least. This could be for various reasons: the subjects may have been unsure as to what clashes and properties were, and thus unable to identify them; subjects may have understood what clashes and properties were, but been unable to translate this understanding to the analysis of the interaction sequence listing and so could not identify them; or, since this is one of the last stages of the methodology, subjects at this point may have been concentrating on finishing the reports and so did not wish to do further analysis.

From the examination of responses to the second questionnaire, it seems that subjects were not happy with the use of clashes and properties, both questions receiving a majority of negative responses. This implies that the subjects had trouble identifying them, or could not identify them from the interaction sequence, rather than the subjects being constrained by time considerations. This is reinforced by the responses for questions ten and eleven, where most subjects gave responses indicating that they found this stage to be difficult and not useful in identifying usability issues.

However, that those clashes and properties that were identified were mostly identified correctly shows that although subjects were reluctant to identify properties and clashes, and were not confident about having done so correctly, they were able to do so correctly when they chose to.

#### 6.5.2.1.2 Usability Issues

When considering the use of the methodology to identify usability issues it must be borne in mind that only eighteen subjects considered themselves to be experienced in HCI. This was despite all subjects having completed at least one previous course in it and currently taking a course in HCI. A majority of the subjects did not feel experienced in writing structured HCI reports, HCI analysis, and in applying HCI and software engineering methodologies. This lack of perceived experience in HCI and in using methodologies may therefore have a bearing on the subjects' expectations of the methodology, and influence their responses.

The majority of the subjects were initially confident about using the methodology, which is interesting given that their answers to the first questionnaire about their previous experience

of applying HCI was at best mixed. This may be because at this point the subjects had no practical experience in applying the methodology. It is encouraging that they felt optimistic, since the lecture had introduced them to a large number of new concepts. This suggests that the methodology appeared to be straightforward on a theoretical basis.

However, the subjects had problems with using the methodology practically, since although thirteen of the nineteen respondents indicated that they found the methodology easy to apply, subjects were divided both as to the straightforwardness of using the methodology, and in using the methodology again. In the breakdown of the ease of use of the various stages, only stage six was found to be difficult, with subjects finding stages one, two and three easy and being undecided about stages four, five, seven and eight. This corresponds to their feelings about the usefulness of the various stages in identifying usability issues, implying that subjects found those stages that were easy to be more useful than those stages that were difficult.

Assessing the quality of particular usability issues, as opposed to sheer quantity of issues identified, is notoriously difficult (Gray and Salzman, 1998). In the examination of the reports for usability issues mentioned, only criticisms and solutions to criticisms were noted, since the aim was to identify usability issues rather than any positive aspects of the particular interfaces. Using merely the number of issues identified or recommendations made is a crude means of assessment, since subjects may have approached the final section of the report in differing ways. Some only felt the need for a brief, overall summary of issues, others went into detail based on everything observed at each stage.

The use of purely numerical values also obscures the ideas themselves in terms of quality and quantity. For example, report {15} talks about the issue of the heating control panel displaying Tuesday as the first day, and having to click the minus button to go back to Monday. However, in the final report, this is counted as two separate items: the issue of the system displaying Tuesday as the first day, and the recommendation that the system display Monday as the first day. Subject {17} spoke about twelve issues in stage eight of the tube ticket machine report, but nowhere mentioned that the task is impossible. Strictly numerical values must therefore be treated with caution, and used merely to gain a rough idea of number of issues and recommendations raised rather than the inherent value of those issues.



Many issues raised at various stages during the methodology were not mentioned at the end, indicating possible confusion in what was expected in the final report. Some of these issues definitely affected the overall usability of the interface, and were not credited as such. One particular example is that some subjects were not able to identify either that the tube ticket machine task could not be successfully completed, or identify it as a usability issue. This suggests that either the methodology and notation did not support the subjects in this realisation, or that the subjects made the realisation but did not consider it to be a usability issue of importance. This may relate back to their experience in HCI and in conducting analyses of this nature.

It seems that some issues, although identified, were simply not considered important enough to warrant inclusion in the final stage of the report. For example, subjects {6,8,19} all mentioned that the tube ticket was not available in the body of the report, but did not make any reference to this at the end. Report {12} explicitly stated at the end of the final stage for the tube ticket machine report: "Note: the reason why I stopped was due to the fact that the ticket could not be purchased so the task was impossible to complete", yet only talked about the need for a bigger screen and bigger buttons and auditory feedback as usability issues. Various subjects in the interaction listing for the web site heating controller wrote about having to click the day back from Tuesday to Monday in order to start setting the times but did not consider this to be an issue of importance. Despite the very long listings for the heating controller interaction listings, many subjects did not see this as a usability issue. For instance, report {19} in the discussion on granularity in stage three of the report wrote: "This list of modalities are quite long so I have simply stated that the steps were repeated for each day. As it would be incredibly long and very repetitive therefore producing problems analysing the system." However, in the final section on usability, the report mentioned only that the instructions were unclear, that there was no feedback for settings, and that there was no feedback for successful completion.

Many issues mentioned at the end as usability issues had no "history", or direct link to the earlier stages of the methodology. The methodology therefore allowed more opportunity for insight than might be apparent from examining the provenance of those issues mentioned at the end, but at the same time many subjects did not necessarily obtain their insights by the

process of using the methodology, and may have come up with equally numerous and valid insights by using a different technique altogether.

## **6.6 General Discussion**

Although case studies are not the only means of assessing the utility or otherwise of a technique, they have the potential to be effective as (Gray and Salzman, 1998) note. By doing this case study it is possible to examine the problems and issues surrounding the practical application of the definition and methodology. Close examination of these issues (here and in later chapters) will make it clear if improvements are needed to EMU, and to the supporting explanatory documentation. This is part of the iterative cycle described by (Blandford et al, 1998a), where a technique is adapted in light of the experience of practitioners, and where practitioners learn the scope and application of the technique. It is not possible within one case study to address all the issues; indeed, in order for a technique to achieve maturity, various iterative stages must be completed. This case study did however raise initial issues regarding the use and usefulness of the definition and methodology overall.

### **6.6.1 Identification of Modalities**

The definition of modality was clear and precise enough for the subjects to successfully identify the critical modalities in the portion of interaction examined in-depth. In order to make this correct identification, a good understanding of the flow of the interaction and which modalities were essential to that flow was required. This also gave opportunity for subjects to gain an insight into the usability issues associated with that flow. Given the initial levels of confidence, and correctness of use, the definition of modality appears to be usable, although the mixed feelings about the identification of modalities in the second questionnaire implies some problems in using the definition.

The subjects were generally able to apply the taxonomy correctly, demonstrating not only that the taxonomy was clear and the categories distinguishable, but that the subjects were able to be precise about the information contained in the interactions. Until the components of the modalities can be established, the usability factors affecting the modalities are more difficult to determine.



The correct identification of modalities and the correct application of the taxonomy were important, since they are central to EMU. However, it seems that subjects occasionally found it difficult to separate out the user's perspective when describing the operation of the system, as shown by the confusion between the identification of visual and haptic modalities, and the problems surrounding the representation of the modalities for button pressing. This is understandable, since it can be difficult to be purely objective and step away from a human perspective; however, it is an extremely relevant point with reference to all modelling where both the user and the system must be represented. As such it is an issue that must be addressed by any further development of the methodology. This may be achieved through the use of practical examples or heuristics in the supporting documentation to stress the need to approach the system modalities from the system's perspective.

With regards to the use of types, subjects were initially confident, and were able to distinguish between user and system modalities without problems. They had some difficulty in separating out the receptive and expressive modalities, and in remembering to include the receptive modalities, and their final confidence levels were mixed. The definition and identification of types is therefore not as clear as it could be, and needs improving. Even so, many subjects used it correctly, and it must be emphasised that this was their first attempt at using this definition. The results are therefore reasonably positive.

The correct identification of types requires a thorough understanding of what is happening in the interaction. That many subjects were initially confident in the use of types, and were able to successfully represent them, demonstrates that most subjects had a clear understanding, thus giving them greater opportunity through increased knowledge of identifying any problems. However, the problems with the use of expressive and receptive types, and the omission of receptive types, shows that subjects had problems with its use and were often confused, which could obscure interaction issues rather than render them opaque. Again, this is something that could be addressed by better supporting documentation.

## 6.6.2 Understanding of Interaction

The subjects were generally able to use the notational part of the methodology, as the examination of the in-depth portion of the tube ticket interaction showed. The apparent tendency of some subjects to segment the interaction into cycles of interaction, following the traditional cycle proposed by (Norman, 1988) or other interpretations, is an encouraging sight since it suggests that the notation supports that kind of representation and could be something to be exploited in further refinements of the methodology. By encouraging practitioners to think in such terms when writing the interaction sequence, it would be possible for them to gain greater insights into the structure of the interaction, and thus the usability issues surrounding the flow of the interaction.

The mixed use of preconditions, with some missed out and some unnecessary, demonstrates that many subjects were unsure about their correct use. Where they were correctly used, it indicated that subjects were aware of the causal flow of the interaction.

The difficulty that many subjects had with precondition types reinforces the need for the definition and application of types to be made clearer. This problem is possibly again related back to the difficulty of achieving an objective view of the interaction separate from the natural identification with the user, which reinforces the need for this issue to be clearly addressed.

That subjects were reluctant to use AND/OR suggests that this is another area where more assistance is required within the methodology, possibly in the form of heuristics and examples. However, where subjects did use it, they generally used it correctly, demonstrating that the overall concept was usable. By using AND/OR, subjects demonstrated an understanding of issues of concurrency and choice, which are of great importance to multi-modal usability, and thus had an opportunity to identify problems with the interaction at these points.

The usability of the notation is a critical issue in the successful application of EMU, and it appears successful in the context of this case study. Problems with the use of preconditions do not significantly detract from the use of EMU. However, that students did not use important concepts such as AND and OR is a cause for concern, since the modelling of



choice and concurrency is central to multi-modal usability evaluation. This may be due to the quality of the supporting documentation, and needs addressing.

### **6.6.3 Identification of Usability Issues**

Subjects were generally unable to successfully identify clashes and properties, and so could not identify the usability issues surrounding them. At present it is not clear why this was so. If it was due to the subjects not understanding what clashes and properties are, this is a matter for clarification within the methodology, with more support needed for their explicit identification. If, however, it was a reflection on the length of time it took to complete the methodology, this is an issue of time taken and payback, as noted by Buckingham Shum & Hammond (1994). As such, it becomes important to determine whether the methodology is too long, and whether it might be appropriate to shorten it, both to make it quicker to use, and so that those aspects which have less value can be dropped in favour of those which are most powerful and can be more readily exploited.

The subjects found three of the methodology stages to be easy and useful, and had mixed responses regarding another four, with stage six perceived as both difficult and not useful. This is not surprising given that this was their first time in using the methodology, and their lack of experience in HCI and in conducting other analyses. However, that there were problems and mixed feelings with some of the stages needs to be addressed, possibly by clearer indications in the methodology as to the purpose of each stage, and more examples of how to do it. This is especially true for stage eight, where subjects were confused as to what was necessary in the final stage of the reports. A report structure could be included so that analysts are explicitly aware that all issues raised in the body of the report should be written there.

That subjects did not include issues of prime relevance, for example that the tube ticket machine task was not possible in the form specified, into the final overall usability assessment is more serious. It suggests either that subjects were unclear as to what constituted a usability issue, or that subjects were reluctant to find fault with the interface. This raises important points about the internal agenda of analysts, and that any analyst may find it difficult to make judgements regarding the overall usability of an interface, regardless of the issues discovered during the analysis. It also raises questions as to what

constitutes a usability issue. Nowhere in the documentation given to the students was the concept of a usability issue explicitly defined. However, there are no absolute standards for usability, since it is often determined by how well the interface performs in the context of use. Therefore another way of determining usability issues, for instance in terms of the behaviour of various parameters such as speed of use, might be needed, with supporting documentation and examples. This will be further discussed in Chapter 8. The issues mentioned at the end of the report but without any direct link to previous stages of the analysis allows issues of craft skill to be separated out. The value of any analysis to give insight is widely known, but what is important here is how powerful this particular methodology is in providing insight into multi-modal issues. By separating out those issues mentioned purely at the end, and those mentioned in the body of the report, it is possible to track where some subjects achieved their understanding of various issues. It is encouraging that so many were identified during the reports, just as it is equally worrying that so many subjects, having identified an issue, did not recognise it as serious enough to affect the overall usability of the interfaces.

It is worrying that students could not identify properties and clashes, given their importance to multi-modal usability evaluation. It is equally disturbing that they were unable to correctly identify usability issues, although this may be due either to the lack of an explicit description of what constitutes a usability issue in the supporting usability documentation, or a lack of understanding of the context of use.

## **6.7 Summary**

This chapter has assessed the usability and usefulness of the modality definition, taxonomy and multi-modal methodology by means of a case study. Despite not being able to use EMU perfectly, the subjects generally applied it well.

The definition and taxonomy were both usable and useful. Subjects were able to use the definition to successfully identify modalities, and most modalities were written using the correct components. The use of modalities and the taxonomy categories gave subjects an opportunity for insight into the interaction, since it forced the subjects to be precise about the exchange of information. However, it was found that some system modalities were



wrongly written due to the tendency of some subjects to be biased towards the user's perspective. Subjects did have problems with the use of types and preconditions, and the definitions need to be strengthened in these areas.

The use of the methodology had mixed results. Subjects were generally able to use the notation and to successfully model the interactions, which allowed insight into the interaction and hence relevant usability issues. However, they had difficulty with the use of preconditions to determine causal relationships, and many subjects did not use AND/OR to model issues of concurrency and choice, although those that did generally did so correctly.

Very few subjects identified any clashes or properties using the methodology, which suggests that either subjects were unclear as to what these might be, and therefore the methodology needs more support in this area, or that the subjects did not have enough time for this stage, and therefore the methodology may be too long and needs condensing.

Subjects had mixed feelings regarding the methodology as a whole, although they found some stages to be both easy and useful. There seemed to be confusion as to the role of the final report in stage eight, with many subjects not mentioning issues identified earlier during the analysis at this stage. Subjects seemed either unclear as to what constituted a usability issue, or reluctant to find fault with interface. However, usability issues were identified through the use of the methodology, and the act of writing the final report concentrated the subjects on the usability of the interactions, with the result that additional issues which had no direct link to previous stages of the analysis were mentioned at this point.

This case study has provided an opportunity to study the application of EMU by non-developers and has highlighted areas where EMU needs improving, particularly with respect to the supporting documentation and what constitutes a usability issue. This chapter has concentrated on issues of usability and usefulness, with respect to the application of the modality definition, taxonomy and multi-modal methodology. The next chapter will approach the methodology from the perspective of its coverage and the kinds of issues that it can identify.

# **7. EVALUATION OF APPROACHES**

## **7.1 Introduction**

The previous chapters have presented the development of EMU and examined its usability by other analysts by means of a case study. However, the usefulness of EMU, in terms of its scope and coverage and the implications of user expertise, have not been assessed. This makes claims for the usefulness of EMU limited, since the individual users might be drawing on their own expertise in order to identify usability issues rather than gaining understanding through the use of the methodology, giving a misleading impression of its scope and coverage. It also raises questions regarding the origin of the issues uncovered by the analyses of the robotic arm using STN, CW, GOMS, PUM and Z.

This is a significant problem. Formal methods are considered to be of use to usability evaluation because they provide a means of communication, a verifiable basis for answers, and guaranteed coverage of the problem space (Hall, 1990). Research has concentrated on how well the problem space is covered, and how effective such methods are (John and Marks 1997, Blandford and Duke 1997). However, there has been little investigation into how well formal or semi-formal approaches actually provide this, and how much of their ability to identify issues is due to the power of the technique by itself as opposed to the inherent knowledge or expertise of the analyst using the technique. Gray and Salzman (1998) have demonstrated the problems of comparing techniques. The comparison of analyst expertise would be even more difficult. Without recognition of the role of skill in usability analysis there is a danger that the properties of the technique can be obscured by the skill of the person using it, and wider claims made for a technique than are actually justifiable.

In order to investigate these problems this chapter first examines empirical data of the robotic arm in use to determine which of the usability issues identified by the six analyses can actually be confirmed, and which were missed. The six approaches are then systematically re-analysed and the role of expertise in deriving usability issues considered



in depth by reference to source material on how to apply each technique. This is not to state that the empirical usability assessment of interfaces is a “gold standard”, but it does provide some kind of benchmark against which issues can be compared.

## 7.2 Video Data Analysis

There has been no previous empirical examination of the robotic arm interface in order to verify the usability issues discovered by EMU and the five other approaches. Unfortunately, the robotic arm was not available for user testing due to flood damage sustained in the summer of 1997. Thus in-depth empirical evaluation of the arm performing the same task as previously analysed was not possible. However, previous video footage of usability testing filmed by B. Parsons, the arm’s developer, was available, and was studied in order to determine whether any of the usability issues identified by the approaches could be supported by empirical evidence.

The video data examined comprised six excerpts, each one showing a user performing a specific task and using a particular means of input, as summarised in table 7-1. For a full description of the video data, users and findings, please refer to Appendix H.

EXCERPT:	One	Two	Three	Four	Five	Six
SECONDS:	123	83	89	50	89	525
INPUT:	Mouse	Voice	Gesture	Gesture	Voice	Mouse
TASK:	Feeding	Feeding	Feeding	Feeding	Feeding	Drinking
USER:	Expert	Expert	Expert	Novice	Novice	Novice
POSITIONS:	Pre-taught	Pre-taught	Pre-taught	Pre-taught	Pre-taught	Mixture

**Table 7-1: Summary table of video data excerpts**

The video data was analysed using a form of Exploratory Sequential Data Analysis (ESDA) (Sanderson & Fisher, 1994a). ESDA techniques are observational and empirical, and include task analysis, protocol analysis and video analysis. Since the robotic arm was no longer available but video data of the arm in use had been recorded, ESDA was one of few

approaches that could be used. This was not exploratory data analysis in its true form since the data analysis was being used to assess existing issues, however it was exploratory in that other usability issues were discovered in the course of the analysis.

The tapes were viewed several times and observable data such as eye movements, speech, and the motion of the robotic arm were listed and then recorded in a graphical format at differing grains of analysis (see figure 7-1 below for an example of one of the graphs produced). Issues and observations directly arising from each tape were also noted. There was no need for statistical analysis to be performed since the data set was so small and variable, with the different tasks and users.

This video data was filmed to provide evidence of the functionality and range of movement of the robotic arm. As such, the original task of switching on a light switch could not be examined, which made the examination of the usability issues surrounding that task difficult to determine. The reliability of the listings and graphs must also be carefully considered since, due to the way in which the video data was recorded, some eye movements could not be observed, and other sounds and small movements were not distinguishable. There is also the problem of observability, or “the distance between what is literally seen and heard and what its significance is for the theoretical framework and question at hand” (Sanderson & Fisher, 1994b) to be taken into account.

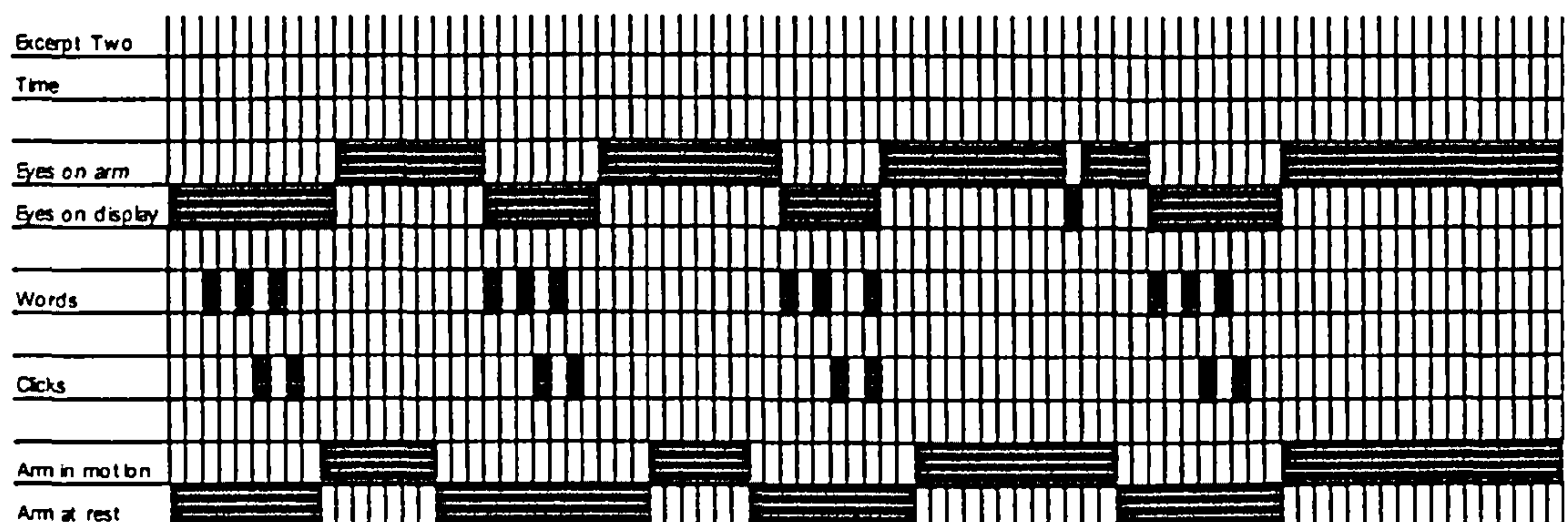


Figure 7-1: Graphical representation of Excerpt Two

### 7.2.1 The users

Of the six excerpts examined, the arm user was the developer of the arm, Bernard Parsons, in excerpts one, two and three. These excerpts are demonstrations of the arm’s capabilities



for movement, using options selected directly from the display and the speech and gesture input devices. Only pre-taught positions are used, and the user was the only person present on these three occasions.

In excerpts four, five and six the arm user was Paul Rocca, who has suffered a previous spinal injury, and lacks full movement, strength and dexterity in his limbs. Excerpts four and five were filmed during Mr Rocca's first session using the arm, and show the gesture and voice input devices being used to control the arm for the feeding exercise demonstrated in excerpts two and three using pre-taught positions. Excerpt six was filmed during Mr Rocca's second session using the arm and uses a combination of pre-taught positions and manual adjustments, by means of mouse input, to fill a glass with water and position it so that the user can drink from it. Excerpts four and five were filmed with another person present, and conversations were held. Excerpt six was filmed with no one else present.

### **7.2.2 Comparison with previous analyses**

The listings and graphs were the basis of the examination of the video data for evidence of the usability issues identified by the previous analyses. The resulting data could be used to empirically support some usability issues, and further issues were identified. Each issue was considered in turn to see whether the video data provided any corroborating evidence. Out of the twenty-three issues examined, video evidence successfully corroborated five issues, as shown in table 7-2. The remaining issues could not be successfully identified for several reasons. The video data was not collected for the specific task of turning on a light that the previous analyses investigated but had other objectives resulting in a lower quality of data, highlighting the importance of the task to the usability evaluation. The users recorded in the footage were not representative of the intended user group, although the second user, Mr. Rocca, shared certain potential characteristics. However, two further usability issues were identified in the course of the analysis.

#### ***7.2.2.1 Successfully Identified Issues***

Video evidence was found to corroborate five of the usability issues identified by the previous analyses. Issues 12 (problems of determining left and right, especially when arm contorted) and 13 (user cannot check direction choice until arm starts to move) concerned

the direction of movement and were applicable only to excerpt six, since the other excerpts used pre-taught positions. From an examination of the video data, there were four instances where either all or part of the arm started to move in one direction, only for it to be stopped and then moved in the exact opposite direction, implying that an error had occurred.

The three issues concerned with difficulty in positioning the arm: 14 (time taken to interact with system to stop arm), 17 (mismatch between way that arm works and way that user would move arm), and 23 (difficulty of judging arm movements); were again only applicable to excerpt six. Evidence to support these issues came from examining the various under- and over-shoots found where the user has had to correct the position of the arm, implying that an error had occurred. The large number of arm movements (forty-two to accomplish this task) show the disparities between how the user would want to move the arm and how the arm can be moved. It also indicates the difficulty of judging arm movements, in that corrections and refinements had to be made so often. The problems with direction choices discussed above corroborate this issue.

#### ***7.2.2.2 Unsatisfactory video evidence***

There were conflicting numbers of operators for the differing excerpts, due to the different menu configurations used and the different input devices, making it difficult to examine issues 1 (long sequence of operators to move arm) and 4 (lack of short cuts) purely on the evidence of the video data alone.

Issues 2 (inability to backtrack), 16 (illegal options), 18 (not clear that End returns user to main menu) and 19 (End having two meanings) were only applicable to excerpt six, since the other excerpts used pre-taught positions where these issues were not relevant.

Issues 3 (difficulty of choosing between Move Arm and Move), 6 (confusion over joint called Arm), 11 (lack of feedback after Move Arm selected, no indication that whole arm is to be moved), and 15 (similarity between moving joint and moving whole arm) related to confusion over a particular joint or the arm as a whole. These again were only applicable to excerpt six, due to the use of pre-taught positions in the other tasks. Excerpt six was checked for long pauses suggesting that the user tried one option then had to try another, or that either the joint or the whole arm had been confused. From the video evidence it seems that there were no occasions to suggest this, or, if so, no corrections were apparent.



No relevant data was found to assess issues 5 (Continue versus Go: Continue seen as redundant) or 9 (if user pauses in middle of saying "Move Arm" ). Evidence to support issue 7 (gesture input with twice as many operations as voice because dependent on cursor movement) was difficult to determine due to there being only two subjects, each of whom used slightly different movement sequences when using the voice and gesture input devices in excerpts two to five, compounded by only part of the sequences being filmed. In the voice excerpts two and five the voice input was prone to error, requiring repetition, thus taking more time than the gesture inputs in excerpts three and four which apparently had no recognition problems. Although the use of voice commands negated the need to look at the display, both users did. This may have been either for reassurance of the particular command needed, or for confirmation of acceptance. Either way this added to the length of the sequences.

Instead of the head gesture system a wrist activated gesture system was used so issue 8 (problem if head moved to look at arm while gesture system operational may be interpreted as a command) could not be examined. Although the same kind of problem could be associated with the wrong movement of the wrist, no data was found in the excerpts of mistaken movement being misinterpreted as a command.

There was no specific data regarding issue 10 (if user engaged in conversation). The only conversation during voice input occurred in excerpt five. This conversation did not contain any of the command words, so the issue cannot be definitively tested. However, it was noted that the user used a different volume and tone of voice for most of the conversation, which suggests that this might not be a problem, depending on the vocal characteristics of the potential user. If a user had a limited vocabulary and range then this problem might be more likely to occur.

Issue 20 (lighting conditions) was not applicable due to the adequate lighting conditions, and issue 21 (difficult for user to move field of vision) due to both users being able to move their heads freely.

There was no significant evidence to suggest that issue 22 (user looking one way, menu options in other direction) was a significant problem, since an examination of the eye movements at the time of the only overshoot in position made in excerpt six, where the arm

moved too far and had to be moved back, shows that the eyes were mostly on the arm. Looking at the display cannot be put forward as a cause, and since this issue cannot be confirmed from the empirical evidence available, it may not be a usability issue.


### *7.2.2.3 New issues identified from the video data*

These two issues relate to different tasks than that of turning on a light switch, yet both could be applicable to that situation, indicating that the original analyses were not complete, and suggesting that empirical examination might have uncovered further usability issues.

The problem of the arm obscuring the user's view was one not previously considered, and was identified from excerpt six where, because of the more complex nature of the task, the user had to move his head significantly on two occasions: to peer under the obstruction of the robotic arm to check the water level of the cup; and to lift his head to check position of both grippers around the cup. On another occasion in excerpt six the user tried to move the dispenser lever using the arm, but the arm was not in the correct position and the user was unable to see the position of the arm's gripper due to the arm obscuring the view.

Another interesting issue was the division of user attention between the robotic arm and display. In excerpt one, which used mouse input, the eyes were on the display whenever the arm was at rest for all but one second. There was only one second when the eyes were on the display when the robotic arm was in motion. However since the arm was using a sequence of pre-taught positions the user did not actually have to look at the arm at any point other than to swallow the food. In excerpt two voice input was used to move the arm through a sequence of pre-taught positions. Theoretically there was no need to look at the display at any time, since the user was an expert who knew the commands. However, the eyes, with very few exceptions, were on the display whenever the arm was at rest, and the voice commands were only spoken when the arm was at rest, with the eyes only once on the arm when a voice command was given. This suggests that the user was not yet comfortable with the natural use of the robotic arm to perform the tasks. This is reinforced by the data for excerpt five, where voice input was also used, and the user's attention was on the display more than on the arm, and only one command spoken when the user's eyes were on the arm. This was a novice user who was less sure of the commands, but even so it suggests



Key issue 

N/A inapplicable issues UNSAT unsatisfactory video evidence

NO DATA no data found EVIDENCE video evidence found


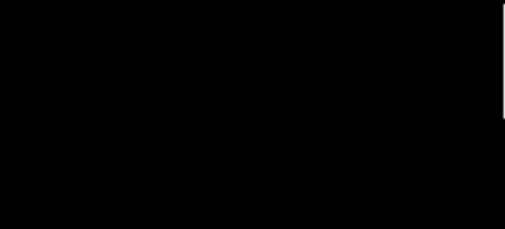





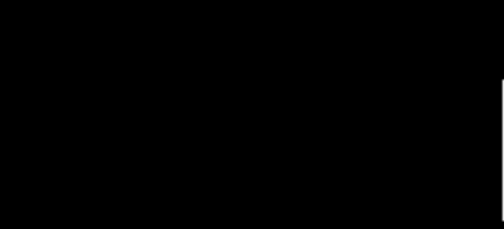
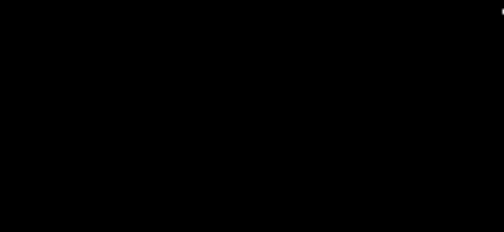
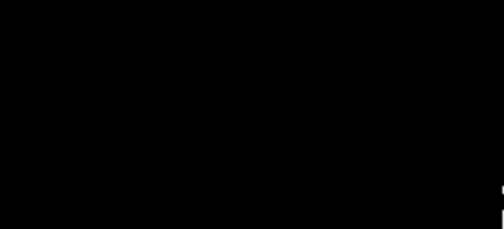


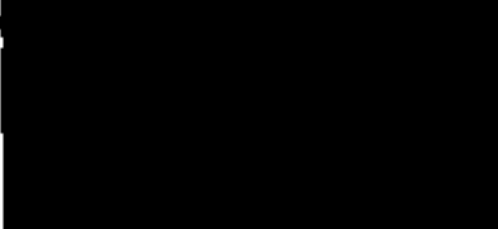
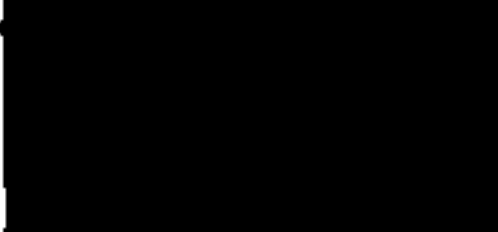

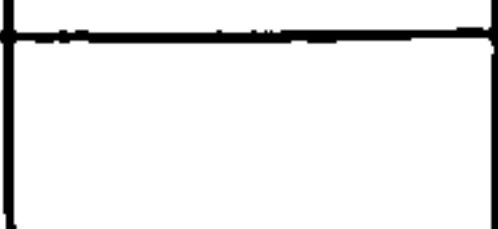

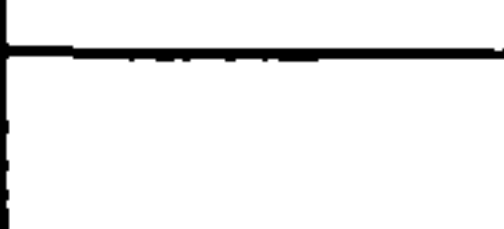
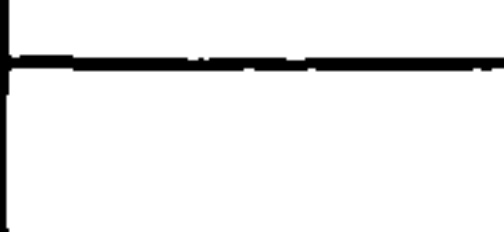
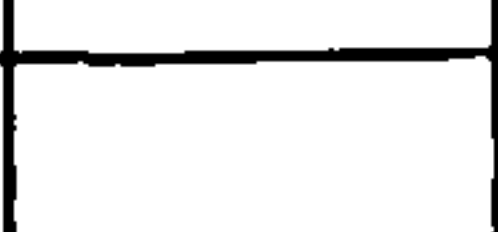
	PROBLEM	N/A	NO DATA	UNSAT	EVIDENCE
1	Long sequence of operators to move arm				
2	Inability to backtrack				
3	Difficulty of choosing between Move Arm or Move				
4	Lack of short cuts				
5	Continue versus Go: Continue seen as redundant				
6	Confusion over joint called Arm				
7	Gesture input with twice as many operations as voice because dependent on cursor movement				
8	Problem if head moved to look at arm while gesture system operational may be interpreted as a command				
9	If user pauses in middle of saying "Move arm"...				
10	If user engaged in conversation...				
11	Lack of feedback after Move Arm selected, no indication that whole arm is to be moved				
12	Problems of determining left and right, especially when arm contorted				
13	User cannot check direction choice until arm starts to move				
14	Time taken to interact with system to stop arm				
15	Similarity between moving joint and moving whole arm				
16	Illegal options				
17	Mismatch between way that arm works and way that user would move arm				
18	Not clear that End returns user to main menu				
19	End having two meanings				
20	Lighting conditions				
21	Difficulty for user to move field of vision				
22	User looking one way, menu options in other direction				
23	Difficulty of judging arm movements				

Table 7-2: Summary table of issues compared to video data

that the voice input did not allow as great a freedom from looking at the display as might be imagined.

### 7.2.3 Discussion

The limitations of the video data, due to the nature of the task performed, have been previously discussed. It is therefore unsurprising that so many of the usability issues could not be confirmed. If the same task had been used, that of turning on a light switch, and the interface had been configured in exactly the same way, then it is likely that more issues could have been examined. However, pure observation alone would not necessarily have been enough to identify issues relating to user confusion or mistakes. Using a think-aloud or co-operative evaluation would allow greater insight into the user's choices and reasoning.

ISSUE	IDENTIFIED BY
12. Problems of determining left and right, especially when arm contorted	CW
13. User cannot check direction choice until arm starts to move	CW
14. Time taken to interact with system to stop arm	CW, EMU
17. Mismatch between way that arm works and way that user would move arm	PUM, EMU
23. Difficulty of judging arm movements	EMU

**Table 7-3: Issues identified by previous approaches and confirmed by video data analysis**

That the video data was able to provide empirical evidence to confirm five issues (12,13,14,17,23) supports the ability of these kind of approaches to assess interface usability (see table 7-3). Three of these issues were originally identified by EMU, confirming the ability of the methodology to successfully identify some usability issues.

Close examination of the video data was unable to show issue 22 (user looking one way, menu options in other direction) as a problem. However in excerpt six user attention switched frequently between the arm and the display even when the arm was at rest, suggesting that the user was having to correlate information from two sources: the position



of the arm, and the display information. Thus although this issue could not be proved, it did raise further points regarding the effective correlation of information.

That two new issues were identified through the video data analysis shows that usability modelling is not a complete way of assessing interface usability. This highlights problems of coverage common to both usability modelling and empirical evaluation, since it is possible for both to miss problems or to assign them less importance than they might actually have. The identification of two new usability issues demonstrates that the six approaches were not comprehensive, although it must be remembered that a different task was being assessed. Empirical evaluation itself has problems, depending on the particular form that it takes. In this case the empirical evidence was incomplete, and using different observational evaluation methods might have been more productive. However, the combination of the two, with empirical evidence being used both to assess issues derived through usability modelling and to identify new issues, was useful, in that it allowed the analyst to focus in on areas that had been identified as potential problems.

### **7.3 Systematic re-analysis of methods**

The previous section has examined empirical evidence of the robotic arm in use in order to confirm the usability issues identified by the original analyses. However, although this can be used to determine the importance or otherwise of the issues identified, it does not give insight into whether or not those issues were identified in a valid manner, according to the actual claims of the techniques used. This section therefore re-examines the original analyses of the robotic arm, including the EMU analysis, with reference to their supporting literature in order to carefully assess the scope of the approaches, and to determine whether the usability issues were identified due to the power of the technique, the skill and knowledge of the analyst, or other factors.

The original study of the robotic arm interface was used to gain insight into the problems surrounding the usability modelling of multi-modal interfaces by formal or semi-formal approaches. Although this study was able to provide interesting insight into questions of modality representation and analysis, it was a naive study, done mainly by one analyst who was a novice in the approaches, with the individual analyses possibly being influenced by

the increased familiarity of the analyst with the portion of interface investigated. Therefore, each of the six approaches was re-examined with careful reference to specific source materials for each particular technique, to determine which issues were identified by the technique, and which were identified due to other factors. The re-analysis of STN used Dix et al (1993), the re-examination of CW used Wharton et al (1994), the investigation of the scope of the CMN GOMS and CPM GOMS analyses was based on John & Kieras (1994), the PUM study used Blandford et al (1998b), the Z re-analysis used Spivey (1989), and the re-examination of EMU was based on the description of the methodology given to the case study subjects (see Appendix F).

Since the robotic arm was not available for re-examination due to the flood damage sustained, each issue previously identified, regardless of which approach identified it, was re-examined in depth. This was to trace whether the issue should have been identified by each technique, and, if identified, whether this was due to the power of the technique or the skill and knowledge of the analyst. Obviously, this was a restricted set of issues and was unable to uncover issues outside that set which should have been identified by the approaches, for example the two issues derived from the video data analysis. Even so, the in-depth examination showed that although many issues were identified due to the approach used, many were identified through the knowledge, skill or other expertise of the analyst. Some issues which should have been identified were not, either due to the level of abstraction of the original analysis, or due to mistakes or omissions by the analyst. For the complete re-analyses, please refer to Appendix I. An example portion of the STN re-analysis is given in figure 7-2.

### **7.3.1 Scope of approaches**

The usability issues originally identified were re-examined in depth to determine which issues were within the scope of the approaches, and which issues were identified for other reasons. Each approach was able to identify some issues purely through use of the technique.

STN was able to identify issues (5,15) relating to unnecessary states for a user to navigate through, and similarity between operations. STN is specifically concerned with the states of the interface and the functional movement between those states.



---

### **1. Long sequence of operators to move arm**

Since the STN shows the number of states that the user has to navigate through before the robotic arm can be moved, this issue should have been identified in the original analysis. That it was not identified shows the extent to which the analysis was dependent on the craft skill (or lack thereof) of the analyst.

### **2. Inability to backtrack**

This issue is apparent from the STN, and was identified as a problem. However, the identification of this issue was possibly influenced by the explicit mention of this kind of problem in a discussion on “undo” in the source materials (Dix et al, 1993, p.291). This shows the effect that the source materials of a technique has on the application of a technique.

### **3. Difficulty of choosing between Move Arm or Move**

The STN concentrates on the actual choice between the system states, rather than on the difficulties the user has in choosing between them. It is therefore not an issue that the STN on its own would be expected to identify, but might have been identified through craft skill.

### **4. Lack of short cuts**

Since the STN explicitly shows the possible path of the interaction through the various states, the lack of short-cuts was an issue that might have become apparent if the analyst had been looking for it. This is therefore an issue that is a combination of craft skill and representation. That it was not noticed was possibly because the analyst's attention was more on obtaining the correct representation of the system states.

### **5. Continue versus Go: Continue seen as redundant**

This was an issue found through drawing the STN, since the use of Continue creates more states for a user to navigate through, and adds little to the functionality of the interface. This is the kind of issue that the use of the STN should make apparent, and did.

### **6. Confusion over joint called Arm**

The STN did not go into the detail of the individual options, so this issue did not arise. If the STN had been done at a different level of abstraction, this issue might have been identified through the craft skill of the analyst. It is not something that the STN would identify however, since it is concerned more with the user understanding of what a particular option choice means rather than with the option choice itself. Thus this issue highlights questions associated with both craft skill and appropriate levels of abstraction.

---

## **Figure 7-2: Portion of STN Re-analysis**

Four of the issues identified using CW (3,6,18,19) concerned misleading option names, and one (11) concerned the lack of feedback. CW explicitly supports the “label-following strategy” used by most users and progress towards user goals through option availability.

The original CPM GOMS analysis was unable to identify any issues other than the difference between the use of voice and gesture operators (issue 7), which is within the bounds of the method since CPM GOMS supports the identification of the critical path of interaction. In the CMN GOMS re-analysis, one issue (7) concerned the number of operators, two issues (1,4) concerned the sequence of operators and lack of short cuts, and two other issues (5,15) concerned the similarity of options. CMN GOMS supports the identification of operators and their sequences.

Four issues identified by PUM (3,6,18,19) concerned misleading option names, and one (17) concerned the mismatch between the way that the robotic arm works, and how a user would move an arm. PUM supports the identification of knowledge needed by a user, and the difference between user and device tasks.

Two of the issues identified by Z (5,15) related to similarity of operations, and were identified by the schemas having identical or nearly identical contents. Issue 16 (illegal options) was identified since the Z specification involves examining in detail each operation and what is allowable for any given state of the interface.


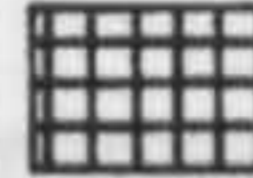
Four of the issues identified by EMU (3,6,22,23) concerned clashes. Three issues (8,20,21) were identified through the comparison of user, system and environmental profiles with the modality listings. EMU explicitly supports the identification of clashes, and the identification of issues arising from incompatibilities of profiles and listings.



Four issues identified by PUM (3,6,18,19) concerned misleading option names, and one (17) concerned the mismatch between the way that the robotic arm works, and how a user would move an arm. PUM supports the identification of knowledge needed by a user, and the difference between user and device tasks.



Two of the issues identified by Z (5,15) related to similarity of operations, and were identified by the schemas having identical or nearly identical contents. Issue 16 (illegal options) was identified since the Z specification involves examining in detail each operation and what is allowable for any given state of the interface.









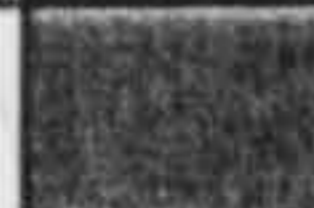















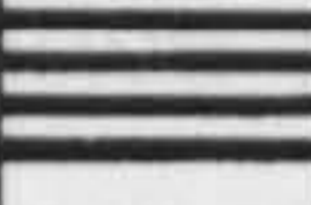





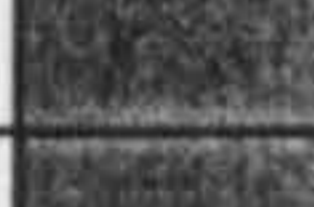
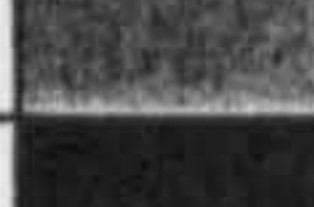

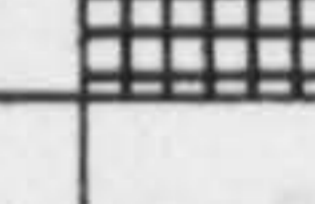


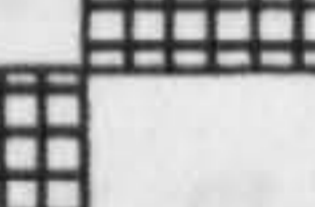


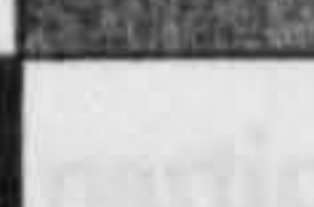

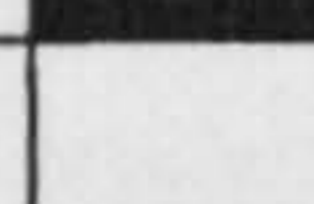

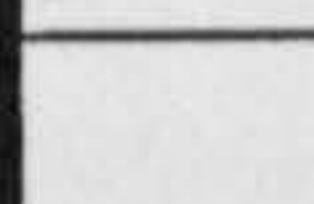
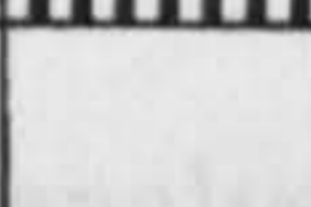


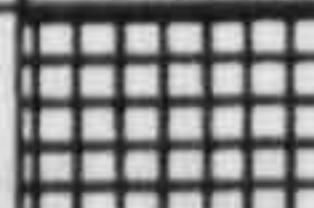

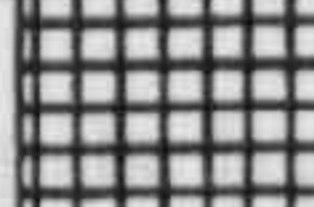

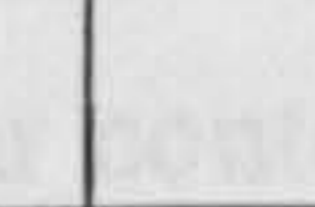





**Key**

Issues identified due to Method:  Issues that should have been identified: 

Issues identified due to Craft Skill:  Level of abstraction issues: 

Issues identified due to Source Materials:  Other: 

CMN-GOMS and CPM-GOMS are included under the heading of GOMS

PROBLEM	STN	CW	GOMS	PUM	Z	EMU
1 Long sequence of operators to move arm						
2 Inability to backtrack						
3 Difficulty of choosing between Move Arm or Move						
4 Lack of short cuts						
5 Continue versus Go: Continue seen as redundant						
6 Confusion over joint called Arm						
7 Gesture input with twice as many operations as voice because dependent on cursor movement						
8 Problem if head moved to look at arm while gesture system operational may be interpreted as a command						
9 If user pauses in middle of saying "Move arm"...						
10 If user engaged in conversation...						
11 Lack of feedback after Move Arm selected, no indication that whole arm is to be moved						
12 Problems of determining left and right, especially when arm contorted						
13 User cannot check direction choice until arm starts to move						
14 Time taken to interact with system to stop arm						
15 Similarity between moving joint and moving whole arm						
16 Illegal options						
17 Mismatch between way that arm works and way that user would move arm						
18 Not clear that End returns user to main menu						
19 End having two meanings						
20 Lighting conditions						
21 Difficulty for user to move field of vision						
22 User looking one way, menu options in other direction						
23 Difficulty of judging arm movements						

**Table 7-4: Table showing issues after re-analysis**



Four of the issues identified by EMU (3,6,22,23) concerned clashes. Three issues (8,20,21) were identified through the comparison of user, system and environmental profiles with the modality listings. EMU explicitly supports the identification of clashes, and the identification of issues arising from incompatibilities of profiles and listings.

In the six analyses, out of a total of 43 identifications, 27 were identified through the particular technique used (see table 7-4), showing the benefits of using formal and semi-formal approaches in evaluating usability. However, that 16 issues were identified by other means shows that the role of other factors is also important and must be considered.

Indeed, one technique, CW, does not actually support all that it aims to do. For example, issue 17 concerns the mismatch between the way that the arm works and the way that a user would move the arm. One of the explicit aims of CW is to find “mismatches between users’ and designers’ conceptualisation of a task” (Wharton et al, 1994). However, despite there being a distinct conceptual gap between how the user would move the arm and how the design of the robotic arm constrains movement, this issue cannot be identified by CW because of the method’s concentration on the step-by-step nature of the task rather than considering the task overall.

One issue (2: inability to backtrack) identified by the STN analysis was found to be outside the direct scope of this technique and was influenced by the explicit mention of this kind of problem in a discussion on “undo” in the source materials (Dix et al, 1993, p.291), demonstrating the effect that the source materials can have on the application of a particular technique.

The seven issues originally identified in the CW analysis that were found to be beyond the boundaries of the method related to the long sequence of operators needed to use the arm (1), the time taken to interact (14), the possible wrong interpretation of commands by the system (8,9,10), and the effects of an option choice in the particular context of use (12,13). Although CW does not examine the long sequences of operators since it is task-based and examines each step one at a time, rather than as a whole, the method does require a list of the task sequence needed to successfully accomplish the task, from which issues related to the length of the interaction can be identified through the expertise of the analyst. The skill of the analyst can also be used in examining issues relating to the system correctly



understanding what the user wishes to accomplish, although CW places its emphasis on the correct interpretation of system actions and options by the user. The failure stories can also be expanded by the analyst to include all conceivable problems rather than just those concerning the lack of a credible success story.

The three issues identified in the CMN GOMS analysis through the analyst's skill related to the inability to backtrack (2), and confusion regarding option names (3,6). The inability to backtrack was not identified by CMN GOMS because although it considers the task sequence in some detail, it only considers limited error recovery. Issues 3 and 6 were not identified by CMN GOMS because the method does not support the identification of issues relating to problems of choosing between options.

Issue (2) was found to be beyond PUM's scope, in that it related to the lack of backtracking opportunities, and was derived from the representation of the interaction, with its emphasis on the heavy ordering of operations. The same issue was identified from the Z specification due to the strongly ordered nature of the specification.

The three issues found to be beyond the boundaries of EMU and identified through the knowledge and skill of the analyst related to the long sequence of operators needed to use the arm (1), the time taken to interact with the system to stop the arm (14), and the mismatch between the way that the arm works and how the user might use the arm (17). EMU does not specifically examine the length of the interaction, but the written interaction sequence produced through following the method is amenable to this kind of analysis. This interaction sequence can also be used to examine issues such as (14). The interaction sequence represented the modalities used for this command in detail, and was able to show that both the gesture and voice commands would use the same number of modalities. EMU is also able to examine specific critical sequences of modalities, whose timing properties can then be investigated. With regards to the mismatch between the arm and the user, stage one of the EMU methodology concentrates attention on the task selected, which would encourage this kind of issue to be identified through the skill of the analyst.

### 7.3.2 Role of expertise

All the techniques (except STN) supported the analyst in identifying other issues. A study of this size and limitations cannot attempt to delve deeply into why this may be so, but a certain amount of speculation is permissible. It is possible that some of the original analyses were contaminated, in that as the analyst gained increasing familiarity with the interface and the possible usability issues, the analyst was able to read more into the techniques than was actually there. It may also be due to what the analyst was bringing to the use of the technique from outside the strict scope of the technique.

Technique	World
Knowledge (of how technique works and how to apply it correctly)	Knowledge (of wider usability potential problems and the effect they might have)
Skill (in applying technique properly and identifying issues)	Skill (in identifying issues based on this knowledge)

**Figure 7-3: Knowledge and skill used in technique and from world**

When using a technique, there is a certain amount of knowledge and skill that are needed to use the technique appropriately, just as the analyst will have other knowledge and skill from “the outside world”, as illustrated in figure 7-3.

In order to successfully apply a technique, an analyst must not only know how the technique works, how to apply it, and what to look for, but also have skill in applying the technique correctly, and in identifying those issues which are relevant. The analyst will also bring knowledge of other usability factors and skill in identifying such potential issues when using the approach. This follows on from how expertise is believed to be developed. Dreyfus and Dreyfus (1985) propose five stages in the development of expertise, from a novice stage where rules are learnt and applied for manipulating context-free elements, to advanced beginner who begins to understand the domain and see meaningful aspects, to competent performer, who learns to set goals and interpret the current situation in terms of what is relevant to achieving those goals, to proficient performer who views a situation as



having a certain significance tending towards a certain outcome and aspects of the situation stand out as salient in relation to that outcome, to the expert, who is able not only to perceive the situation but is able to rapidly generate appropriate solutions. Klein (1998) has examined the development of experience and expertise in depth, and suggests that experienced people rely on pattern-matching to a large degree, where they identify familiar elements in a situation, and that expertise can be developed by having many experiences, as well as quick and accurate feedback, and time to reflect and learn from the experiences.

In the hands of different analysts, exactly the same techniques will produce different results, due to novices not being able to apply the technique correctly and identify everything, and experts able to use the technique as a lever to gain awareness of many other potential issues. Nielsen (1992) has explored this in relation to Heuristic evaluation. As (Dreyfus, 1992) states: "It seems that when a person has enough experience to make him or her an expert in any domain, the fields of experiences becomes structured so that one directly experiences which events and things are relevant and how they are relevant. ...objects appear to an involved participant not in isolation and with context-free properties but as things that solicit responses by their significance." This supports the pattern-matching strategy proposed by (Klein, 1998) above. This suggests that the value of a technique is not only in what it is able to represent, but also in what it can contribute to an experienced analyst. The investigation of the re-analyses found that this was indeed the case, and that the analyst was able to identify other issues through the process of applying the techniques. This can explain why the Z representation could be used to reason about aspects of usability beyond what might be expected, since the analyst was using world knowledge and skill to identify interesting usability issues. This leads into deep questions regarding expertise and the use of representations to guide usability engineering. Humans do not use facts and rules to guide how they think, since their "background knowledge consists largely of skills for dealing with things and people rather than facts about them" (Dreyfus, 1992).

### **7.3.3 Accuracy and completeness**

One proposed advantage of a formal or semi-formal approach is that it provides an accurate representation, and that this accuracy allows many issues to be identified. However, obtaining an accurate representation is far from straightforward. Mistakes in the

representation can lead to the omission of important problems and the wrong identification of problems.

This was corroborated by re-examining which issues were not identified by the original analyses. Issue 1 (long sequence of operators to move arm) was not identified by the STN analysis, despite being within the scope of STN and with the diagram clearly showing the large number of states that the user has to navigate through before the robotic arm can be moved. Issue 14 (time taken to interact with system to stop arm) should have been identified by CMN GOMS: although CMN GOMS identified the number of operators for both voice and gesture, and found that gesture had twice as many as voice, the implications for this with regards to the stopping of the arm were not considered. Two issues: 11 (lack of feedback after Move Arm selected) and 13 (user cannot check direction choice until arm starts to move); were not identified in the original PUM analysis. Although there was no output implemented for the robotic arm at this point, the method should still have identified these issues since PUM traces what information users need and where it comes from. Issue 6 (confusion over joint called Arm) should have been identified in the Z analysis when the type ARMPART was declared, since the whole arm had to be called “wholearm” rather than “arm”, because there was already a joint called “arm”. Issues 18 (not clear that End returns user to main menu) and 19 (End having two meanings), both potential mismatches, should have been identified by EMU but were omitted.

These omissions demonstrate how the identification of any issue is dependent upon the skill and expertise of the analyst, and that mistakes and omissions can occur. As discussed earlier, the use of a technique is dependent not only on the knowledge of the analyst but also on their skill and expertise in applying it. This can only be gained through experience, which places more emphasis on appropriate training of analysts and how analysts can gain and make good use of that experience. However, one of the popular conceptions of formal or semi-formal approaches is that they minimise the need for training. That issues were not identified when they should have been shows that this is not the case, and that reliance on these approaches without taking into consideration the development of skill in applying them can lead to significant issues being omitted from the analysis, as mentioned by Scapin & Basien, (1997). This issue was highlighted by Gray & Salzman (1998) in their review of usability evaluation methods, where one of the factors identified as a cause for concern was



that of causal construct validity: were the experimenters actually using the usability evaluation method they claimed to be using, or was it poorly defined and reliant more on the skill of the analyst to achieve the results than on clearly defined procedures?

There was also a problem of incomplete representation with the STN diagram. On re-examination of issue 16 (illegal options) it was found that the STN diagram had no state showing that the arm had reached its limit of movement, nor was there an end option leading from the 'travel until stop' state.

The level of detail used is another factor that is determined by skill, and is important since it may obscure issues by being too low or high for issues to be apparent. For example, the STN was not written at a level of detail which could identify issue 7 (gesture input with twice as many operations as voice because dependent on cursor movement). A more detailed diagram would have shown the gesture input as a linear series of states with transitions between them, to indicate the movement of the cursor beneath the options. Voice input would have been drawn as only one state with multiple transitions coming from it, since it is not reliant on the cursor moving beneath the options.

With the PUM analysis, issue 1 (long sequence of operators to move arm) was mentioned in passing in the original analysis, but not strongly enough for it to be identified as a usability issue. Issue 7 (gesture input with twice as many operations as voice because dependent on cursor movement) was not identified in the original analysis since it was not written at a low enough level of detail. If the precondition that the cursor had to be under the correct option before it could be selected by the user had been written then the differing number of preconditions for the two input devices would have been recognised and this issue identified. The same issue was not identified by the Z analysis because the specification was not written at a low enough level of detail to represent the cursor movement.

#### **7.3.4 Representation form**

The notation or representation used by a particular usability modelling technique has a major impact on its use. Not only does it constrain what can be described, but it is an additional burden on the analyst in developing expertise which is often overlooked. Emphasis is more often placed on identifying problems, and correctly modelling the

interface, than on issues concerned with the correct use of the notation. The effect of developing expertise in the representation and its effects on the analysis is an area that is poorly understood.

Some notations are easier to learn and use than others, which can affect the amount of time and effort spent by the analyst on obtaining the representation as opposed to then examining that representation for usability issues. As previous research has discussed (see Hall, 1990) the process of creating a representation can be as useful in finding usability issues as the final artefact. However, if a notation or representation is very formal or very complex, until a high degree of expertise is achieved, a large amount of effort must be focused on the representation, rather than on considering usability problems that may lie outside the direct scope of a particular technique. By contrast a notation that is simple and requires less effort may release attention which can then identify important potential usability problems tangential to the scope of the method.

Thus it may be that the form of the technique can influence the number of additional issues that can be uncovered through the skill of the analyst. Although only a small study with limitations, the re-analyses of the six techniques demonstrated that the CW analysis had a large proportion of issues identified due to craft skill in addition to those issues which fell within its scope, whereas the GOMS and PUM analyses identified fewer. CW has a loose form of representation, allowing the analyst freedom to concentrate on usability issues rather than on correctly representing the interaction. By contrast, GOMS and PUM both have a strict representation which strongly constrains the analyst, and the amount of effort to derive the representation is greater. The analyst thereby concentrates more on the correct form of the representation than on observing usability issues, and it may be that the craft skill of the analyst therefore has less opportunity to emerge.

EMU combines the looseness of CW with a precise representation of the interaction. The case study reported in a previous chapter illustrated how subjects concentrated so hard on obtaining the correct representation that they were unable to gain much insight into the usability issues derived from it. Therefore it may be that by constraining analysts to a particular type of representation understanding is blocked and the analyst concentrates more



on creating a representation which appears to be correct rather than one from which understanding can be gained.

However, looseness of representation, although it may help with promoting the application of craft skill, can lead to more problems with analysts becoming uncertain as to the actual limits of the notation and how it should be used correctly, as Blandford et al (1998) observe in a study of training software engineers in PUM. This tension is shown in the development of CW, which has iterated through various cycles in an attempt to provide a simple and flexible notation to aid the analyst and still remain true to the underlying cognitive theory.

### **7.3.5 Discussion**

The original aim of the rational re-analyses was to try and distinguish between issues identified through craft skill based on general HCI expertise and familiarity with the problem, and issues identified due to the core of the method for each approach. This was to gain a realistic understanding of the validity of claims made by each method, and to identify how else issues might be identified, and the consequent implications.

It was apparent that not all issues were identified strictly by the use of a particular method. Out of forty-three original issues, sixteen were identified due to other reasons, demonstrating that craft skill, whether consciously used or otherwise, plays an extremely important role in usability analysis. Issues were identified through source materials explicitly discussing that kind of issue, the kind of artefact produced (for example, the CW and EMU listings which enabled issues such as those relating to length to be identified), and the craft skill or existing usability expertise of the analyst. Not taking into account such factors means that a method's claims can be inflated, and actual issues supposed to be identified by it are actually identified for other reasons, giving a misleading impression of its scope. The final scope of a method can be wider than its "official" scope due to the expertise of the analyst.

Using a method is not straightforward, but relies on the skill and knowledge of the technique and also of the wider usability possibilities that may be relevant. Developing such expertise takes time, as the analyst progresses through several stages. If this is not taken into account, then novices will find that methods cannot be used to their full potential

immediately. There will be a large difference between analyses done of the same interface depending on whether the analyst is a novice or expert in the technique, and in identifying usability problems generally. The implications are that differences between novices and experts must be explicitly recognised by both method developers and interface developers, since they will identify different problems and therefore cannot guarantee complete coverage. This problem is explicitly recognised by (Cockton & Woolrych, 2001), who propose better HCI education for analysts to aid effective usability evaluation.

The amount of expertise necessary to correctly apply a technique may lead to issues being omitted, incomplete representations, and representations produced at inappropriate levels of detail. The tension between accuracy and completeness places stress on the analyst who then misses issues.

The form of representation used not only constrains the types of issues which can be identified, but also how much practise is needed in learning how to use it. There is an expertise component to obtaining 'good' representations that give leverage on 'real' usability problems. This is unavoidable, in that there appears no alternative to gaining a good representation other than through skill, recognition, practise and appropriate feedback (Klein 1998). Attention which is focused on applying the representation may miss issues outside the direct scope of the method. However using a loose representation may detract from the underlying theory of the method, as demonstrated with the problems surrounding the development of CW.

## **7.4 Summary**

This chapter has closely examined the usability issues identified to determine their source and validity. Empirical data of the robotic arm in use was used to verify the usability issues uncovered by EMU and the five approaches discussed in previous chapters. The limitations of the video data available meant that it was only possible to verify some of the issues. Two new issues were identified through examination of the video data, highlighting the problem of coverage of any usability modelling evaluation, in that no single technique is fully comprehensive.



The systematic re-analysis of the scope of the six techniques raised further questions regarding whether the original issues were identified through the power of the technique or the craft skill of the analyst. A significant number of usability issues were identified due to reasons other than the strict scope of the techniques. The development of expertise was discussed in relation to the role of craft skill, and it was suggested that the analyst was able to make use of the technique to stimulate the identification of further issues. The importance of analyst expertise was discussed further in relation to the accuracy and completeness of evaluations and the role of representation.

## **8. IMPLICATIONS FOR EMU**

### **8.1 Introduction**

Methodologies do not spring fully-formed from the minds of the developers, but iterate as alterations are made in the light of practitioner experience (Blandford et al, 1998a). Once a methodology has been designed it needs to be examined in use in order that issues of usefulness, in terms of what it can actually do, and usability, in terms of how easily it can be used, can be examined. The usability of EMU was investigated by a case study examining how subjects used EMU to analyse two interactive interfaces. The case study involved computer science students applying EMU to two multi-modal interfaces, in order to ground claims that the modality definition and taxonomy were usable, and that the methodology enabled usability issues to be identified. By using non-originators it was possible to examine problems involving the practical application of the methodology. The usefulness of EMU was considered in its application to the robotic arm interface, and again in the rational re-analysis of the robotic arm. This chapter assesses these findings and examines the implications for the improvement of EMU, in terms of its scope, content, representation and length.

### **8.2 Scope**

The re-analysis of EMU has implications for the scope of the methodology. Of the ten usability issues identified by EMU relating to the robotic arm interface seven were within the strict boundaries of EMU. This demonstrates that the methodology has clear objectives and is capable of identifying usability issues. That this is not due solely to the expertise of the analyst is reinforced by the case study where subjects were also able to identify usability issues through the use of EMU. However, both the case study and the re-analysis of the robotic arm show that some issues were identified by means other than the strict use of EMU. It is important to examine what was identified to see whether these issues should or could be included within the specific scope of EMU.



Of the three issues identified through the craft skill of the analyst in the analysis of the robotic arm interface, two issues (issues 1 and 17) could be included within the scope of EMU without significantly adding to the time or difficulty of using EMU. Issue 1 (long sequence of operators needed to use the arm) was identified through examining the interaction sequence produced in stage five of EMU. Since this sequence is already explicitly part of the methodology and thus requires no extra effort to produce, placing the examination of its length as part of the methodology at this or a later stage would draw the analyst's attention to this issue, and enable greater insight using this existing material.

The difference between how a user might perform a task, and how the task is actually completed using an interface, as exemplified in issue 17 (mismatch between the way that the arm works and how the user might use the arm), could be highlighted by the methodology. The task could be considered on this basis in stage one, where the task is defined and justified. This would necessitate there being models of the expected goal and activity structures to make this explicit.

However, the third issue, (issue 14: time taken to interact with the system to stop the arm), which would involve the intense examination of each part of the sequence, would place heavy time overheads on the methodology with no guarantee of results. Therefore it should not be included explicitly. This may be something that could be delegated to another method such as CPM-GOMS.

The case study of computer science students using the EMU methodology found that many of the usability issues identified by the subjects were identified by the craft skill of the subjects rather than through the use of EMU.

	Total number of issues raised in Stage 8	Issues raised in Stage 8 and also mentioned earlier	Issues raised in Stage 8 and not mentioned earlier	Uncertain as to mentioned earlier	Mentioned earlier but not included in Stage 8
Tube ticket machine task	91	32	59	0	44
Heating control panel task	114	22	82	10	41

**Table 8-1: Summary table of derivation of usability issues in case study**

As table 8-1 shows, out of 91 usability issues identified for the tube ticket machine task, only 32 could be traced explicitly to the use of EMU. 59 were issues that had no relation to EMU, and were derived from the outside knowledge of the subjects. The usability issues identified for the heating control panel task follow a similar pattern.

This is hardly surprising, given that these subjects were still very much in the novice stage of using the methodology, whereas they all had previous experience of HCI so were able to draw on their wider knowledge to identify possible areas of concern. Yet there seemed to be confusion as to what actually constituted a usability issue. The supporting documentation did not specifically state what a usability problem might be, or how to identify it. This is a difficult issue, since what is a usability problem in one situation may not be a problem elsewhere, due to the different context of use. Therefore an explicit description is not possible. A better approach may be to examine the context of use and establish particular parameters against which the system could be assessed (Cockton & Lavery, 1999), making use of the relevant ISO standards (Fitzpatrick & Higgins, 1998). In addition, some kind of standard format for reporting usability problems at every stage would make the recording of these issues more proceduralised (Lavery & Cockton, 1997).

### **8.3 Omissions**

EMU has been specifically developed to handle issues relating to multi-modal interaction, such as the identification of modalities and their types, modality properties, and clashes caused by cognitive processing constraints. Since these areas are fundamental to the scope of the methodology, any problems are a cause for concern.

The rational re-analysis of the EMU analysis of the robotic arm interface found two issues (18 and 19, relating to potential mismatches) that were within the scope of the methodology but had not been originally identified. Mismatches occur when the expressed modality is incorrectly received, or when a modality is received that was not intended. The user or the system interprets information one way when the information actually means something else. Mismatches can seriously affect the interaction by distorting the information flow through omission or addition. They should be identified at stage six of EMU, where modality properties are identified from the interaction sequence listing. Mismatches are therefore a



fundamental aspect of multi-modal interaction, and their identification crucial to the usability of a given interface. That an expert user, the actual developer of EMU, did not identify these two issues demonstrates that expert users can make omissions, and that the methodology requires more support in ensuring that all aspects are considered.

The case-study revealed that the subjects found identifying modality properties and clashes difficult, with very few occurrences identified, and eight reports not mentioning any, despite ample opportunity. The identification of clashes and properties in stage six of EMU is a critical part of the methodology, since it is this section which identifies potential problems relating to the combination and use of modalities.

The second questionnaire revealed that subjects were unhappy with the use of clashes and properties. However, that those clashes and properties that were identified were mostly identified correctly shows that although subjects were reluctant to identify properties and clashes, and were not confident about having done so correctly, they were able to do so correctly when they chose to. This implies that it is not a problem with the definition of clashes and properties so much as one relating to the development of expertise and confidence in novice users.

Since the development of expertise follows a series of stages of increasing competence, there is a need for clearer guidelines and explanations with regards to what clashes and properties are, why they are important, and how they might be identified, so that analysts are supported in the development of expertise in this area. The effective production of such material would require further testing and examination.

## **8.4 Representation**

The notation used in the EMU methodology appears to be successful, in that the case study subjects made few errors. This is extremely promising given that this was their first opportunity to use it, and made use of a notation that was unfamiliar. In the light of this, no alterations are considered necessary.

The case study subjects were generally able to identify modalities in terms of their component parts. They made occasional errors in distinguishing between visual and haptic modalities from the system's perspective, finding it difficult not to take the user's

perspective when describing the modalities involved in the operation of the system. There were also minor errors with the correct representation of modalities for button-pressing. Given the complexity of the concepts and the novice status of the users, this is very reassuring, and all that is required in the methodology to address this is the provision of a selection of typical examples of user and system modalities in the supporting documentation to clearly demonstrate the differences.

The subjects found distinguishing between receptive and expressive modalities difficult, and frequently omitted receptive modalities. The supporting documentation needs to make the differences between the types clearer, most easily through the use of typical examples, and provide some way of reminding the analyst to include receptive modalities. This could be achieved through some form of automation which prompts for a receptive modality after each expressive modality. However, this is not ideal, since the existence of an expressive modality does not guarantee an equivalent receptive modality.

Another area that needs improvement is in the use of preconditions, which subjects found confusing. Emphasis needs to be placed on developing the flow of the interaction through the use of EMU, demonstrating the causal relationships between modalities. More information about how to write the interaction listings may help, perhaps by breaking it down into several stages, identifying blocks of interaction, and writing them in natural language first, before trying to identify them in terms of modalities. Again, the use of types for preconditions needs improving, probably through the provision of examples.

Placing more emphasis on the interaction flow as a precursor to the identification of modalities would also assist the use of AND and OR. In the case study many subjects did not use them. However it is essential to EMU that concurrency and choice between modalities, fundamental aspects of multi-modal usability, be represented adequately.

The robotic arm analysis using the methodology revealed two aspects that EMU is unable to adequately represent at this time: the effects of system-based changes on the system; and transitions, or the start, stop and absence of movement. Although the methodology seeks to represent both the user and the system, more emphasis is currently placed upon the interaction between the user and the system, with no way of adequately modelling the effect of system-based changes based on non-user actions such as the arm stopping automatically



when reaching the limit of movement. The representation does not support the issue of transition, showing when modalities actually start and stop. In the section of the interaction sequence listing detailing the stopping of the arm it was necessary to include a modality for the arm stopping as a modality showing the arm's absence of movement. The need to mark the start and end points of modalities had not previously been considered.

These are important concepts that need to be addressed, since without them the interaction listing is limited in the range of information about the interaction displayed. However, these issues are complex and cannot be incorporated into EMU without significant further research.

## **8.5 Length**

The length of the methodology can be considered in terms of both the time taken to complete it, and the number of component stages. The length of time taken to use EMU cannot be accurately quantified, since it depends on variables such as the expertise of the analyst, the particular interaction under assessment, and the level of detail examined. However, the experience of the case study found that using EMU for a whole interaction involved a considerable amount of time and effort, with some stages done poorly or not at all. The number of different stages to the methodology, and the necessity for producing a complete and accurate interaction sequence, places a high demand on the analyst. As noted by (Buckingham Shum & Hammond, 1994) there comes a point when there will be a trade-off between comprehensiveness and unnecessary detail.

Since the main reason for the development of EMU is to specifically address multi-modal usability issues, one way of reducing the load on the analyst and make the methodology quicker and more productive is to concentrate on segments of interaction in considerable detail. By such selective application of EMU to micro-tasks at a low level of granularity, issues that cannot be considered by other modelling techniques can be considered. No single methodology is complete, and it is unrealistic for EMU to attempt this. However, such concentration of effort would undoubtedly pay dividends in assessing the strictly multi-modal usability issues of a critical stage of an interaction.

## 8.6 Discussion

EMU has been developed to address issues of multi-modal usability. Thus any changes should be made with an eye to ensuring that it fulfils its original purpose, rather than attempting to become an all-encompassing methodology which consequently does nothing effectively. Improvements must not be made at the expense of scope.

The support of the development of expertise in EMU by novice users is something that should be considered most carefully. It is reassuring that the novice users studied were able to identify modalities correctly, and use the notation without significant error. However, their problems with types, preconditions, AND/OR, properties and clashes, are a more serious concern, and suggests that the supporting documentation needs significant improvement in these areas, to support the development of appropriate levels of expertise. These are at the heart of EMU as a methodology for identifying multi-modal usability issues. The flow of interaction, the identification of clashes, the importance of modelling concurrency and choice, and possible instances of redundancy, must as a minimum be able to be accurately represented if the methodology is to have substantial benefits. The correct use of preconditions is of less importance, but still adds to the overall understanding of the flow of interaction. The use of worked examples, preferably dealing with many different forms of interaction, would give the novice user confidence in how to apply the technique and how to recognise potential issues. The use of a thresholding approach to the identification of usability issues, using parameters based on the system's context of use (Cockton & Lavery, 1999), would make it easier to be sure which problems identified were of most importance, and allow the analyst to focus attention on major issues rather than be side-tracked by minor concerns.

The length of the methodology is another area of concern. Rather than attempting to deal with all aspects relating to the modelling of an interaction, concentration on smaller segments of interaction may be more fruitful, with other methods used as appropriate. The analyst could be more selective about which parts of the methodology are applied, with more emphasis being placed on the interaction sequence listing and the identification of clashes and properties. This would reduce the overheads of time and effort, and allow close



examination of those areas of interaction where multi-modal usability issues are most critical.

## **8.7 Summary**

This chapter has assessed the scope, content, representation and length of EMU in the light of the rational re-analysis of the methodology and the use of EMU by non-practitioners. Some changes to EMU have been proposed, relating to assisting the development of expertise by users, particularly through changes in the supporting documentation, and the length of the methodology.

# **9. CONCLUSIONS**

## **9.1 Introduction**

This research has argued that there is a lack of theoretical understanding of multi-modal usability, and a corresponding lack of appropriate ways of applying such theory. It has also examined issues of applying usability modelling techniques. This chapter brings these strands of research together. The impact of the research overall is assessed with regards to the original aims, both to examine the strengths and weaknesses of the research, and to determine the resulting contribution to knowledge. This is then related to current HCI research in order to provide a relevant context. Areas where the research can be expanded and possible future work are examined and discussed.

## **9.2 Thesis summary**

There were two original research aims: the development of theory for multi-modal usability; and the application of such theory. Resulting from the work undertaken in this thesis, a new area of interest was identified, relating to the scope and use of usability modelling techniques. This section summarises the research undertaken for each aim.

### **9.2.1 Multi-modal theory**

The main aim of this research was to address the lack of theoretical understanding of multi-modal usability. This has been achieved by identifying and clarifying those theoretical issues relevant to multi-modal usability, and through developing a new definition and associated taxonomy.

Chapter two identified the lack of theoretical understanding of multi-modal usability as a cause for concern. A case study, involving the application of five modelling approaches (STN, CW, GOMS, PUM, Z) to a portion of an interface of a robotic arm, confirmed that existing approaches are unable to adequately represent issues of multi-modal usability. This study identified the need to clearly identify and represent modalities, requiring a coherent, consistent and useful definition of modality; and



modality constraints, in terms of timing, combination and ordering, necessitating a clear understanding of the way in which modalities interact.

Existing multi-modal research was critically assessed in chapter three. By examining definitions of modality, based around elements of sense, device, mode, form, and combinations of these elements, their limitations were exposed and discussed. The requirements for a definition of modality within an HCI context were investigated, and the need for a new definition, using the human senses in combination with other elements, established. Five different approaches to the usability of multi-modal and multi-media systems were examined and assessed. Four areas of critical relevance to multi-modal usability were identified: identification and granularity; user and system; load and combination; and time and ordering. To understand the mental and physical capabilities of the user in relation to multi-modal usability four cognitive models were examined with regards to the four issues identified above, resulting in the derivation of several cognitive issues to be addressed by a theory of multi-modal usability.

In chapter four the original aim of identifying and clarifying theoretical issues relevant to multi-modal usability was addressed by developing the issues discussed in chapter three into a new definition of modality, a new taxonomy, and the clarification of other properties and cognitive issues relevant to modality.

A new three-part definition of modality was proposed: a modality is a temporally based instance of information perceived by a particular sensory channel. This definition is based on the elements of sense, information form, and temporal nature, and is novel in that it explicitly includes the temporal element, identified as being important in terms of both cognitive processing and the loading effects of modalities.

A twenty-seven cell taxonomy was developed to support this definition by providing a structure for identifying modalities. The taxonomy has three groups of components: the sensory components, categorised as visual, acoustic and haptic; the information form components, categorised as lexical, symbolic and concrete; and the temporal components, categorised as discrete, continuous and dynamic. Each cell represents one modality, allowing for a large coverage of the interaction space and precise identification of modalities, while the classification is small enough to be easily applied.

The work on the four issues identified in chapter three: identification and granularity; user and system; combination; and time and ordering; were applied to the new definition

of modality. Modalities are identified as User or System, and Receptive or Expressive. The granularity of modalities is addressed by their definition as composite, atomic and dependent. Issues of mismatch and redundancy are explored.

The cognitive restrictions established in chapter three were developed into five types of clash: physical, relating to sensory processing restrictions; lexical, relating to language processing restrictions; temporal, relating to temporal processing constraints; semantic, relating to clashes of meaning due to the automatic cognitive processes involved; and clash unless expert, relating to the constraints associated with levels of expertise.

### **9.2.2 Application of theory**

The second research aim was to apply the new theory of multi-modal usability, which necessitated the development and use of a new methodology, Evaluating Multi-modal Usability (EMU). Three areas were considered: which properties and elements should be incorporated; what form of representation should be used within the methodology; and the overall structure of the methodology, guiding its application and the kind of issues to be addressed.

The results of the research examined and presented in the previous chapters was used to identify those matters which the methodology should address, and the definition of modality was applied to the robotic arm interface examined in chapter two to provide a starting point for development. The robotic arm study identified the need for well-defined procedures and a more meaningful notation.

The investigation of the needs of representation found that a notational form was best able to clearly represent issues of ordering, combination, preconditions and choice, as well as more complex interfaces. Various representation formats were tried, including tabular representations such as User Action Notation, and diagrammatic representations such as CPM and Petri Nets. These were found to be unable to represent the complexity of the modalities and their properties. A notational representation was developed which could show the temporal flow of the interaction, as a whole and in smaller sections, as well as make explicit the preconditions applying to the modalities and the choice between them. It was also able to handle the complex representation of large interfaces. The final representation can clearly show modalities in terms of individual modalities, preconditions, and choice, as well as different levels of granularity.



The structure of the methodology was developed by iterating around the example of the telephone, and resulted in a methodology of eight stages to support the analyst in identifying usability issues, both of a general nature and those specifically relating to multi-modal usability. The approaches examined in chapter two, namely STN, CW, GOMS, PUM and Z, were used extensively both as sources of inspiration and as examples of what not to do. The methodology stages encourage the analyst to iterate round the interaction, and specifically encourages the investigation of multi-modal usability issues. The methodology overall can handle issues of task, profiles, and clashes, as well as the listing of the interaction sequence.

The developed methodology was then applied to the portion of the robotic arm studied previously in chapter two. This was to examine issues relating to the coverage of the new methodology when used by an experienced analyst. The analysis found various usability problems which were different in nature to those uncovered earlier. Issues surrounding the effective use of the methodology were also raised and discussed.

In order to determine the usability and usefulness of the new definition, taxonomy and methodology in a wider context, a case-study using computer science students to assess the usability issues associated with two tasks performed on two multi-modal interfaces was carried out, and reported in chapter six.

This case-study found that with regards to the theoretical aspects of multi-modality, the definition and taxonomy were both usable and useful, with subjects able to use the definition to successfully identify modalities, and most modalities written correctly. There were minor problems with some system modalities wrongly written, and occasional difficulty with the use of types and preconditions.

There were problems with other theoretical aspects of multi-modal usability, in that very few subjects identified any clashes or properties. This might be due to confusion over what these are, which requires clarification of the relevant theory, or that the subjects did not have enough time for this stage, and therefore the methodology may be too long and needs condensing.

Subjects had mixed feelings regarding the methodology. They were generally able to use the notation and to successfully model the interactions. However, they had difficulty with the use of preconditions, and many subjects did not use AND/OR to model issues of concurrency and choice. Subjects seemed either unclear as to what constituted a

usability issue, or reluctant to find fault with the interface. This may be due to problems about how to define a usability problem, given their contextual basis and the lack of absolutes. This could be remedied through the use of context-based approaches using particular parameters to establish usability thresholds (Cockton & Lavery, 1999) However, usability issues were identified through the use of the methodology.

In chapter seven empirical video data of the robotic arm in use was analysed to determine if any of the usability issues identified by the approaches could be supported by empirical evidence. Many of the usability issues could not be confirmed due to the limitations of the video data. That the video data was able to provide empirical evidence to confirm five issues supports the ability of all six approaches to assess interface usability. Three of these issues were originally identified by EMU, confirming the ability of the methodology to successfully identify some usability issues.

In chapter eight the limitations of the theory and methodology were examined in order to assess future implications. The subjects' problems with types, preconditions, AND/OR, properties and clashes, suggests that the methodology needs significant improvement in these areas critical to multi-modal usability in terms of the supporting documentation, and in the identification of usability issues. Another area of concern is the length of EMU. At present it tries to be too comprehensive. Rather than attempting to deal with all aspects relating to the modelling of an interaction, concentration on smaller segments of interaction with more emphasis being placed on particular multi-modal issues may be a better approach, thus reducing overheads of time and effort.

### **9.2.3 Scoping of usability modelling techniques**

The final aim of this research was to examine problems in applying modelling techniques, with regards to their scope, and the development of the necessary expertise to use them correctly.

This theme became apparent as a result of the original analyses of the robotic arm in chapter two, which found that the five techniques are very different in terms of the level of support they provide for the analyst in identifying usability issues, and that many usability issues can be identified through the craft skill of the particular analyst. Despite STN and Z being representations and thus providing no support, being used in software engineering for purposes other than evaluating interfaces, many usability issues were identified through their use. GOMS, PUM and CW, being designed for usability



assessment, have well defined procedures; however, they also rely significantly on the ability of the analyst to identify areas of potential concern, demonstrating that even methods which purport to support the analyst and focus on user behaviour are dependent to some extent on the analyst's inherent skill.

The study also found that the representation can affect ease of use. GOMS, PUM and Z require the analyst to learn particular notations and means of application, whereas STN and CW allow for more freedom of representation. In the study, STN and CW were the most straightforward to apply, due to their simple representational style, followed by GOMS, with PUM and Z hampered by both the degree of formalism necessary and the need to learn a specific notation. This suggests that effective techniques are those which utilise a clear representational style, and allow for straightforward application, with well defined procedures and parameters. However, it must be observed that there are other versions of GOMS, such as NGOMSL (Kieras, 1997), which might produce different results. These conclusions must be read in this light.

Chapter seven re-examined the original analyses of the robotic arm, including the EMU analysis, with reference to their supporting literature in order to carefully assess the scope of the approaches, and to determine whether the usability issues were identified due to the power of the technique, the skill and knowledge of the analyst, or other factors. The study found that not all issues were identified strictly by the use of a particular method. Issues were identified through source materials, the artefact produced, and the craft skill or existing usability expertise of the analyst. The video data analysis found that there were some usability issues that had not been uncovered by the techniques which should have been. This has wide implications, since not taking these factors into account means that a method's claims can be inflated, and a misleading impression given of its scope, with the final scope possibly being wider than intended due to the expertise of analyst.

The development of expertise was discussed as a factor in the successful application of techniques. There will be a large difference in analyses done of the same interface depending on whether the analyst is a novice or expert in the technique, and in identifying usability problems generally (see Nielsen, 1992). The implications are that differences between novices and experts must be explicitly recognised by both method developers and interface developers, since they will identify different problems and therefore cannot guarantee complete coverage.

Other aspects examined discussed how the amount of expertise necessary to correctly apply a technique may lead to issues being omitted, incomplete representations, and representations produced at inappropriate levels of detail. The form of representation used not only constrains the types of issues which can be identified, but also how much practise is needed in learning how to use it. Attention which is focused on applying the representation may miss issues outside the direct scope of the method. However using a loose representation may detract from the underlying theory of the method.

The role of expertise and the need to strengthen aspects of EMU to accommodate this were discussed in chapter eight, as noted above.

### **9.3 Contributions and limitations**

The previous section discussed how this thesis has addressed the research aims. The strands of research, although connected, have had differing levels of success, which is unsurprising given their breadth and complexity. In this section the overall contributions to knowledge from the research, and the research limitations, are presented.

The original aim was to provide a theoretical understanding of multi-modal usability. This was primarily motivated by the confusion resulting from the various contradictory definitions for modality current within the HCI community. In this, the research has been very successful. The issue of definition has been carefully examined and a new definition and taxonomy proposed. This new three-part definition of modality is a fundamental contribution. The idea of combining elements into one definition is itself not novel, but this particular combination with its inclusion of sense, information form, and time, allows for a better and more effective description of those information entities involved in the interaction. In contrast to many other theoretical definitions, this has been practically applied in the usability assessment of the ticket machine and heating controller interfaces, and found to be successful in identifying modalities in this manner. The associated taxonomy is also a novel contribution. It is of manageable size, with twenty-seven individual modality types, and the categories of elements are clear. Its use by non-developers substantiates claims of coverage, clarity and usability. They have found the definition and taxonomy remarkably straightforward to use, although minor problems with regards to particular sense and information form categories have been noted.



The investigation of the wider issues surrounding multi-modal interaction has resulted in four important issues being identified, relating to: identification and granularity; user and system; combination; and time and ordering. The cognitive restrictions surrounding multi-modal usability have also been examined, and developed into five types of clash: physical, relating to sensory processing restrictions; lexical, relating to language processing restrictions; temporal, relating to temporal processing constraints; semantic, relating to clashes of meaning due to the automatic cognitive processes involved; and clash unless expert, relating to the constraints associated with levels of expertise.

The identification of these additional issues has enabled modalities to be seen within a wider context, in terms of their interaction, rather than merely identified by the definition and taxonomy. Many of these concepts are not completely novel, for example the identification of modalities as User or System was presaged by the CARE properties (Coutaz et al, 1995). However, this is one of the more complete attempts at examining and defining the properties of multi-modal interaction, and is thus an important contribution to research.

The second aim was to apply the theory to the evaluation of multi-modal usability practically. There are problems with EMU as discussed earlier, and there needs to be further work in resolving them. It is unable to deal with some important aspects of multi-modal usability, relating to transitions, and at present it is too long and complicated. This is common to most methodologies which undergo various iterations during development in the light of practitioner experience. However, EMU is a positive contribution to the research community given the lack of approaches which specifically look at multi-modal usability. In addition to being developed, it has also been investigated on non-practitioners, which means that it has been tested and is not a purely theoretical construct.

The investigation of usability modelling is a pertinent contribution to the research community given that there has been little previous investigation into the scope of techniques and the role of expertise and the development of expertise in methodologies. Since this was not the primary aim of the research it has not been able to investigate this in depth, but the possible problems raised are still valid and require further elucidation.

## 9.4 Current research

The problems with multi-modal usability have not gone away, they are simply relevant in different ways from five years ago. The rise of the Web as an information source capable of multimodal interaction is pushing research. The issue is now not so much about technology, but the effective use of that technology in the manipulation of information. Information is now the biggest resource in the modern world, and the corollary to this is that access to this information is essential. Given the amount of information available, the challenge is to effectively combine it, so that the user is not overwhelmed, and to resolve issues of time, load and combination, as well as cognitive restrictions. The appropriate theory and methodologies to support this are still required, since development is outstripping theory in this field.

The past few years has seen significant effort in combining modalities, and research into the problems associated with this. Much of this has happened within the field of multimedia, prompted not only by the rise of CD-ROM packages for stand-alone personal computers, but also by the use of the Web for transmitting text, video and audio information. For example, (Faraday & Sutcliffe, 1998) with their work on providing contact points or co-references between images and text in multimedia, argue that having to integrate disparate sources of information interferes with attention and imposes a heavy cognitive load, but where material can be properly combined, and attention therefore concentrated rather than split, performance is improved. This kind of research is important, for as Ghinea & Thomas (1998) state: "Users have difficulty in absorbing audio, visual and textual information concurrently...Highly dynamic scenes, although expensive in resources, have a negative impact on user understanding and information assimilation." New ways of manipulating information require combinations of modalities, for example the CueVideo system (Ponceleon et al, 1998) which is designed to browse catalogued video which has been annotated using speech as well as text. There has been work on adapting the content of multimedia presentations by the user, such as that by Bultermann (1998) who uses "channels", a "grouping abstraction for a set of media items that share some common attributes", for example text colour, natural language, layout and background, which the user can switch off if not needed. Other work, for instance that by Keates & Robinson (1999) on multimodal gestural input, shows that multimodal input is still a very important research area.



The growth of ubiquitous, mobile, personal gadget-style computing is also influencing current research. Handheld web access is proposed as the next big thing, with use set to explode over the next three to five years (Stenton, 1999) (Jones et al, 1999). Some sites have been developed for particular devices, e.g. Psion's PocketInfo, but as Jones et al (1999) point out, this goes against the platform independent ethos of the Web, and "places a heavy burden on content providers who want as wide as possible access to their information." There are obvious constraints as to size of screen, problems with the display of information, the use of greyscale and the representation of colour on devices, however, the possible effects on interaction time and user effort is as yet under-researched. People using smaller handheld devices possibly have different goals from those using the Web on big devices: more searching for specific information than browsing (Jones et al 1999). Given these differing constraints, different modalities may be appropriate: for example, the use of graded auditory cues (Schmandt, 1999), or even making use of proxemics, the space around the body (Susani, 1999). Work is being done on taking into account the context of the user in a physical sense (see Jose & Davies, 1999, Couderc & Kermarrec, 1999, Schmidt et al, 1999); work should also be done on the appropriateness of the modalities used.

The increase of information is paralleled by the need to understand and combine its attributes in meaningful ways. The rise of Digital Libraries which combine various forms of data has seen a demand for increasingly sophisticated retrieval systems which do not just get obvious attributes such as shape, motion, texture (Huang et al, 1998). The work on descriptive standards, for example the work on MPEG7 "Multimedia Content Description Interface", begun in 1996, will result in a standard set of descriptors that can be used to specify various types of multimedia (Ponceleon et al, 1998).

Therefore, the challenge of combining these modalities in positive ways is, if anything, greater than the situation five years ago, with development, as always, leading the field, waiting for theory and techniques to catch up and provide a stable support.

## **9.5 Future work**

The research presented in this thesis has potential for extension in several areas. The new definition of modality and taxonomy provide a vocabulary which can be used to re-examine existing research, placing it within a common context. This will enable

previous work to be reassessed in relation to other research and provide additional insight into multi-modal usability. Candidates for particular attention include the research of Bensen (1995a) and Coutaz et al (1993, 1995) as well as a wider spectrum of research in the field of multimedia and virtual reality. Existing cognitive work such as that by Barnard and May (1993) on the structure of interfaces might also be amenable to incorporation. Together, this would provide a substantial and coherent basis of necessary theory for the developers of multi-modal systems.

One main area which requires future exploration is that relating to the nature of expertise, and the implications this has for usability modelling in an HCI context. This theme unintentionally arose from the work conducted in this thesis, and became a subsidiary focus of interest. As such, it was not possible or practical to investigate it in any great depth. However, there is a need to understand the effect of expertise on analysts when modelling usability issues, both in the wider context of HCI, and in the more immediate context of EMU and multi-modal usability.

Klein (1998) argues that expertise is most quickly and effectively developed when immediate feedback can be given. This needs to be considered when deciding how best to teach EMU, and under what circumstances.

Related to the development of expertise is the need to investigate the needs of designers. Although EMU and the associated theory have been examined theoretically and on a small scale, there may need to be changes based on a fuller understanding of the constraints under which designers normally use (or don't use) HCI techniques. This would involve examining in more detail the work that designers do, and investigating how EMU might help them in developing multi-modal interfaces.

As mentioned earlier, EMU has not yet been applied on a large scale. Since it is a methodology which supports various levels of detail, thus potentially requiring several applications at different levels to gain a full understanding, this needs to be investigated further. The application of EMU to novel devices should also be considered in depth.

A current limitation of EMU is that it is very detailed on aspects which are available in other techniques, which suggests that it should be modified to concentrate more on those aspects relating to multi-modal issues. Various stages of the methodology could be revised to accommodate this, by, for example, merging stages 1 and 3. Other notations can be examined for the leverage they might give to representing multi-modal



interaction, such as the handling of pre-conditions in Event CSP (Alexander, 1990). The terminology needs clarifying, and the issue of transitions, at present a major gap in the methodology's capabilities, needs addressing. It also needs to be determined whether EMU is concerned more with modalities as discrete events or as ongoing sub-processes, which might help to address the notion of duration (Dix & Abowd, 1996). Lastly, EMU needs revising to explicitly address in what context usability problems should be identified, and how threshold parameters might be established.

## **10. BIBLIOGRAPHY**

ALEXANDER, H. (1990) Structuring Dialogues Using CSP. In HARRISON, M. & THIMBLEBY, H. (Eds.): *Formal Methods in Human-Computer Interaction*. Cambridge University Press, UK

ALPERT, S. R., SINGLEY, M. K., CARROLL, J. M. (1995) Multiple multimodal mentors: delivering computer-based instruction via specialised anthropomorphic advisors. In: *Behaviour and Information Technology*, 14(2), pp. 69-79

ALTY, J. L. (1991) Multi-media - what it is and how we exploit it. In DIAPER, D. & HAMMOND, N. (Eds.): *Proceedings of HCI '91*. Cambridge University Press, pp. 31-46

ANZAI, Y. (1994) Human-robot-computer interaction: a new paradigm of research in robotics. In: *Advanced Robotics*, Vol. 8, No. 4, pp. 357-369

BAILEY, K. D. (1994) *Typologies and taxonomies: an introduction to classification techniques*. (Sage University Paper Series on Quantitative Applications in the Social Sciences, series no. 07-102) Sage publications, USA

BARNARD, P. J. (1996) Cognitive theory and the design of advanced interfaces. In: *Proceedings of APCHI'96*, pp 49-63

BARNARD, P.J. & TEASDALE, J. (1991). Interacting Cognitive Subsystems: A systemic approach to cognitive affective interaction and change. *Cognition and Emotion* , 5, 1-39.

BARNARD, P. AND MAY, J. (1993) Cognitive modelling for user requirements. In P.F. BYERLEY, P. J. BARNARD, AND J. MAY, (Eds.): *Computers, Communication and Usability: Design issues, research and methods for integrated service*. North Holland Series in Telecommunication pp.101-145. Elsevier: Amsterdam.

BASTIDE, R. & PALANQUE, P. (1990) Petri net objects for the design, validation and prototyping of user-driven interfaces. In DIAPER, D. GILMORE, D., COCKTON, G. &



SHACKEL, B. (Eds.): Human-Computer Interaction - INTERACT'90. pp.625-631, Elsevier Science Publications, North Holland, Netherlands.

BEARNE, M., JONES, S., SAPSFORD-FRANCIS, J. (1994) Towards usability guidelines for multimedia systems. In: Proceedings of Multimedia '94, ACM Press, pp 105-110

BERNSEN, N. O. (1995a) A revised generation of the taxonomy of output modalities. In BERNSEN, N. O., JENSAGER, F., LU, S., VERJANS, S. (eds): Information theory and information mapping, Amodeus project deliverable D15

BERNSEN, N. O. (1995b) A toolbox of output modalities: representing output information in multi-modal interfaces. In BERNSEN, N. O., JENSAGER, F., LU, S., VERJANS, S. (eds): Information theory and information mapping, Amodeus project deliverable D15

BERNSEN, N. O. (1995c) A taxonomy of input modalities. In BERNSEN, N. O., JENSAGER, F., LU, S., VERJANS, S. (eds): Information theory and information mapping, Amodeus project deliverable D15

BLANDFORD, A. E., BUCKINGHAM SHUM, S. J. & YOUNG, R. M. (1998a) Training software engineers in a novel usability evaluation technique. In: International Journal Human-Computer Studies, 49, pp. 245-279

BLANDFORD, A. E., GOOD, J., YOUNG, R. M. (1998b) Programmable User Modelling Analysis for usability evaluation. PUMA project working paper WP11. (available from [www.cs.mdx.ac.uk/puma](http://www.cs.mdx.ac.uk/puma))

BLANDFORD, A. E. & DUKE, D. J. (1997) Integrating user and computer system concerns in the design of interactive systems. In: International Journal Human-Computer Studies 46, 653-679.

BLANDFORD, A. E. & YOUNG, R. M. (1996) Specifying user knowledge for the design of interactive systems. In: Software Engineering Journal. 11.6, 323-333.

BRETAN, I. & KARLGREN, J. (1993) Synergy effects in natural language-based multimodal interaction. In: Proc. of 1993 ERCIM Workshop on Multimodal HumanComputer Interaction, INRIA, Nancy

- BUCKINGHAM SHUM, S. & HAMMOND, N. (1994) Transferring HCI modelling and design techniques to practitioners: a framework and empirical work. In: *People and Computers IX: Proceedings of HCI'94*, Cambridge University Press, pp 21-36
- BULTERMANN, D. C. A. (1998) User-centered abstractions for adaptive hypermedia presentations. In: *Proceedings of ACM Multimedia '98*, ACM New York
- CARD, S. K., MORAN, T. P. AND NEWELL, A. (1983). *The Psychology of Human Computer Interaction*, Hillsdale : Lawrence Erlbaum.
- COCKTON, G. (1998) Let's Get It all Together: Literate Development and the Integration of HCI Research. In *CHI 98 Basic Research Symp.*
- COCKTON, G. & LAVERY, D. (1999) A Framework for Usability Problem Extraction. In: SASSE, M. A., JOHNSON, C. (Eds.): *Human-Computer Interaction - INTERACT'99*, IOS Press, The Netherlands, pp. 344 – 352.
- COCKTON, G. & WOOLRYCH, A. (2001) Understanding Inspection Methods: Lessons from an Assessment of Heuristic Evaluation. In: *People and Computers XV: Proceedings of HCI'01*, Springer, pp 171-191.
- COHEN, P. R. (1992) The role of natural language in a multimodal interface. In: *Proceedings of UIST'92*, Monterey, California, November 15-18 1992, pp 143-149
- COHEN, P. R. (1984) The pragmatics of referring and the modality of communication. In: *Computational Linguistics*, Vol 10, No 2, April-June 1984
- COUDERC, P. & KERMARREC, A. -M. (1999) Enabling context-awareness from network-level location tracking. In: GELLERSON, H-W. (ed.) *Proceedings of Handheld and Ubiquitous Computing '99*, Springer-Verlag, Berlin
- COUTAZ, J., NIGAY, L. & SALBER, D. (1993) The MSM Framework: A Design Space for Multi-Sensory-Motor Systems. In L. BASS, J. GORNOSTAEV, & C. UNGER (eds): *Lecture Notes in Computer Science 753 (EWHCI'93) Selected Papers*. pp. 231-241. Springer-Verlag.



- COUTAZ, J., NIGAY, L., SALBER, D., BLANDFORD, A., MAY, J., YOUNG, R. (1994) Four easy pieces for assessing the usability of multimodal interaction: the CARE properties. Amodeus project document ID/WP37
- COUTAZ, J., MAY, J., YOUNG, R., BLANDFORD, A., NIGAY, L., SALBER, D. (1995) Integrating system and user modelling through abstraction: the CARE properties for reasoning about multimodality. In: NORDBY, K., HELMERSEN, P., GILMORE, D. J., AND ARNESEN, S. (eds): Human-Computer Interaction: Interact'95. Chapman and Hall, pp. 115-120
- DIX, A. & ABOWD, G. (1996) Modelling Status and Event Behaviour of Interactive Systems. *Software Engineering Journal* 11 (6) pp. 334-346
- DIX, A., FINLAY, J., ABOWD, G. & BEALE R. (1993) *Human-Computer Interaction*, Prentice Hall International
- DOWELL, J., LIFE, A., SALTER, I. (1994) The design space for a multimodal multimedia travel facility. In: Proceedings of ECCE7, European Society for Cognitive Ergonomics, September 5-8 1994, Bonn
- DOWELL, J., LUKAU, W., SHMUELI, Y., SALTER, I., LIFE, A. (1995) Designing the Multimodal Speech Interface to a Public Travel Facility. In: DE MORAES, A. & MARINHO, S. (eds). Proc. International Ergonomics Association Conference, Rio de Janeiro, Brazil, 16-20 October, pp. 288-291
- DREYFUS, H. L. (1992) *What computers still can't do: a critique of artificial reason*. Revised edition, MIT Press, USA
- DREYFUS, H. L. & DREYFUS, S. E. (1985). *Mind over machine*. New York : Macmillan / The Free Press.
- DUKE, D. J. & HARRISON, M. D. (1995) Event model of human-system interaction. In: *Software Engineering Journal*, January 1995
- DUKE, D. J., BARNARD, P. J., DUCE, D. A., MAY, J. (1998) Syndetic Modelling. In: *Human-Computer Interaction*, 13, 4, pp. 337-393
- EBERTS, R. E. (1994) *User Interface Design*. Prentice-Hall, Inc. USA

EYSENCK, M. & KEANE, M. T. (1990) *Cognitive psychology: a student's handbook*. Lawrence Erlbaum Associates Ltd, UK

FARADAY, P. & SUTCLIFFE, A. (1998) Making contact points between text and images. In: *Proceedings of ACM Multimedia '98*, ACM New York

FITZPATRICK, R. & HIGGINS, C. (1998) Usable Software and its Attributes: a Synthesis of Software Quality, European Community Law and Human-Computer Interaction. In: *People and Computers XIII, Proceedings of HCI'98*, Springer, pp.3-21

FROLICH, D. M. (1993) History and future of direct manipulation. In: *Behaviour and Information Technology*, 12, 6, pp 315-329

GHINEA, G. & THOMAS, J. P. (1998) QoS impact on user perception and understanding of multimedia video clips. In: *Proceedings of ACM Multimedia '98*, ACM New York

GOULD, S. J. (1983) *Hen's teeth and horse's toes*. Penguin Books, USA

GRAY, W. D., & SALZMAN, M. C. (1998) Damaged merchandise? A review of experiments that compare usability methods. In: *Human Computer Interaction*, 13, pp. 203-261

HALL, A. (1990) Seven Myths of Formal Methods *IEEE Software*, September, pp. 11-19

HARTSON, R., SIOCHI & HIX (1990) The UAN: a user-oriented representation for direct manipulation interface designs. In: *ACM Transactions on Information Systems*, 8, pp. 181-203

HINCHEY, M. G. & BOWEN, J. P. (1995) Applications of formal methods FAQ. In HINCHEY, M. G. & BOWEN, J. P. (eds) *Applications of formal methods*, Prentice-Hall International, U.K.

HUANG, J., KUMAR, S. R. & ZABIH, R. (1998) An automatic hierarchical image classification scheme. In: *Proceedings of ACM Multimedia '98*, ACM New York

HUTCHINS, E. L., HOLLAN, J. D., NORMAN, D. A. (1985) Direct manipulation interfaces. In: NORMAN, D. A., DRAPER, S. W. (eds.) *User Centred System Design*. Erlbaum



HYDE, J. K. (1995) Fusion engines and melting pots: an examination of a multi-modal model using a formal notation. Dissertation presented as partial fulfillment of MSc in Information Processing and the Man-Machine Interface, Department of Computer Science, University of York

HYDE, J. K., BLANDFORD, A. E., GOODMAN, H. S. (1998) A comparison of five techniques for investigating general and multi-modal usability issues. Middlesex University School of Computing Science technical report ISSN 1462-0871

JOHN, B. & KIERAS, D. (1994) The GOMS family of analysis techniques: tools for design and evaluation. Human-Computer Interaction Institute Technical Report CMU-HCII-94-181.ps, Carnegie Mellon University, Pittsburgh, USA.

JOHN, B. & KIERAS, D. (1996) The GOMS family of user interface analysis techniques: comparison and contrast. In: ACM Transactions on CHI., 3, 320-351.

JOHN, B. & MARKS, S. (1997) Tracking the effectiveness of usability evaluation methods. In: Behaviour and Information Technology 16, No. 4/5, 188-202.

JOHNSON, P. & NEMETZ, F. (1998) Towards principles for the design and evaluation of multimedia systems. In JOHNSON, H., NIGAY, L., ROAST, C. (eds): People and Computers X111 (Proceedings of HCI'98), Springer-Verlag

JOHNSON, C. (1995) The application of Petri Nets to represent and reason about human factors problems during accident analyses. In Palanque, P. and Bastide, R. (eds): Design, Specification and Verification of Interactive Systems '95, Springer-Verlag/Wien, Austria

JOHNSON, C. (1996) Evaluation of user interface notations. In BODART, F. AND VANDERDONCKT, J. (eds): Design, Specification and Verification of Interactive Systems '96, Springer-Verlag/Wien, Austria

JONES, M., MARSDEN, G., MOHD-NASIR, L., BOONE, K. & BUCHANAN, G. (1999) Improving Web interaction on small displays. In: Computer Networks 31, 1129-1137

JOSE, R. & DAVIES, N. (1999) Scalable and flexible location-based services for ubiquitous information access. In: GELLERSON, H-W. (ed.) Proceedings of Handheld and Ubiquitous Computing '99, Springer-Verlag, Berlin

- KUAR, K., MAIDEN, N., SUTCLIFFE, A. (1999) Interacting with virtual environments: an evaluation of a model of interaction. In: *Interacting with computers*, Vol 11, No 4, April 1999
- KEATES, S. & ROBINSON, P. (1999) Gestures and multimodal input. In: *Behaviour and Information Technology*, Vol. 18, No. 1, pp. 36 44
- KIERAS, D. E. (1997) A Guide to GOMS Model Usability Evaluation Using NGOMSL. In HELANDER, M. G., LANDAUER, T. K., PRABHU, P. V. (Eds.): *Handbook of Human-Computer Interaction*. Second Edition, Elsevier Science, North-Holland, chapter 31, pp. 733-766
- KIERAS, D. E., MEYER, D. E. (1995) An overview of the EPIC architecture for cognition and performance with application to human-computer interaction. In: *Human-Computer Interaction*, 12, 391-438
- KLEIN, G. (1998) *Sources of power: how people make decisions*. MIT Press, UK
- KRAPLICHER, C., HAUBNER, M., LOSCH, A., SCHUHMANN, D., SEEMANN, M., ENGLEMEIER, K-H. (1999) Physicians in virtual environments - mulitmodal human-computer interaction. In: *Interacting with computers*, Vol 11, No 4, April 1999
- LAVERY, D. & COCKTON, G. (1997) Representing Predicted and Actual Usability Problems. In: *Proc. Int. Workshop on Representations in Interactive Software Development*, QMW London, pp. 97-108
- MARCELLA, R., & NEWTON, R. (1994) *A new manual of classification*. Gower, UK
- MAY, J., & BARNARD, P.J (1995) Towards supportive evaluation during design. In: *Interacting with Computers*, 7, 115-143
- MCKENZIE MILLS, K. & ALTY, J. (1997) Investigating the role of redundancy in multimodal input systems. In: WACHSMUTH, I. AND FROHLICH, M. (eds): *Gesture and Sign Language in Human-Computer Interaction*, Proceedings of the Gesture Workshop 1997, pp 159-171. Springer Verlag, Germany
- MELLOR, P. (1994) CAD: computer aided disaster. In: *High Integrity Systems*, Vol 1, No 2, 1994



- MILLS, S. & NOYES, J. (1999) Virtual reality: an overview of user-related design issues. In: *Interacting with computers*, Vol 11, No 4, April 1999
- NEAL, J. G. & SHAPIRO, S. C. (1991) Intelligent multi-media interface technology. In SULLIVAN, J. W. & TYLER, S. W. (eds): *Intelligent user interfaces*. ACM, Addison Wesley, America
- NIELSEN, J. (1992) Finding Usability Problems Through Heuristic Evaluation. ACM Conference on Human Factors in Computing, CHI'92, pp.373-380
- NIGAY, L., COUTAZ, J., SALBER, D. (1993) MATIS: a Multimodal Airline Travel Information System. Amodeus Project Document: system modelling/WP10
- NIGAY, L. & COUTAZ, J. (1995) A generic platform for addressing the multimodal challenge. In: *Proceedings CHI'95*, ACM Press
- NORMAN, D. A. (1981) Categorization of action slips. In: *Psychological Review*, 88, pp. 1-15
- NORMAN, D. A. (1988) *The psychology of everyday things*. Basic Books, New York
- OVIATT, S. (1997) Multimodal interactive maps: designing for human performance. In: *Human-Computer Interaction*, Vol 12, pp 93-129
- PALMER, E. A., HUTCHINS, E. L., RITTER, R. D., VANCLEEMPUT, I. (1993) Altitude deviations: breakdowns of an error tolerant system. NASA Technical Memorandum, DOT/FAA/RD-92/7
- PANKHURST, R. J. (1978) *Biological identification: the principles and practice of identification methods in biology*. Edward Arnold, London
- PARSONS, B., PRIOR, S. D & WARNER, P. R. (1995) An Approach to Designing Manipulator Controller Software Employing Context Dependent Command Interpretation. In: *European Conference on Assistive and Rehabilitation Technology*, Lisbon, October 1995
- PEARSALL, J. (ed.) (1999) *Concise Oxford Dictionary*. Tenth Edition, Oxford University Press

- PONCELEON, D., SRINIVASAN, S., AMIR, A., PETKOVIC, D. & DIKLIC, D. (1998) Key to effective video retrieval: effective cataloging and browsing. In: Proceedings of ACM Multimedia '98, ACM New York
- PREECE, J., ROGERS, Y., SHARP, H., BENYON, D., HOLLAND, S., CAREY, T. (1994) Human-computer interaction. Addison Wesley, UK
- PURCHASE, H. (1999) A semiotic definition of multimedia communication. In: *Semiotica*, vol 123-3/4, pp 247-259
- RANGANATHAN, S. R. (1962) Elements of library classification. Asia Publishing House, Bombay.
- ROTH, S. F. & HEFLEY, W. E. (1993) Intelligent multimedia presentation systems: research and principles. In MAYBURY, M. T. (ed): *Intelligent Multi Media Interfaces*, MIT Press, USA, pp 13-58
- ROTH, S. F., CHUAH, M. C., KERPEDJIEV, S. & KOLOJEJCHICK, J. A. (1997) Toward an information visualization workspace: combining multiple means of expression. In: *Human-Computer Interaction*, Vol 12, pp 131-185
- SANDERSON, P. E., & FISHER, C. (1994a) Introduction to this special issue on Exploratory Sequential Data Analysis. In: *Human-Computer Interaction*, Vol. 9, pp. 247-250
- SANDERSON, P. E., & FISHER, C. (1994b) Exploratory Sequential Data Analysis: foundations. In: *Human-Computer Interaction*, Vol. 9
- SASSE, M. A. (1997) Eliciting and describing users' models of computer systems. PhD thesis, School of Computer Science, University of Birmingham
- SCAPIN, D. L. & BASTIEN, J. M. C. (1997) Ergonomic Criteria For Evaluating the Ergonomic Quality of Interactive Systems. In: *Behaviour & Information Technology*, Vol. 16, No. 4/5, pp. 220-231
- SCHIELE, F. & GREEN, T. R. G. (1990) HCI formalisms and cognitive psychology. In Harrison, M. & Thimbleby, H. (eds.): *Formal Methods in Human-Computer Interaction*, Cambridge University Press, England.



- SCHMANDT, C. (1999) Everywhere messaging. In: GELLERSON, H-W. (ed.) Proceedings of Handheld and Ubiquitous Computing '99, Springer-Verlag, Berlin
- SCHMIDT, A., AIDOO, K. A., TAKALUOMA, A., TUOMELA, U., LAERHOVEN, K. & VELDE, W. (1999) Advanced interaction in context. In: GELLERSEN, H-W. (ed.) Proceedings of Handheld and Ubiquitous Computing '99, Springer-Verlag, Berlin
- SPIVEY, J.M. (1989). The Z Notation: A Reference Manual, Prentice-Hall International.
- STENNING, K. & TOBIN, R. (1997) Assigning information to modalities: comparing graphical treatments of the syllogism. In: EJERHED, E. & LINDSTROM S. (eds) Logic, Action and Cognition, Kluwer Academic Publishers, Dordrecht, pp. 211-228
- STENTON, P. (1999) Maintaining context and control in a digital world. In: GELLERSEN, H-W. (ed.) Proceedings of Handheld and Ubiquitous Computing '99, Springer-Verlag, Berlin
- SUSANI, M. (1999) Mobile, ubiquitous and the sense of space. In: GELLERSEN, H-W. (ed.) Proceedings of Handheld and Ubiquitous Computing '99, Springer-Verlag, Berlin
- TAKEUCHI, A. & NAGAO, K. (1993) Communicating facial displays as a new conversational modality. In ASHLUND, S., MULLET, K., HENDERSON, A., HOLLNAGEL, E., WHITE, T. (eds): Human factors in computing systems: Interchi'93 conference proceedings, ACM New York
- VERJANS, S. (1995) State-of-the-art for input modalities. In BERNSEN, N. O., JENSAGER, F., LU, S., VERGANS, S. (eds) Information theory and information mapping, Amodeus project deliverable D15
- WHARTON, C., RIEMAN, J., LEWIS, C. & POLSON, P. (1994). The Cognitive Walkthrough Method: A Practitioner's Guide. In NIELSEN, J., & MACK, R. (ed): Usability Inspection Methods, pp105-140. Wiley: New York.
- WICKENS, C. D., SANDRY, D. L., VIDULICH, M. (1983) Compatibility and resource competition between modalities of input, central processing, and output. In: Human Factors, 25, 2, pp 227-248

WOODS, D. D., JOHANNESEN, L. J., COOK, R. I., SARTER, N. B. (1994) Behind human error: cognitive systems, computers, and hindsight. CSERIAC SOAR 94-01



# **11. APPENDIX A: A SURVEY OF FOUR**

## **COGNITIVE MODELS**

### **11.1 Human Information Processing Model**

The Human Information Processing Model (HIPM) proposed by (Wickens et al., 1983) utilizes compatibilities between stimuli and responses, on the basis that if they have the same form there will be fewer stages of processing needed, resulting in a quicker response. Interfaces can therefore be designed to exploit these compatibilities.

The model has three main stages: perception, cognition and motor response. Stimuli are stored in a short-term sensory store, which is iconic for visual and echoic for auditory. If they are not further processed, the stimuli will decay, at different rates depending on the store. Long-term memory will identify which stimuli to process further, at the perception stage. This stage is limited by the amount of attentional resources available. In the cognitive stage, decisions and response selections are made, using working memory, which is limited, and long-term memory, which is permanent. The mental operations at this stage are limited by the amount of attentional resources available, but iterations may occur. The motor response stage carries out the decision made in the cognitive stage, such as to move a hand or a head, and also requires attentional resources.

This theory has been expanded into stimulus-central processing-response compatibility (S-C-R), which emphasises the role of the cognitive processing. Differing tasks are represented using different codes. Wickens defines four main types of input (Eberts, 1994), which can be verbal or spatial, and auditory or visual, as seen in the table below:

	VERBAL	SPATIAL
AUDITORY	speech	sound localisation and pitch
VISUAL	text	analog picture

**Table 11-1: Types of Input**

Tasks can be either verbal or spatial, and if inputs can be matched with a suitable representation, then the processing of the task can be made quicker. There are two kinds of output defined: manual response, which is spatial in nature; and speech, which is verbal in nature. This theory is dependent on being able to specify the representation code of the task. Tasks which combine verbal and spatial representations cannot at present be modelled, and only auditory and visual input, and auditory and haptic output, are considered, which makes its application more limited. There may also be things which affect compatibility of stimuli and response other than representation.

## **11.2 Model Human Processor**

The Model Human Processor (MHP) (Card et al, 1983) is a simplistic representation of the human mind using the information processing paradigm. It is not intended to be complete and the authors explicitly define the limits of the MHP as being for expert users and concentrating on the temporal issues relating to tasks. It comprises a set of memories and processors, and a set of principles of operation. There are three sub-systems: the perceptual system, the motor system, and the cognitive system, each with its own memories and processors.

The perceptual system takes information from the sensory systems, stores it in the appropriate image store, and passes the information on to the cognitive processor. The time it takes to do this depends on the particular sense and stimuli involved. The motor system controls the muscles that are used to carry out instructions from the cognitive processor. With regard to HCI, the arm-hand-finger and the head-eye systems are considered to be the



most important. The cognitive system has two parts: the memory, and the processor. The memory itself is divided into working memory, which holds the information currently being considered, and long-term memory, which stores knowledge for future use. The recognize-act cycle is the basic function of the cognitive processor. In the recognize phase, the contents of working memory are examined by long-term memory, modifying the contents of working memory ready for the next cycle.

The perceptual and motor systems can act in parallel, but the cognitive system is parallel only in the recognize phase. In the action phase it is serial. Therefore, although the cognitive system can be receiving many inputs, it can only initiate one output at a time. Multiple activities are performed by allocating small serial amounts of time to each task, so that they appear to be done in parallel.

This model is to establish the best possible times for performing a task, by timing the flow of information through the system. As such, it can model expert user's actions.

### **11.3 Interacting Cognitive Subsystems**

Interacting Cognitive Subsystems (ICS) (Barnard and May, 1993) is a parallel cognitive architecture made up of nine distinct subsystems, each with different capabilities, although similar in structure. There is no central processor or memory store. The architecture operates as a system, in that behaviour is considered as the result of more than one process. It is claimed to be more comprehensive than other architectures because of the way that it handles perception, cognition, emotion, and body control issues.

The nine subsystems have a common structure of an input channel, an image record which transfers input information without changing it, and transforming processes which transform the information from one mental code into another. All these processes can happen in parallel.

There are three types of subsystem. The sensory subsystems handle incoming sensory data, and comprises the acoustic, visual and body-state subsystems. The two effector subsystems handle outgoing sensory data: the articulatory subsystem handling the information needed in order to speak, and the limbic subsystem processing the information needed to move. The four central subsystems are concerned with high-order representations. The

morphonological subsystem deals with speech forms, the object subsystem deals with entities and relationships in visual space, the propositional subsystem handles entities and relationships in semantic space (or the “knowing something”), and the implicational subsystem handles schematic models of experience.

The architecture takes a highly dynamic view of processing, with information being repeatedly transformed, interpreted and re-represented, in order to make predictions about patterns of cognitive activity. Because a transformation process can only deal with one single data stream at a time, if there are two data streams that do not make a whole representation, then only one of them will be processed, which explains phenomena such as the cocktail party effect.

#### **11.4 Executive Process-Interactive Control**

Executive Process-Interactive Control (EPIC) (Kieras and Meyer, 1995) is a cognitive architecture especially designed to model multi-task human activity, by representing the perceptual, motor and cognitive constraints on humans as they carry out activities. The architecture consists of various motor and sensory processors, and a production-rule cognitive processor. EPIC takes into account the way that various sensory and motor processes influence activity, by including constraints and timing data based on current empirical evidence, resulting in a more realistic model.

The architecture consists of: three sensory processors (auditory, visual and tactile); three motor processors (ocular, vocal and manual); and a cognitive processor which is made up of working memory and a production rule interpreter. Long-term memory and production memory are separate to the cognitive processor and feed into the production rule interpreter.

Information from the senses is processed through the relevant perceptual processors to the cognitive processor, and then information from the cognitive processor is passed to the relevant motor processors. Each of the different processors have different timing settings, which are based on existing empirical data.

The cognitive processor is programmed in terms of production rules in the format (<rule-name> IF <condition> THEN <actions>). If the condition is in the production system working memory then the actions can be performed, which might be to manipulate the



contents of working memory, or to send a command to one or more of the motor processors. The cognitive processor works in cycles of activity, and incoming inputs have to wait until a new cycle begins before they can be processed.

Working memory in EPIC is split into four main areas: the visual, auditory and tactile working memory; the motor working memory (current state of motor processes); the control store containing items that represent current goals, and current steps for accomplishing goals; and general, which contains the miscellaneous task information.

Motor processors complete movements in two parts. The preparation phase is when the cognitive processor sends the commands to the motor processor, which transforms the information into a set of movement features, and prepares the features. The execution phase is where the motor processor actually carries out the movement features. Features remain in the motor-processor's memory, thus features of future movements can be performed more quickly if identical to the current movement. Also, if the next movement or movements can be decided in advance, the motor processor can be commanded to prepare all the motor features and save them in motor memory, ready to be executed.

The cognitive processor can set off any number of rules simultaneously, but the overall system is limited by the structure of the sense organs and effectors, for instance, eyes can only look at one place at a time, and the hands have a bottleneck of one processor. Therefore, although the cognitive requirements of a task might be trivial, the actual performance might be hampered by the nature of response required. EPIC is used in a similar fashion to the MHP, to demonstrate the sequence of actions required by a particular task, and how long these actions might take.

# 12. APPENDIX B: TAXONOMY AS APPLIED TO

## ROBOTIC ARM

User task: Turn on light switch.

The action sequence for the task was that the arm needed to be moved to the right, then the arm part moved up and the gripper moved out.

When arm is at rest, there is an initial menu. There is a choice for the user between **MoveArm**, which governs the movement of the arm as a whole, and **Move**, which governs the movement of individual joints.

System Expressive:

Menu Display

(vis-lex-cont)

Cursor

(vis-sym-dyn)

Position of arm

(hap-con-cont)

User Receptive:

Menu display

(vis-lex-cont)

Cursor

(vis-sym-dyn)

Position of arm

(vis-con-cont)

**MoveArm** chosen:

User Expressive:

Head movement

(hap-sym-dis)

or name option

(aco-lex-dis)

Menu display alters to show following choices:

**Left, Right, Up, Down, Forward, Back, End**

System Expressive:

Menu Display

(vis-lex-cont)

Cursor

(vis-sym-dyn)

Position of arm

(hap-con-cont)

User Receptive:

Menu display

(vis-lex-cont)

Cursor

(vis-sym-dyn)

Position of arm

(vis-con-cont)



**Right chosen:**

User Expressive:  
Head movement  
(hap-sym-dis)  
or name option  
(aco-lex-dis)

Menu display alters to show following choices:

**Go, Speed Level, End**

System Expressive:  
Menu Display  
(vis-lex-cont)  
Cursor  
(vis-sym-dyn)  
Position of arm  
(hap-con-cont)

User Receptive:  
Menu display  
(vis-lex-cont)  
Cursor  
(vis-sym-dyn)  
Position of arm  
(vis-con-cont)

**Speed Level chosen:**

User Expressive:  
Head movement  
(hap-sym-dis)  
or name option  
(aco-lex-dis)

Menu display alters to show following choices:

**Min, Slow, Medium, Fast, Max**

System Expressive:  
Menu Display  
(vis-lex-cont)  
Cursor  
(vis-sym-dyn)

User Receptive:  
Menu display  
(vis-lex-cont)  
Cursor  
(vis-sym-dyn)

**Min chosen:**

User Expressive:  
Head movement  
(hap-sym-dis)  
or name option  
(aco-lex-dis)

Menu display alters to show following choices:

**Go, Speed Level, End**

System Expressive:

Menu Display

(vis-lex-cont)

Cursor

(vis-sym-dyn)

User Receptive:

Menu display

(vis-lex-cont)

Cursor

(vis-sym-dyn)

**Go** chosen:

User Expressive:

Head movement

(hap-sym-dis)

or name option

(aco-lex-dis)

The arm will continue moving until **Stop** is selected:

System Expressive:

Menu Display

(vis-lex-cont)

Cursor

(vis-sym-cont)

Position of arm

(hap-con-dyn)

User Receptive:

Menu display

(vis-lex-cont)

Cursor

(vis-sym-cont)

Position of arm

(vis-con-dyn)

Note: cursor will be under the only option, so has changed from dynamic to continuous.

**Stop** chosen:

User Expressive:

Head movement

(hap-sym-dis)

or name option

(aco-lex-dis)

or the arm reaches the limit of movement.

Menu display alters to show following choices:

**Continue, Speed Level, End**



System Expressive:

Menu Display

(vis-lex-cont)

Cursor

(vis-sym-dyn)

Position of arm

(hap-con-cont)

User Receptive:

Menu display

(vis-lex-cont)

Cursor

(vis-sym-dyn)

Position of arm

(vis-con-cont)

**End chosen:**

User Expressive:

Head movement

(hap-sym-dis)

or name option

(aco-lex-dis)

Menu display alters to show following choices:

**Move, MoveArm (amongst others)**

System Expressive:

Menu Display

(vis-lex-cont)

Cursor

(vis-sym-dyn)

Position of arm

(hap-con-cont)

User Receptive:

Menu display

(vis-lex-cont)

Cursor

(vis-sym-dyn)

Position of arm

(vis-con-cont)

**Move chosen:**

User Expressive:

Head movement

(hap-sym-dis)

or name option

(aco-lex-dis)

Menu display alters to show following choices:

**Base, Arm, Shoulder, Elbow, Hand, Wrist, Palm, Grip, End**

System Expressive:

Menu Display

(vis-lex-cont)

Cursor

(vis-sym-dyn)

Position of arm

(hap-con-cont)

User Receptive:

Menu display

(vis-lex-cont)

Cursor

(vis-sym-dyn)

Position of arm

(vis-con-cont)

**Arm chosen:**

User Expressive:  
Head movement  
(hap-sym-dis)  
or name option  
(aco-lex-dis)

Menu display alters to show following choices:

**In, Out, End**

System Expressive:  
Menu Display  
(vis-lex-cont)  
Cursor  
(vis-sym-dyn)  
Position of arm  
(hap-con-cont)

User Receptive:  
Menu display  
(vis-lex-cont)  
Cursor  
(vis-sym-dyn)  
Position of arm  
(vis-con-cont)

**Out chosen:**

User Expressive:  
Head movement  
(hap-sym-dis)  
or name option  
(aco-lex-dis)

Menu display alters to show following choices:

**Go, Speed Level, End**

System Expressive:  
Menu Display  
(vis-lex-cont)  
Cursor  
(vis-sym-dyn)

User Receptive:  
Menu display  
(vis-lex-cont)  
Cursor  
(vis-sym-dyn)

**Speed Level chosen:**

User Expressive:  
Head movement  
(hap-sym-dis)  
or name option  
(aco-lex-dis)



Menu display alters to show following choices:

**Min, Slow, Medium, Fast, Max**

System Expressive:

Menu Display

(vis-lex-cont)

Cursor

(vis-sym-dyn)

User Receptive:

Menu display

(vis-lex-cont)

Cursor

(vis-sym-dyn)

**Min** chosen:

User Expressive:

Head movement

(hap-sym-dis)

or name option

(aco-lex-dis)

Menu display alters to show following choices:

**Go, Speed Level, End**

System Expressive:

Menu Display

(vis-lex-cont)

Cursor

(vis-sym-dyn)

User Receptive:

Menu display

(vis-lex-cont)

Cursor

(vis-sym-dyn)

**Go** chosen:

User Expressive:

Head movement

(hap-sym-dis)

or name option

(aco-lex-dis)

The arm will continue moving until **Stop** is selected:

System Expressive:

Menu Display

(vis-lex-cont)

Cursor

(vis-sym-cont)

Position of arm

(hap-con-dyn)

User Receptive:

Menu display

(vis-lex-cont)

Cursor

(vis-sym-cont)

Position of arm

(vis-con-dyn)

Note: cursor will be under the only option, so has changed from dynamic to continuous.

**Stop chosen:**

User Expressive:

Head movement

(hap-sym-dis)

or name option

(aco-lex-dis)

or the arm reaches the limit of movement.

Menu display alters to show following choices:

**Continue, Speed Level, End**

System Expressive:

Menu Display

(vis-lex-cont)

Cursor

(vis-sym-dyn)

Position of arm

(hap-con-cont)

User Receptive:

Menu display

(vis-lex-cont)

Cursor

(vis-sym-dyn)

Position of arm

(vis-con-cont)

**End chosen:**

User Expressive:

Head movement

(hap-sym-dis)

or name option

(aco-lex-dis)

Menu display alters to show following choices:

**Move, MoveArm (amongst others)**

System Expressive:

Menu Display

(vis-lex-cont)

Cursor

(vis-sym-dyn)

Position of arm

(hap-con-cont)

User Receptive:

Menu display

(vis-lex-cont)

Cursor

(vis-sym-dyn)

Position of arm

(vis-con-cont)



**Move chosen:**

User Expressive:  
Head movement  
(hap-sym-dis)  
or name option  
(aco-lex-dis)

Menu display alters to show following choices:

**Base, Arm, Shoulder, Elbow, Hand, Wrist, Palm, Grip, End**

System Expressive:  
Menu Display  
(vis-lex-cont)  
Cursor  
(vis-sym-dyn)  
Position of arm  
(hap-con-cont)

User Receptive:  
Menu display  
(vis-lex-cont)  
Cursor  
(vis-sym-dyn)  
Position of arm  
(vis-con-cont)

**Grip chosen:**

User Expressive:  
Head movement  
(hap-sym-dis)  
or name option  
(aco-lex-dis)

Menu display alters to show following choices:

**In, Out, End**

System Expressive:  
Menu Display  
(vis-lex-cont)  
Cursor  
(vis-sym-dyn)  
Position of arm  
(hap-con-cont)

User Receptive:  
Menu display  
(vis-lex-cont)  
Cursor  
(vis-sym-dyn)  
Position of arm  
(vis-con-cont)

**Out chosen:**

User Expressive:  
Head movement  
(hap-sym-dis)  
or name option  
(aco-lex-dis)

Menu display alters to show following choices:

**Go, Speed Level, End**

System Expressive:  
Menu Display  
(vis-lex-cont)  
Cursor  
(vis-sym-dyn)

User Receptive:  
Menu display  
(vis-lex-cont)  
Cursor  
(vis-sym-dyn)

**Speed Level chosen:**

User Expressive:  
Head movement  
(hap-sym-dis)  
or name option  
(aco-lex-dis)

Menu display alters to show following choices:

**Min, Slow, Medium, Fast, Max**

System Expressive:  
Menu Display  
(vis-lex-cont)  
Cursor  
(vis-sym-dyn)

User Receptive:  
Menu display  
(vis-lex-cont)  
Cursor  
(vis-sym-dyn)

**Min chosen:**

User Expressive:  
Head movement  
(hap-sym-dis)  
or name option  
(aco-lex-dis)

Menu display alters to show following choices:

**Go, Speed Level, End**

System Expressive:  
Menu Display  
(vis-lex-cont)  
Cursor  
(vis-sym-dyn)

User Receptive:  
Menu display  
(vis-lex-cont)  
Cursor  
(vis-sym-dyn)



**Go chosen:**

User Expressive:  
Head movement  
(hap-sym-dis)  
or name option  
(aco-lex-dis)

The arm will continue moving until **Stop** is selected:

System Expressive:  
Menu Display  
(vis-lex-cont)  
Cursor  
(vis-sym-cont)  
Position of arm  
(hap-con-dyn)

User Receptive:  
Menu display  
(vis-lex-cont)  
Cursor  
(vis-sym-cont)  
Position of arm  
(vis-con-dyn)

Note: cursor will be under the only option, so has changed from dynamic to continuous.

**Stop chosen:**

User Expressive:  
Head movement  
(hap-sym-dis)  
or name option  
(aco-lex-dis)

or the arm reaches the limit of movement.

Menu display alters to show following choices:

**Continue, Speed Level, End**

System Expressive:  
Menu Display  
(vis-lex-cont)  
Cursor  
(vis-sym-dyn)  
Position of arm  
(hap-con-cont)

User Receptive:  
Menu display  
(vis-lex-cont)  
Cursor  
(vis-sym-dyn)  
Position of arm  
(vis-con-cont)

**End chosen:**

User Expressive:

Head movement

(hap-sym-dis)

or name option

(aco-lex-dis)

Menu display alters to show following choices:

**Move, MoveArm** (amongst others)



# **13. APPENDIX C: METHODOLOGY APPLIED TO ROBOTIC ARM**

## **13.1 Stage 1. Define the task that is to be analysed**

The task set was to turn on a light switch.

The action sequence for the task was that the robotic arm needed to be moved to the right, then the arm part moved up and the gripper moved out.

The robotic arm is intended to be used in a domestic context for everyday tasks such as feeding and grooming, so therefore this is the kind of simple task that the robotic arm is designed to support, and can shed considerable light on the usability issues inherent in this interface. This is a non-trivial task which involves moving both the arm as a whole, and two of the individual joints, giving an opportunity to examine the use of the arm in action.

## **13.2 Stage 2. Modality lists**

### **13.2.1 User modalities:**

[UE hap-sym-dis] \*head movement\*

[UE aco-lex-dis] \*name option\*

[UR vis-lex-cont] \*menu display\*

[UR vis-sym-dyn] \*moving cursor\*

[UR vis-con-cont] \*position of arm\*

[UR vis-sym-cont] \*stopped cursor\*

[UR vis-con-cont] \*arm moving\*

### 13.2.2 System modalities:

[SE vis-lex-cont] \*menu display\*

[SE vis-sym-dyn] \*moving cursor\*

[SE hap-con-cont] \*position of arm\*

[SE hap-con-cont] \*moving arm\*

[SE hap-sym-cont] \*stopped cursor\*

[SE hap-con-cont] \*arm stopped\*

[SR hap-sym-dis] \*head movement\*

[SR aco-lex-dis] \*name option\*

### 13.2.3 Granularity:

The level of granularity used in this report is of a high level. The user and system actions are recorded, but not at a low level. For instance, the movement of the cursor is not represented as a series of discrete modalities, where the cursor appears under one option, then appears under another, then appearing under another, but as an ongoing dynamic one. In this way, the movement of the cursor is compressed so that instead of using several stages of interaction to represent it, it can be represented in just one. By choosing this level of granularity the overall usability issues can be regarded. This still gives scope for areas to be investigated at a lower level of granularity if needed. For example, if the motion of the cursor was felt to be of importance, it could be represented as discrete modalities.



### **13.3 Stage 3. Define the user, system and environment profiles**

#### **13.3.1 User Profile:**

The users of the system are people with severe motor disabilities, unable to do elementary physical tasks unaided. These people may also have restricted head movements and voice capabilities. Their intelligence will range right across the scale, from below average to well above average. They may well be inexperienced in interacting with computer interfaces, given the nature of their disabilities, and since there is a lack of such robotic arms, they may not be sure about how to go about operating one. Therefore both in terms of the task and the style of interface the users will probably be inexperienced. However, given that the robotic arm is designed to help them live a more independent life, they will probably be very motivated to learn how to use the system. They are moderately familiar with how the arm operates, and that the arm is composed of several joints, each of which can move separately. They know how to use the input devices.

#### **13.3.2 System Profile:**

The robotic arm consists of eight joints, powered by motors, which can move either individual joints, or the whole arm at once, via the input devices. In the portion of the interface examined, the user is able to set the direction and speed of the arm, chose whether to manipulate all or part of the arm, and initiate and stop arm movement, by means of menu-based software by two possible methods of interaction, that of gesture and speech. At the time the analyses were conducted, none of the input or output devices for the arm controller had been implemented. The interface was to be implemented as a Windows application, with menu options selected from a small LCD display attached to the base of the arm using either voice recognition or gesture. The analysis operates on the assumption that these have been implemented.

The gesture input system is based on a baseball cap with two sensors: one allowing movement forwards and backwards to be detected, the other allowing movement left and right to be detected. The gesture system is presently implemented so that a cursor moves along underneath the menu options continuously in turn, and if the correct gesture is made

when the cursor is underneath a particular option, then that option is selected. A final gesture is an escape option, which automatically stops the arm if the arm is moving, and returns the user to the main menu.

The voice recognition system allows direct menu option selection simply by saying the menu option out loud. It is designed to be trained to individual voices, and needs resetting over time due to the way that voices change. The system has been trained for this particular user.

### **13.3.3 Environment Profile:**

The environment envisioned for this task is that of a quiet room. Since the task is to turn on a light switch, the levels of lighting must be considered potentially poor. It is assumed that there is enough space for the arm to be moved.

## **13.4 Stage 4. Profiles compared to modality listings**

The profiles of the user, system and environment are compared against the lists of modalities in order to check for obvious discrepancies.

### **13.4.1 User profile compared with System modalities:**

May be difficult for user, due to poor movement control, to switch from menu display to arm.

### **13.4.2 System profile compared with User modalities:**

The movement of the user turning from looking at the arm to the menu display, or vice versa, may be mis-interpreted by the system as a gesture command.

### **13.4.3 Environment profile compared with User and System modalities:**

The user needs enough light to be able to see the menus, unless they are lit from the screen, as well as the arm. The motion of the arm may need space: in the environment profile it was assumed that this was available.



## 13.5 Stages 5 and 6. Interaction Sequence listing and clashes etc.

1.

[SE vis-lex-cont]      and      [SE vis-sym-dyn]      and      [SE hap-con-cont]  
\*menu display\*                      \*moving cursor\*                      \*position of arm\*

[UR vis-lex-cont]      and      [UR vis-sym-dyn]  
\*menu display\*                      \*moving cursor\*  
precon: [SE vis-lex-cont]      precon: [SE vis-sym-dyn]  
\*menu display\*                      \*moving cursor\*  
precon: looking at display      precon: looking at display

or

[UR vis-con-cont]  
\*position of arm\*  
precon: [SE hap-con-cont]  
\*position of arm\*  
precon: looking at arm

The system is displaying a menu, underneath which is a flashing cursor moving from one option in turn to another. The arm not moving and is at rest. The user is either looking at the menu with the cursor, or at the arm.

There is a choice for the user between **MoveArm**, which governs the movement of the arm as a whole, and **Move**, which governs the movement of individual joints.

---

2.

[SE vis-lex-cont]      and      [SE vis-sym-dyn]      and      [SE hap-con-cont]  
\*menu display\*                      \*moving cursor\*                      \*position of arm\*

[UR vis-lex-cont]	and	[UR vis-sym-dyn]
*menu display*		*moving cursor*
precon: [SE vis-lex-cont]		precon: [SE vis-sym-dyn]
*menu display*		*moving cursor*
precon: looking at display		precon: looking at display

[UE hap-sym-dis]	or	[UE aco-lex-dis]
*head movement*		*name option*
precon:[UR vis-sym-dyn]		precon: [UR vis-lex-cont]
*moving cursor*		*menu display*
precon: looking at display		

[SR hap-sym-dis]	or	[SR aco-lex-dis]
*head movement*		*name option*
precon: [UE hap-sym-dis]		precon: [UE aco-lex-dis]
*head movement*		*name option*

The menu is still displaying the menu choices, with the cursor moving from underneath one option to underneath another. The arm is at rest and is not moving. The user is looking at the screen and cursor in order to choose the option **MoveArm**. There is no need for a precondition of looking at the display for the UE modality of naming the option because the user may have stopped looking at the display at that point.

3.

[SE vis-lex-cont]	and	[SE vis-sym-dyn]	and	[SE hap-con-cont]
*menu display*		*moving cursor*		*position of arm*
precon:[UE hap-sym-dis]				
*head movement*				

or



[UE aco-lex-dis]

\*name option\*

[UR vis-lex-cont]      and      [UR vis-sym-dyn]

\*menu display\*                      \*moving cursor\*

precon: [SE vis-lex-cont]      precon: [SE vis-sym-dyn]

\*menu display\*                      \*moving cursor\*

precon: looking at display      precon: looking at display

or

[UR vis-con-cont]

\*position of arm\*

precon: [SE hap-con-cont]

\*position of arm\*

precon:              looking at arm

The system is displaying a new menu, underneath which is a flashing cursor moving from one option in turn to another. The arm not moving and is at rest. The user is either looking at the menu with the cursor, or at the arm.

There is a choice for the user between: **Left, Right, Up, Down, Forward, Back, End**

---

4.

[SE vis-lex-cont]                      and      [SE vis-sym-dyn]                      and      [SE hap-con-cont]

\*menu display\*                                      \*moving cursor\*                                      \*position of arm\*

precon:[UE hap-sym-dis]

\*head movement\*

or

[UE aco-lex-dis]

\*name option\*

[UR vis-lex-cont]	and	[UR vis-sym-dyn]
*menu display*		*moving cursor*
precon: [SE vis-lex-cont]		precon: [SE vis-sym-dyn]
*menu display*		*moving cursor*
precon: looking at display		precon: looking at display

[UE hap-sym-dis]	or	[UE aco-lex-dis]
*head movement*		*name option*
precon:[UR vis-sym-dyn]		precon: [UR vis-lex-cont]
*moving cursor*		*menu display*
precon: looking at display		

[SR hap-sym-dis]	or	[SR aco-lex-dis]
*head movement*		*name option*
precon: [UE hap-sym-dis]		precon: [UE aco-lex-dis]
*head movement*		*name option*

The menu is still displaying the menu choices, with the cursor moving from underneath one option to underneath another. The arm is at rest and is not moving. The user is looking at the screen and cursor in order to choose the option **Right**. There is no need for a precondition of looking at the display for the UE modality of naming the option because the user may have stopped looking at the display at that point.

---

5.

[SE vis-lex-cont]	and	[SE vis-sym-dyn]	and	[SE hap-con-cont]
*menu display*		*moving cursor*		*position of arm*
precon:[UE hap-sym-dis]				
*head movement*				

or



[UE aco-lex-dis]

\*name option\*

[UR vis-lex-cont]      and      [UR vis-sym-dyn]

\*menu display\*                      \*moving cursor\*

precon: [SE vis-lex-cont]      precon: [SE vis-sym-dyn]

\*menu display\*                      \*moving cursor\*

precon: looking at display      precon: looking at display

or

[UR vis-con-cont]

\*position of arm\*

precon: [SE hap-con-cont]

\*position of arm\*

precon: looking at arm

The system is displaying a new menu, underneath which is a flashing cursor moving from one option in turn to another. The arm not moving and is at rest. The user is either looking at the menu with the cursor, or at the arm.

There is a choice for the user between: **Go, Speed Level, End**

---

6.

[SE vis-lex-cont]                      and      [SE vis-sym-dyn]                      and      [SE hap-con-cont]

\*menu display\*                                      \*moving cursor\*                                      \*position of arm\*

precon:[UE hap-sym-dis]

\*head movement\*

or

[UE aco-lex-dis]

\*name option\*

[UR vis-lex-cont]	and	[UR vis-sym-dyn]
*menu display*		*moving cursor*
precon: [SE vis-lex-cont]		precon: [SE vis-sym-dyn]
*menu display*		*moving cursor*
precon: looking at display		precon: looking at display

[UE hap-sym-dis]	or	[UE aco-lex-dis]
*head movement*		*name option*
precon:[UR vis-sym-dyn]		precon: [UR vis-lex-cont]
*moving cursor*		*menu display*
precon: looking at display		

[SR hap-sym-dis]	or	[SR aco-lex-dis]
*head movement*		*name option*
precon: [UE hap-sym-dis]		precon: [UE aco-lex-dis]
*head movement*		*name option*

The menu is still displaying the menu choices, with the cursor moving from underneath one option to underneath another. The arm is at rest and is not moving. The user is looking at the screen and cursor in order to choose the option **Speed Level**. There is no need for a precondition of looking at the display for the UE modality of naming the option because the user may have stopped looking at the display at that point.

7.

[SE vis-lex-cont]	and	[SE vis-sym-dyn]	and	[SE hap-con-cont]
*menu display*		*moving cursor*		*position of arm*
precon:[UE hap-sym-dis]				
*head movement*				

or



[UE aco-lex-dis]

\*name option\*

[UR vis-lex-cont]      and      [UR vis-sym-dyn]

\*menu display\*                      \*moving cursor\*

precon: [SE vis-lex-cont]      precon: [SE vis-sym-dyn]

\*menu display\*                      \*moving cursor\*

precon: looking at display      precon: looking at display

or

[UR vis-con-cont]

\*position of arm\*

precon: [SE hap-con-cont]

\*position of arm\*

precon: looking at arm

The system is displaying a new menu, underneath which is a flashing cursor moving from one option in turn to another. The arm not moving and is at rest. The user is either looking at the menu with the cursor, or at the arm.

There is a choice for the user between: **Min, Slow, Medium, Fast, Max**

---

8.

[SE vis-lex-cont]                      and      [SE vis-sym-dyn]                      and      [SE hap-con-cont]

\*menu display\*                                      \*moving cursor\*                                      \*position of arm\*

precon:[UE hap-sym-dis]

\*head movement\*

or

[UE aco-lex-dis]

\*name option\*

[UR vis-lex-cont]	and	[UR vis-sym-dyn]
*menu display*		*moving cursor*
precon: [SE vis-lex-cont]		precon: [SE vis-sym-dyn]
*menu display*		*moving cursor*
precon: looking at display		precon: looking at display

[UE hap-sym-dis]	or	[UE aco-lex-dis]
*head movement*		*name option*
precon:[UR vis-sym-dyn]		precon: [UR vis-lex-cont]
*moving cursor*		*menu display*
precon: looking at display		

[SR hap-sym-dis]	or	[SR aco-lex-dis]
*head movement*		*name option*
precon: [UE hap-sym-dis]		precon: [UE aco-lex-dis]
*head movement*		*name option*

The menu is still displaying the menu choices, with the cursor moving from underneath one option to underneath another. The arm is at rest and is not moving. The user is looking at the screen and cursor in order to choose the option **Min**. There is no need for a precondition of looking at the display for the UE modality of naming the option because the user may have stopped looking at the display at that point.

---

9.

[SE vis-lex-cont]	and	[SE vis-sym-dyn]	and	[SE hap-con-cont]
*menu display*		*moving cursor*		*position of arm*
precon:[UE hap-sym-dis]				
*head movement*				

or



[UE aco-lex-dis]

\*name option\*

[UR vis-lex-cont]      and      [UR vis-sym-dyn]

\*menu display\*                      \*moving cursor\*

precon: [SE vis-lex-cont]      precon: [SE vis-sym-dyn]

\*menu display\*                      \*moving cursor\*

precon: looking at display      precon: looking at display

or

[UR vis-con-cont]

\*position of arm\*

precon: [SE hap-con-cont]

\*position of arm\*

precon: looking at arm

The system is displaying a new menu, underneath which is a flashing cursor moving from one option in turn to another. The arm not moving and is at rest. The user is either looking at the menu with the cursor, or at the arm.

There is a choice for the user between: **Go, Speed Level, End**

---

10.

[SE vis-lex-cont]                      and      [SE vis-sym-dyn]                      and      [SE hap-con-cont]

\*menu display\*                                      \*moving cursor\*                                      \*position of arm\*

precon:[UE hap-sym-dis]

\*head movement\*

or

[UE aco-lex-dis]

\*name option\*

[UR vis-lex-cont]	and	[UR vis-sym-dyn]
*menu display*		*moving cursor*
precon: [SE vis-lex-cont]		precon: [SE vis-sym-dyn]
*menu display*		*moving cursor*
precon: looking at display		precon: looking at display

[UE hap-sym-dis]	or	[UE aco-lex-dis]
*head movement*		*name option*
precon:[UR vis-sym-dyn]		precon: [UR vis-lex-cont]
*moving cursor*		*menu display*
precon: looking at display		

[SR hap-sym-dis]	or	[SR aco-lex-dis]
*head movement*		*name option*
precon: [UE hap-sym-dis]		precon: [UE aco-lex-dis]
*head movement*		*name option*

The menu is still displaying the menu choices, with the cursor moving from underneath one option to underneath another. The arm is at rest and is not moving. The user is looking at the screen and cursor in order to choose the option **Go**. There is no need for a precondition of looking at the display for the UE modality of naming the option because the user may have stopped looking at the display at that point.

---

11.

[SE vis-lex-cont]	and	[SE vis-sym-cont]	and	[SE hap-con-cont]
*menu display*		*stopped cursor*		*arm moving*
precon:[UE hap-sym-dis]				
*head movement*				

or



[UE aco-lex-dis]

\*name option\*

[UR vis-lex-cont]      and      [UR vis-sym-cont]

\*menu display\*                      \*stopped cursor\*

precon: [SE vis-lex-cont]      precon: [SE vis-sym-cont]

\*menu display\*                      \*stopped cursor\*

precon: looking at display      precon: looking at display

or

[UR vis-con-cont]

\*arm moving\*

precon: [SE hap-con-cont]

\*arm moving\*

precon: looking at arm

The arm is in motion. The arm will continue moving until **Stop** is selected. The cursor is now under the only option, so has changed from dynamic to continuous. The user is either looking at the movement of the arm, or at the display.

---

12.

[SE vis-lex-cont]      and      [SE vis-sym-cont]      and      [SE hap-con-cont]

\*menu display\*                      \*stopped cursor\*                      \*arm moving\*

precon:[UE hap-sym-dis]

\*head movement\*

or

[UE aco-lex-dis]

\*name option\*

[UR vis-lex-cont]	and	[UR vis-sym-cont]
*menu display*		*stopped cursor*
precon: [SE vis-lex-cont]		precon: [SE vis-sym-cont]
*menu display*		*stopped cursor*
precon: looking at display		precon: looking at display

or

[UR vis-con-cont]  
 \*arm moving\*  
 precon: [SE hap-con-cont]  
 \*arm moving\*  
 precon: looking at arm

[UE hap-sym-dis]	or	[UE aco-lex-dis]
*head movement*		*name option*
		precon: [SE vis-lex-cont]
		*menu display*

[SR hap-sym-dis]	or	[SR aco-lex-dis]
*head movement*		*name option*
precon:[UE hap-sym-dis]		precon: [UE aco-lex-dis]
*head movement*		*name option*

or

[SE hap-con-cont]  
 \*arm stops\*  
 precon: [SE hap-con-cont]  
 \*arm moving\*

The arm is in motion until the user either chooses **Stop** or the arm reaches the limit of movement. The UE modality of head movement has no precondition, because the user does not have to be looking at the menu display, since there is only option available at this point. The user can therefore be looking at the arm if needed. Also, since only one option



available, the user may not have to read the menu interface to say “Stop”, since again there is only one option. Therefore the precondition for the UE modality of saying “Stop” is not definite. The system will either accept the UE modalities or the arm will reach the limit of motion.

---

13.

[SE vis-lex-cont]	and	[SE vis-sym-dyn]	and	[SE hap-con-cont]
*menu display*		*moving cursor*		*position of arm*
precon:[SR hap-sym-dis]		precon:[SR hap-sym-dis]		precon:[SR hap-sym-dis]
*head movement*		*head movement*		*head movement*
or		or		or
[SR aco-lex-dis]		[SR aco-lex-dis]		[SR aco-lex-dis]
*name option*		*name option*		*name option*
or		or		or
[SE hap-con-cont]		[SE hap-con-cont]		[SE hap-con-cont]
*arm stops *		*arm stops *		*arm stops *

[UR vis-lex-cont]	and	[UR vis-sym-dyn]
*menu display*		*moving cursor*
precon: [SE vis-lex-cont]		precon: [SE vis-sym-dyn]
*menu display*		*moving cursor*
precon: looking at display		precon: looking at display

or

[UR vis-con-cont]
*position of arm*
precon: [SE hap-con-cont]
*position of arm*
precon: looking at arm

The arm has stopped moving and the system is displaying a new menu, underneath which the flashing cursor is once more moving from one option in turn to another. The user is either looking at the menu with the cursor, or at the arm.

There is a choice for the user between **Continue, Speed Level, End**

---

14.

[SE vis-lex-cont]	and	[SE vis-sym-dyn]	and	[SE hap-con-cont]
*menu display*		*moving cursor*		*position of arm*

precon:[UE hap-sym-dis]  
 \*head movement\*

or

[UE aco-lex-dis]  
 \*name option\*

[UR vis-lex-cont]	and	[UR vis-sym-dyn]
*menu display*		*moving cursor*
precon: [SE vis-lex-cont]		precon: [SE vis-sym-dyn]
*menu display*		*moving cursor*
precon: looking at display		precon: looking at display

[UE hap-sym-dis]	or	[UE aco-lex-dis]
*head movement*		*name option*
precon:[UR vis-sym-dyn]		precon: [UR vis-lex-cont]
*moving cursor*		*menu display*
precon: looking at display		

[SR hap-sym-dis]	or	[SR aco-lex-dis]
*head movement*		*name option*
precon: [UE hap-sym-dis]		precon: [UE aco-lex-dis]
*head movement*		*name option*



The menu is still displaying the menu choices, with the cursor moving from underneath one option to underneath another. The arm is at rest and is not moving. The user is looking at the screen and cursor in order to choose the option **End**. There is no need for a precondition of looking at the display for the UE modality of naming the option because the user may have stopped looking at the display at that point.

---

15.

[SE vis-lex-cont]                      and      [SE vis-sym-dyn]                      and      [SE hap-con-cont]  
 \*menu display\*                                              \*moving cursor\*                                              \*position of arm\*

precon:[UE hap-sym-dis]

\*head movement\*

or

[UE aco-lex-dis]

\*name option\*

[UR vis-lex-cont]              and      [UR vis-sym-dyn]

\*menu display\*                                              \*moving cursor\*

precon: [SE vis-lex-cont]              precon: [SE vis-sym-dyn]

\*menu display\*                                              \*moving cursor\*

precon: looking at display              precon: looking at display

or

[UR vis-con-cont]

\*position of arm\*

precon: [SE hap-con-cont]

\*position of arm\*

precon: looking at arm

The system is displaying the original menu, underneath which is a flashing cursor moving from one option in turn to another. The arm not moving and is at rest. The user is either looking at the menu with the cursor, or at the arm.





precondition of looking at the display for the UE modality of naming the option because the user may have stopped looking at the display at that point.

---

17.

[SE vis-lex-cont]                      and      [SE vis-sym-dyn]                      and      [SE hap-con-cont]  
\*menu display\*                                              \*moving cursor\*                                              \*position of arm\*

precon:[UE hap-sym-dis]

\*head movement\*

or

[UE aco-lex-dis]

\*name option\*

[UR vis-lex-cont]                      and      [UR vis-sym-dyn]

\*menu display\*                                              \*moving cursor\*

precon: [SE vis-lex-cont]                      precon: [SE vis-sym-dyn]

\*menu display\*                                              \*moving cursor\*

precon: looking at display                      precon: looking at display

or

[UR vis-con-cont]

\*position of arm\*

precon: [SE hap-con-cont]

\*position of arm\*

precon: looking at arm

The system is displaying a new menu, underneath which is a flashing cursor moving from one option in turn to another. The arm not moving and is at rest. The user is either looking at the menu with the cursor, or at the arm.

There is a choice for the user between: **Base, Arm, Shoulder, Elbow, Hand, Wrist, Palm, Grip, End**

---

18.

[SE vis-lex-cont]                      and    [SE vis-sym-dyn]                      and    [SE hap-con-cont]  
\*menu display\*                                              \*moving cursor\*                                              \*position of arm\*

precon:[UE hap-sym-dis]  
\*head movement\*

or

[UE aco-lex-dis]  
\*name option\*

[UR vis-lex-cont]                      and    [UR vis-sym-dyn]  
\*menu display\*                                              \*moving cursor\*  
precon: [SE vis-lex-cont]                      precon: [SE vis-sym-dyn]  
\*menu display\*                                              \*moving cursor\*  
precon: looking at display                      precon: looking at display

[UE hap-sym-dis]                      or    [UE aco-lex-dis]  
\*head movement\*                                              \*name option\*  
precon:[UR vis-sym-dyn]                      precon: [UR vis-lex-cont]  
\*moving cursor\*                                              \*menu display\*  
precon: looking at display

[SR hap-sym-dis]                      or    [SR aco-lex-dis]  
\*head movement\*                                              \*name option\*  
precon: [UE hap-sym-dis]                      precon: [UE aco-lex-dis]  
\*head movement\*                                              \*name option\*

The menu is still displaying the menu choices, with the cursor moving from underneath one option to underneath another. The arm is at rest and is not moving. The user is looking at the screen and cursor in order to choose the option Arm. There is no need for a precondition of looking at the display for the UE modality of naming the option because the user may have stopped looking at the display at that point.



---

19.

[SE vis-lex-cont]                      and    [SE vis-sym-dyn]                      and    [SE hap-con-cont]  
\*menu display\*                                              \*moving cursor\*                                              \*position of arm\*

precon:[UE hap-sym-dis]  
\*head movement\*

or

[UE aco-lex-dis]  
\*name option\*

[UR vis-lex-cont]                      and    [UR vis-sym-dyn]  
\*menu display\*                                              \*moving cursor\*  
precon: [SE vis-lex-cont]                      precon: [SE vis-sym-dyn]  
\*menu display\*                                              \*moving cursor\*  
precon: looking at display                      precon: looking at display

or

                    [UR vis-con-cont]  
                    \*position of arm\*  
                    precon: [SE hap-con-cont]  
                    \*position of arm\*  
                    precon: looking at arm

The system is displaying a new menu, underneath which is a flashing cursor moving from one option in turn to another. The arm not moving and is at rest. The user is either looking at the menu with the cursor, or at the arm.

There is a choice for the user between: **In, Out, End**

---

19.

[SE vis-lex-cont]                      and    [SE vis-sym-dyn]                      and    [SE hap-con-cont]

*menu display*		*moving cursor*		*position of arm*
precon:[UE hap-sym-dis]				
*head movement*				
or				
[UE aco-lex-dis]				
*name option*				
[UR vis-lex-cont]	and	[UR vis-sym-dyn]		
*menu display*		*moving cursor*		
precon: [SE vis-lex-cont]		precon: [SE vis-sym-dyn]		
*menu display*		*moving cursor*		
precon: looking at display		precon: looking at display		
[UE hap-sym-dis]	or	[UE aco-lex-dis]		
*head movement*		*name option*		
precon:[UR vis-sym-dyn]		precon: [UR vis-lex-cont]		
*moving cursor*		*menu display*		
precon: looking at display				
[SR hap-sym-dis]	or	[SR aco-lex-dis]		
*head movement*		*name option*		
precon: [UE hap-sym-dis]		precon: [UE aco-lex-dis]		
*head movement*		*name option*		

The menu is still displaying the menu choices, with the cursor moving from underneath one option to underneath another. The arm is at rest and is not moving. The user is looking at the screen and cursor in order to choose the option **Out**. There is no need for a precondition of looking at the display for the UE modality of naming the option because the user may have stopped looking at the display at that point.



[SE vis-lex-cont]                      and    [SE vis-sym-dyn]                      and    [SE hap-con-cont]  
\*menu display\*                                              \*moving cursor\*                                              \*position of arm\*

precon:[UE hap-sym-dis]

\*head movement\*

or

[UE aco-lex-dis]

\*name option\*

[UR vis-lex-cont]                      and    [UR vis-sym-dyn]

\*menu display\*                                              \*moving cursor\*

precon: [SE vis-lex-cont]                      precon: [SE vis-sym-dyn]

\*menu display\*                                              \*moving cursor\*

precon: looking at display                      precon: looking at display

or

[UR vis-con-cont]

\*position of arm\*

precon: [SE hap-con-cont]

\*position of arm\*

precon: looking at arm

The system is displaying a new menu, underneath which is a flashing cursor moving from one option in turn to another. The arm not moving and is at rest. The user is either looking at the menu with the cursor, or at the arm.

There is a choice for the user between: **Go, Speed Level, End**

---

21.

[SE vis-lex-cont]                      and    [SE vis-sym-dyn]                      and    [SE hap-con-cont]  
\*menu display\*                                              \*moving cursor\*                                              \*position of arm\*

precon:[UE hap-sym-dis]

\*head movement\*

or

[UE aco-lex-dis]

\*name option\*

[UR vis-lex-cont]	and	[UR vis-sym-dyn]
*menu display*		*moving cursor*
precon: [SE vis-lex-cont]		precon: [SE vis-sym-dyn]
*menu display*		*moving cursor*
precon: looking at display		precon: looking at display

[UE hap-sym-dis]	or	[UE aco-lex-dis]
*head movement*		*name option*
precon:[UR vis-sym-dyn]		precon: [UR vis-lex-cont]
*moving cursor*		*menu display*
precon: looking at display		

[SR hap-sym-dis]	or	[SR aco-lex-dis]
*head movement*		*name option*
precon: [UE hap-sym-dis]		precon: [UE aco-lex-dis]
*head movement*		*name option*

The menu is still displaying the menu choices, with the cursor moving from underneath one option to underneath another. The arm is at rest and is not moving. The user is looking at the screen and cursor in order to choose the option **Speed Level**. There is no need for a precondition of looking at the display for the UE modality of naming the option because the user may have stopped looking at the display at that point.

---

22.

[SE vis-lex-cont]	and	[SE vis-sym-dyn]	and	[SE hap-con-cont]
*menu display*		*moving cursor*		*position of arm*



precon:[UE hap-sym-dis]

\*head movement\*

or

[UE aco-lex-dis]

\*name option\*

[UR vis-lex-cont]      and      [UR vis-sym-dyn]

\*menu display\*

\*moving cursor\*

precon: [SE vis-lex-cont]

precon: [SE vis-sym-dyn]

\*menu display\*

\*moving cursor\*

precon: looking at display

precon: looking at display

or

[UR vis-con-cont]

\*position of arm\*

precon: [SE hap-con-cont]

\*position of arm\*

precon: looking at arm

The system is displaying a new menu, underneath which is a flashing cursor moving from one option in turn to another. The arm not moving and is at rest. The user is either looking at the menu with the cursor, or at the arm.

There is a choice for the user between: **Min, Slow, Medium, Fast, Max**

---

23.

[SE vis-lex-cont]

and

[SE vis-sym-dyn]

and

[SE hap-con-cont]

\*menu display\*

\*moving cursor\*

\*position of arm\*

precon:[UE hap-sym-dis]

\*head movement\*

or

[UE aco-lex-dis]

\*name option\*

[UR vis-lex-cont]      and      [UR vis-sym-dyn]

\*menu display\*                      \*moving cursor\*

precon: [SE vis-lex-cont]      precon: [SE vis-sym-dyn]

\*menu display\*                      \*moving cursor\*

precon: looking at display      precon: looking at display

[UE hap-sym-dis]      or      [UE aco-lex-dis]

\*head movement\*                      \*name option\*

precon:[UR vis-sym-dyn]      precon: [UR vis-lex-cont]

\*moving cursor\*                      \*menu display\*

precon: looking at display

[SR hap-sym-dis]      or      [SR aco-lex-dis]

\*head movement\*                      \*name option\*

precon: [UE hap-sym-dis]      precon: [UE aco-lex-dis]

\*head movement\*                      \*name option\*

The menu is still displaying the menu choices, with the cursor moving from underneath one option to underneath another. The arm is at rest and is not moving. The user is looking at the screen and cursor in order to choose the option **Min**. There is no need for a precondition of looking at the display for the UE modality of naming the option because the user may have stopped looking at the display at that point.

---

24.

[SE vis-lex-cont]                      and      [SE vis-sym-dyn]                      and      [SE hap-con-cont]

\*menu display\*                                      \*moving cursor\*                                      \*position of arm\*

precon:[UE hap-sym-dis]

\*head movement\*



or

[UE aco-lex-dis]

\*name option\*

[UR vis-lex-cont]      and      [UR vis-sym-dyn]

\*menu display\*                      \*moving cursor\*

precon: [SE vis-lex-cont]      precon: [SE vis-sym-dyn]

\*menu display\*                      \*moving cursor\*

precon: looking at display      precon: looking at display

or

[UR vis-con-cont]

\*position of arm\*

precon: [SE hap-con-cont]

\*position of arm\*

precon: looking at arm

The system is displaying a new menu, underneath which is a flashing cursor moving from one option in turn to another. The arm not moving and is at rest. The user is either looking at the menu with the cursor, or at the arm.

There is a choice for the user between: **Go, Speed Level, End**

---

25.

[SE vis-lex-cont]                      and      [SE vis-sym-dyn]                      and      [SE hap-con-cont]

\*menu display\*                                      \*moving cursor\*                                      \*position of arm\*

precon:[UE hap-sym-dis]

\*head movement\*

or

[UE aco-lex-dis]

\*name option\*

[UR vis-lex-cont]	and	[UR vis-sym-dyn]
*menu display*		*moving cursor*
precon: [SE vis-lex-cont]		precon: [SE vis-sym-dyn]
*menu display*		*moving cursor*
precon: looking at display		precon: looking at display

[UE hap-sym-dis]	or	[UE aco-lex-dis]
*head movement*		*name option*
precon:[UR vis-sym-dyn]		precon: [UR vis-lex-cont]
*moving cursor*		*menu display*
precon: looking at display		

[SR hap-sym-dis]	or	[SR aco-lex-dis]
*head movement*		*name option*
precon: [UE hap-sym-dis]		precon: [UE aco-lex-dis]
*head movement*		*name option*

The menu is still displaying the menu choices, with the cursor moving from underneath one option to underneath another. The arm is at rest and is not moving. The user is looking at the screen and cursor in order to choose the option **Go**. There is no need for a precondition of looking at the display for the UE modality of naming the option because the user may have stopped looking at the display at that point.

---

26.

[SE vis-lex-cont]	and	[SE vis-sym-cont]	and	[SE hap-con-cont]
*menu display*		*stopped cursor*		*arm moving*
precon:[UE hap-sym-dis]				
*head movement*				
	or			
[UE aco-lex-dis]				



\*name option\*

[UR vis-lex-cont]      and      [UR vis-sym-cont]  
\*menu display\*                      \*stopped cursor\*  
precon: [SE vis-lex-cont]      precon: [SE vis-sym-cont]  
\*menu display\*                      \*stopped cursor\*  
precon: looking at display      precon: looking at display

or

[UR vis-con-cont]  
\*arm moving\*  
precon: [SE hap-con-cont]  
\*arm moving\*  
precon: looking at arm

The arm is in motion. The arm will continue moving until **Stop** is selected. The cursor is now under the only option, so has changed from dynamic to continuous. The user is either looking at the movement of the arm, or at the display.

---

27.

[SE vis-lex-cont]      and      [SE vis-sym-cont]      and      [SE hap-con-cont]  
\*menu display\*                      \*stopped cursor\*                      \*arm moving\*  
precon:[UE hap-sym-dis]  
\*head movement\*

or

[UE aco-lex-dis]  
\*name option\*

[UR vis-lex-cont]      and      [UR vis-sym-cont]

<i>*menu display*</i>	<i>*stopped cursor*</i>
precon: [SE vis-lex-cont]	precon: [SE vis-sym-cont]
<i>*menu display*</i>	<i>*stopped cursor*</i>
precon: looking at display	precon: looking at display

or

[UR vis-con-cont]  
*\*arm moving\**  
precon: [SE hap-con-cont]  
*\*arm moving\**  
precon: looking at arm

[UE hap-sym-dis]	or	[UE aco-lex-dis]
<i>*head movement*</i>		<i>*name option*</i>
		precon: [SE vis-lex-cont]
		<i>*menu display*</i>

[SR hap-sym-dis]	or	[SR aco-lex-dis]
<i>*head movement*</i>		<i>*name option*</i>
precon:[UE hap-sym-dis]		precon: [UE aco-lex-dis]
<i>*head movement*</i>		<i>*name option*</i>

or

[SE hap-con-cont]  
*\*arm stops\**  
precon: [SE hap-con-cont]  
*\*arm moving\**

The arm is in motion until the user either chooses **Stop** or the arm reaches the limit of movement. The UE modality of head movement has no precondition, because the user does not have to be looking at the menu display, since there is only option available at this point. The user can therefore be looking at the arm if needed. Also, since only one option available, the user may not have to read the menu interface to say “Stop”, since again there



is only one option. Therefore the precondition for the UE modality of saying “Stop” is not definite. The system will either accept the UE modalities or the arm will reach the limit of motion.

---

28.

[SE vis-lex-cont]	and	[SE vis-sym-dyn]	and	[SE hap-con-cont]
*menu display*		*moving cursor*		*position of arm*
precon:[SR hap-sym-dis]		precon:[SR hap-sym-dis]		precon:[SR hap-sym-dis]
*head movement*		*head movement*		*head movement*
or		or		or
[SR aco-lex-dis]		[SR aco-lex-dis]		[SR aco-lex-dis]
*name option*		*name option*		*name option*
or		or		or
[SE hap-con-cont]		[SE hap-con-cont]		[SE hap-con-cont]
*arm stops *		*arm stops *		*arm stops *

[UR vis-lex-cont]	and	[UR vis-sym-dyn]
*menu display*		*moving cursor*
precon: [SE vis-lex-cont]		precon: [SE vis-sym-dyn]
*menu display*		*moving cursor*
precon: looking at display		precon: looking at display

or

[UR vis-con-cont]
*position of arm*
precon: [SE hap-con-cont]
*position of arm*
precon: looking at arm

The arm has stopped moving and the system is displaying a new menu, underneath which the flashing cursor is once more moving from one option in turn to another. The user is either looking at the menu with the cursor, or at the arm.

There is a choice for the user between **Continue**, **Speed Level**, **End**

---

29.

[SE vis-lex-cont]                      and    [SE vis-sym-dyn]                      and    [SE hap-con-cont]  
\*menu display\*                                              \*moving cursor\*                                              \*position of arm\*

precon:[UE hap-sym-dis]  
\*head movement\*

or

[UE aco-lex-dis]  
\*name option\*

[UR vis-lex-cont]                      and    [UR vis-sym-dyn]  
\*menu display\*                                              \*moving cursor\*  
precon: [SE vis-lex-cont]                      precon: [SE vis-sym-dyn]  
\*menu display\*                                              \*moving cursor\*  
precon: looking at display                      precon: looking at display

[UE hap-sym-dis]                      or    [UE aco-lex-dis]  
\*head movement\*                                              \*name option\*  
precon:[UR vis-sym-dyn]                      precon: [UR vis-lex-cont]  
\*moving cursor\*                                              \*menu display\*  
precon: looking at display

[SR hap-sym-dis]                      or    [SR aco-lex-dis]  
\*head movement\*                                              \*name option\*  
precon: [UE hap-sym-dis]                      precon: [UE aco-lex-dis]  
\*head movement\*                                              \*name option\*

The menu is still displaying the menu choices, with the cursor moving from underneath one option to underneath another. The arm is at rest and is not moving. The user is looking at



the screen and cursor in order to choose the option **End**. There is no need for a precondition of looking at the display for the UE modality of naming the option because the user may have stopped looking at the display at that point.

---

30.

[SE vis-lex-cont]                      and    [SE vis-sym-dyn]                      and    [SE hap-con-cont]  
\*menu display\*                                              \*moving cursor\*                                              \*position of arm\*

precon:[UE hap-sym-dis]

\*head movement\*

or

[UE aco-lex-dis]

\*name option\*

[UR vis-lex-cont]                      and    [UR vis-sym-dyn]

\*menu display\*                                              \*moving cursor\*

precon: [SE vis-lex-cont]                      precon: [SE vis-sym-dyn]

\*menu display\*                                              \*moving cursor\*

precon: looking at display                      precon: looking at display

or

[UR vis-con-cont]

\*position of arm\*

precon: [SE hap-con-cont]

\*position of arm\*

precon: looking at arm

The system is displaying the original menu, underneath which is a flashing cursor moving from one option in turn to another. The arm not moving and is at rest. The user is either looking at the menu with the cursor, or at the arm.

There is a choice for the user between: **Move**, **MoveArm** (amongst others)

---

31.

[SE vis-lex-cont]                      and    [SE vis-sym-dyn]                      and    [SE hap-con-cont]  
\*menu display\*                                              \*moving cursor\*                                              \*position of arm\*

precon:[UE hap-sym-dis]  
\*head movement\*

or

[UE aco-lex-dis]  
\*name option\*

[UR vis-lex-cont]	and	[UR vis-sym-dyn]
*menu display*		*moving cursor*
precon: [SE vis-lex-cont]		precon: [SE vis-sym-dyn]
*menu display*		*moving cursor*
precon: looking at display		precon: looking at display

[UE hap-sym-dis]	or	[UE aco-lex-dis]
*head movement*		*name option*
precon:[UR vis-sym-dyn]		precon: [UR vis-lex-cont]
*moving cursor*		*menu display*
precon: looking at display		

[SR hap-sym-dis]	or	[SR aco-lex-dis]
*head movement*		*name option*
precon: [UE hap-sym-dis]		precon: [UE aco-lex-dis]
*head movement*		*name option*

The menu is still displaying the menu choices, with the cursor moving from underneath one option to underneath another. The arm is at rest and is not moving. The user is looking at the screen and cursor in order to choose the option **Move**. There is no need for a



precondition of looking at the display for the UE modality of naming the option because the user may have stopped looking at the display at that point.

---

32.

[SE vis-lex-cont]                      and    [SE vis-sym-dyn]                      and    [SE hap-con-cont]  
\*menu display\*                                              \*moving cursor\*                                              \*position of arm\*

precon:[UE hap-sym-dis]

\*head movement\*

or

[UE aco-lex-dis]

\*name option\*

[UR vis-lex-cont]                      and    [UR vis-sym-dyn]

\*menu display\*                                              \*moving cursor\*

precon: [SE vis-lex-cont]                      precon: [SE vis-sym-dyn]

\*menu display\*                                              \*moving cursor\*

precon: looking at display                      precon: looking at display

or

[UR vis-con-cont]

\*position of arm\*

precon: [SE hap-con-cont]

\*position of arm\*

precon: looking at arm

The system is displaying a new menu, underneath which is a flashing cursor moving from one option in turn to another. The arm not moving and is at rest. The user is either looking at the menu with the cursor, or at the arm.

There is a choice for the user between: **Base, Arm, Shoulder, Elbow, Hand, Wrist, Palm, Grip, End**

32.

[SE vis-lex-cont]                      and    [SE vis-sym-dyn]                      and    [SE hap-con-cont]  
\*menu display\*                                              \*moving cursor\*                                              \*position of arm\*

precon:[UE hap-sym-dis]

\*head movement\*

or

[UE aco-lex-dis]

\*name option\*

[UR vis-lex-cont]	and	[UR vis-sym-dyn]
*menu display*		*moving cursor*
precon: [SE vis-lex-cont]		precon: [SE vis-sym-dyn]
*menu display*		*moving cursor*
precon: looking at display		precon: looking at display

[UE hap-sym-dis]	or	[UE aco-lex-dis]
*head movement*		*name option*
precon:[UR vis-sym-dyn]		precon: [UR vis-lex-cont]
*moving cursor*		*menu display*
precon: looking at display		

[SR hap-sym-dis]	or	[SR aco-lex-dis]
*head movement*		*name option*
precon: [UE hap-sym-dis]		precon: [UE aco-lex-dis]
*head movement*		*name option*

The menu is still displaying the menu choices, with the cursor moving from underneath one option to underneath another. The arm is at rest and is not moving. The user is looking at the screen and cursor in order to choose the option **Grip**. There is no need for a



precondition of looking at the display for the UE modality of naming the option because the user may have stopped looking at the display at that point.

---

33.

[SE vis-lex-cont]                      and    [SE vis-sym-dyn]                      and    [SE hap-con-cont]  
\*menu display\*                                              \*moving cursor\*                                              \*position of arm\*

precon:[UE hap-sym-dis]  
\*head movement\*

or

[UE aco-lex-dis]  
\*name option\*

[UR vis-lex-cont]                      and    [UR vis-sym-dyn]  
\*menu display\*                                              \*moving cursor\*  
precon: [SE vis-lex-cont]                      precon: [SE vis-sym-dyn]  
\*menu display\*                                              \*moving cursor\*  
precon: looking at display                      precon: looking at display

or

[UR vis-con-cont]  
\*position of arm\*  
precon: [SE hap-con-cont]  
\*position of arm\*  
precon: looking at arm

The system is displaying a new menu, underneath which is a flashing cursor moving from one option in turn to another. The arm not moving and is at rest. The user is either looking at the menu with the cursor, or at the arm.

There is a choice for the user between: **In, Out, End**

---

34.

[SE vis-lex-cont]                      and    [SE vis-sym-dyn]                      and    [SE hap-con-cont]  
\*menu display\*                                              \*moving cursor\*                                              \*position of arm\*

precon:[UE hap-sym-dis]  
\*head movement\*

or

[UE aco-lex-dis]  
\*name option\*

[UR vis-lex-cont]	and	[UR vis-sym-dyn]
*menu display*		*moving cursor*
precon: [SE vis-lex-cont]		precon: [SE vis-sym-dyn]
*menu display*		*moving cursor*
precon: looking at display		precon: looking at display

[UE hap-sym-dis]	or	[UE aco-lex-dis]
*head movement*		*name option*
precon:[UR vis-sym-dyn]		precon: [UR vis-lex-cont]
*moving cursor*		*menu display*
precon: looking at display		

[SR hap-sym-dis]	or	[SR aco-lex-dis]
*head movement*		*name option*
precon: [UE hap-sym-dis]		precon: [UE aco-lex-dis]
*head movement*		*name option*

The menu is still displaying the menu choices, with the cursor moving from underneath one option to underneath another. The arm is at rest and is not moving. The user is looking at the screen and cursor in order to choose the option **Out**. There is no need for a precondition



of looking at the display for the UE modality of naming the option because the user may have stopped looking at the display at that point.

---

35.

[SE vis-lex-cont]                      and      [SE vis-sym-dyn]                      and      [SE hap-con-cont]  
\*menu display\*                                              \*moving cursor\*                                              \*position of arm\*

precon:[UE hap-sym-dis]

\*head movement\*

or

[UE aco-lex-dis]

\*name option\*

[UR vis-lex-cont]                      and      [UR vis-sym-dyn]  
\*menu display\*                                              \*moving cursor\*  
precon: [SE vis-lex-cont]                      precon: [SE vis-sym-dyn]  
\*menu display\*                                              \*moving cursor\*  
precon: looking at display                      precon: looking at display

or

[UR vis-con-cont]

\*position of arm\*

precon: [SE hap-con-cont]

\*position of arm\*

precon: looking at arm

The system is displaying a new menu, underneath which is a flashing cursor moving from one option in turn to another. The arm not moving and is at rest. The user is either looking at the menu with the cursor, or at the arm.

There is a choice for the user between: **Go, Speed Level, End**

---

36.

[SE vis-lex-cont]                      and    [SE vis-sym-dyn]                      and    [SE hap-con-cont]  
\*menu display\*                                              \*moving cursor\*                                              \*position of arm\*

precon:[UE hap-sym-dis]

\*head movement\*

or

[UE aco-lex-dis]

\*name option\*

[UR vis-lex-cont]                      and    [UR vis-sym-dyn]  
\*menu display\*                                              \*moving cursor\*  
precon: [SE vis-lex-cont]                      precon: [SE vis-sym-dyn]  
\*menu display\*                                              \*moving cursor\*  
precon: looking at display                      precon: looking at display

[UE hap-sym-dis]                      or    [UE aco-lex-dis]  
\*head movement\*                                              \*name option\*  
precon:[UR vis-sym-dyn]                      precon: [UR vis-lex-cont]  
\*moving cursor\*                                              \*menu display\*  
precon: looking at display

[SR hap-sym-dis]                      or    [SR aco-lex-dis]  
\*head movement\*                                              \*name option\*  
precon: [UE hap-sym-dis]                      precon: [UE aco-lex-dis]  
\*head movement\*                                              \*name option\*



The menu is still displaying the menu choices, with the cursor moving from underneath one option to underneath another. The arm is at rest and is not moving. The user is looking at the screen and cursor in order to choose the option **Speed Level**. There is no need for a precondition of looking at the display for the UE modality of naming the option because the user may have stopped looking at the display at that point.

---

37.

[SE vis-lex-cont]                      and      [SE vis-sym-dyn]                      and      [SE hap-con-cont]  
\*menu display\*                                              \*moving cursor\*                                              \*position of arm\*

precon:[UE hap-sym-dis]

\*head movement\*

or

[UE aco-lex-dis]

\*name option\*

[UR vis-lex-cont]                      and      [UR vis-sym-dyn]

\*menu display\*                                              \*moving cursor\*

precon: [SE vis-lex-cont]                      precon: [SE vis-sym-dyn]

\*menu display\*                                              \*moving cursor\*

precon: looking at display                      precon: looking at display

or

[UR vis-con-cont]

\*position of arm\*

precon: [SE hap-con-cont]

\*position of arm\*

precon: looking at arm

The system is displaying a new menu, underneath which is a flashing cursor moving from one option in turn to another. The arm not moving and is at rest. The user is either looking at the menu with the cursor, or at the arm.

There is a choice for the user between: **Min, Slow, Medium, Fast, Max**

---

38.

[SE vis-lex-cont]	and	[SE vis-sym-dyn]	and	[SE hap-con-cont]
*menu display*		*moving cursor*		*position of arm*

precon:[UE hap-sym-dis]

\*head movement\*

or

[UE aco-lex-dis]

\*name option\*

[UR vis-lex-cont]	and	[UR vis-sym-dyn]
*menu display*		*moving cursor*
precon: [SE vis-lex-cont]		precon: [SE vis-sym-dyn]
*menu display*		*moving cursor*
precon: looking at display		precon: looking at display

[UE hap-sym-dis]	or	[UE aco-lex-dis]
*head movement*		*name option*
precon:[UR vis-sym-dyn]		precon: [UR vis-lex-cont]
*moving cursor*		*menu display*
precon: looking at display		

[SR hap-sym-dis]	or	[SR aco-lex-dis]
*head movement*		*name option*
precon: [UE hap-sym-dis]		precon: [UE aco-lex-dis]



\*head movement\*

\*name option\*

The menu is still displaying the menu choices, with the cursor moving from underneath one option to underneath another. The arm is at rest and is not moving. The user is looking at the screen and cursor in order to choose the option **Min**. There is no need for a precondition of looking at the display for the UE modality of naming the option because the user may have stopped looking at the display at that point.

---

39.

[SE vis-lex-cont]                      and    [SE vis-sym-dyn]                      and    [SE hap-con-cont]  
\*menu display\*                                              \*moving cursor\*                                              \*position of arm\*

precon:[UE hap-sym-dis]

\*head movement\*

or

[UE aco-lex-dis]

\*name option\*

[UR vis-lex-cont]                      and    [UR vis-sym-dyn]

\*menu display\*                                              \*moving cursor\*

precon: [SE vis-lex-cont]                      precon: [SE vis-sym-dyn]

\*menu display\*                                              \*moving cursor\*

precon: looking at display                      precon: looking at display

or

[UR vis-con-cont]

\*position of arm\*

precon: [SE hap-con-cont]

\*position of arm\*

precon: looking at arm

The system is displaying a new menu, underneath which is a flashing cursor moving from one option in turn to another. The arm not moving and is at rest. The user is either looking at the menu with the cursor, or at the arm.

There is a choice for the user between: **Go, Speed Level, End**

---

40.

[SE vis-lex-cont]                      and    [SE vis-sym-dyn]                      and    [SE hap-con-cont]  
\*menu display\*                                              \*moving cursor\*                                              \*position of arm\*

precon:[UE hap-sym-dis]

\*head movement\*

or

[UE aco-lex-dis]

\*name option\*

[UR vis-lex-cont]                      and    [UR vis-sym-dyn]  
\*menu display\*                                              \*moving cursor\*  
precon: [SE vis-lex-cont]                      precon: [SE vis-sym-dyn]  
\*menu display\*                                              \*moving cursor\*  
precon: looking at display                      precon: looking at display

[UE hap-sym-dis]                      or    [UE aco-lex-dis]  
\*head movement\*                                              \*name option\*  
precon:[UR vis-sym-dyn]                      precon: [UR vis-lex-cont]  
\*moving cursor\*                                              \*menu display\*  
precon: looking at display

[SR hap-sym-dis]                      or    [SR aco-lex-dis]





The arm is in motion. The arm will continue moving until **Stop** is selected. The cursor is now under the only option, so has changed from dynamic to continuous. The user is either looking at the movement of the arm, or at the display.

---

42.

[SE vis-lex-cont]      and      [SE vis-sym-cont]      and      [SE hap-con-cont]  
\*menu display\*                      \*stopped cursor\*                      \*arm moving\*

precon:[UE hap-sym-dis]

\*head movement\*

or

[UE aco-lex-dis]

\*name option\*

[UR vis-lex-cont]      and      [UR vis-sym-cont]

\*menu display\*                      \*stopped cursor\*

precon: [SE vis-lex-cont]      precon: [SE vis-sym-cont]

\*menu display\*                      \*stopped cursor\*

precon: looking at display      precon: looking at display

or

[UR vis-con-cont]

\*arm moving\*

precon: [SE hap-con-cont]

\*arm moving\*

precon: looking at arm

[UE hap-sym-dis]      or      [UE aco-lex-dis]

\*head movement\*                      \*name option\*

precon:              [SE vis-lex-cont]

\*menu display\*

[SR hap-sym-dis]                      or                      [SR aco-lex-dis]



*head movement*		*name option*
precon:[UE hap-sym-dis]		precon: [UE aco-lex-dis]
*head movement*		*name option*

or

[SE hap-con-cont]  
 \*arm stops\*  
 precon: [SE hap-con-cont]  
 \*arm moving\*

The arm is in motion until the user either choses **Stop** or the arm reaches the limit of movement. The UE modality of head movement has no precondition, because the user does not have to be looking at the menu display, since there is only option available at this point. The user can therefore be looking at the arm if needed. Also, since only one option available, the user may not have to read the menu interface to say “Stop”, since again there is only one option. Therefore the precondition for the UE modality of saying “Stop” is not definate. The system will either accept the UE modalities or the arm will reach the limit of motion.

---

43.

[SE vis-lex-cont]	and	[SE vis-sym-dyn]	and	[SE hap-con-cont]
*menu display*		*moving cursor*		*position of arm*
precon:[SR hap-sym-dis]		precon:[SR hap-sym-dis]		precon:[SR hap-sym-dis]
*head movement*		*head movement*		*head movement*
or		or		or
[SR aco-lex-dis]		[SR aco-lex-dis]		[SR aco-lex-dis]
*name option*		*name option*		*name option*
or		or		or
[SE hap-con-cont]		[SE hap-con-cont]		[SE hap-con-cont]
*arm stops *		*arm stops *		*arm stops *

[UR vis-lex-cont]      and      [UR vis-sym-dyn]  
 \*menu display\*                      \*moving cursor\*  
 precon: [SE vis-lex-cont]      precon: [SE vis-sym-dyn]  
 \*menu display\*                      \*moving cursor\*  
 precon: looking at display      precon: looking at display

or

[UR vis-con-cont]  
 \*position of arm\*  
 precon: [SE hap-con-cont]  
 \*position of arm\*  
 precon: looking at arm

The arm has stopped moving and the system is displaying a new menu, underneath which the flashing cursor is once more moving from one option in turn to another. The user is either looking at the menu with the cursor, or at the arm.

There is a choice for the user between **Continue, Speed Level, End**

---

44.

[SE vis-lex-cont]                      and      [SE vis-sym-dyn]                      and      [SE hap-con-cont]  
 \*menu display\*                                      \*moving cursor\*                                      \*position of arm\*  
 precon:[UE hap-sym-dis]  
 \*head movement\*

or

[UE aco-lex-dis]  
 \*name option\*

[UR vis-lex-cont]      and      [UR vis-sym-dyn]  
 \*menu display\*                      \*moving cursor\*



precon: [SE vis-lex-cont]	precon: [SE vis-sym-dyn]
*menu display*	*moving cursor*
precon: looking at display	precon: looking at display

[UE hap-sym-dis]	or	[UE aco-lex-dis]
*head movement*		*name option*
precon:[UR vis-sym-dyn]		precon: [UR vis-lex-cont]
*moving cursor*		*menu display*
precon: looking at display		

[SR hap-sym-dis]	or	[SR aco-lex-dis]
*head movement*		*name option*
precon: [UE hap-sym-dis]		precon: [UE aco-lex-dis]
*head movement*		*name option*

The menu is still displaying the menu choices, with the cursor moving from underneath one option to underneath another. The arm is at rest and is not moving. The user is looking at the screen and cursor in order to choose the option **End**. There is no need for a precondition of looking at the display for the UE modality of naming the option because the user may have stopped looking at the display at that point.

---

45.

[SE vis-lex-cont]	and	[SE vis-sym-dyn]	and	[SE hap-con-cont]
*menu display*		*moving cursor*		*position of arm*
precon:[UE hap-sym-dis]				
*head movement*				
	or			
[UE aco-lex-dis]				
*name option*				
[UR vis-lex-cont]	and	[UR vis-sym-dyn]		

*menu display*	*moving cursor*
precon: [SE vis-lex-cont]	precon: [SE vis-sym-dyn]
*menu display*	*moving cursor*
precon: looking at display	precon: looking at display
or	
[UR vis-con-cont]	
*position of arm*	
precon: [SE hap-con-cont]	
*position of arm*	
precon: looking at arm	

The system is displaying the original menu, underneath which is a flashing cursor moving from one option in turn to another. The arm not moving and is at rest. The user is either looking at the menu with the cursor, or at the arm.

There is a choice for the user between: **Move**, **MoveArm** (amongst others)

### **13.6 Stage 7. Assess the use of modalities**

There was one potential physical clash, which occurred many times during the analysis, resulting from the user having to potentially look in one direction to observe the arm, and in another to observe the menu options. However, this was not a problem for the most critical part of the interaction, that is, giving input for the arm to stop, since providing that the user knows that the command is called stop, they are able to say it without looking at the menu. If the command is by gesture, then stop is the only command available at that point, so the user does not have to check that the cursor is under it or not.

There are no temporal, lexical or clash unless expert clashes, with the exception of the clash unless expert clash of having to judge the distances the arm needs to move or extend and possibly becoming confused.



There may be one or more semantic clashes relating to the terms **move/movearm**, **arm** and **arm**. The user may not understand the difference between what move and what movearm does, since there is a joint called arm, in addition to the arm as a whole.

The interface makes no use of acoustic feedback at any point, although it provides visual feedback via both the arm and the display.

### **13.7 Stage 8. Final report**

By defining the task to be analysed the difference between what is needed to do, and how that is to be done, is made apparent. The task of switching on the light switch in this instance involves moving the arm as a whole as well as two of the arm joints. This is therefore not a quick and easy task, but one which requires planning and careful thought in order to establish how best to move the arm and joints. This points to a possible conceptual problem for the user, since their idea of the task, and the actual way in which the task may be accomplished, might be different.

In the Environment Profile, it was noticed that the task was to switch on a light switch. Therefore, the level of lighting already in existence had to be considered, as did what lighting might be available from the display.

The comparing of the profiles to the lists of modalities raised various issues. The comparison of user profile with system modalities suggested that it might be difficult for the user, due to disabilities of movement, to move their field of vision from display to arm and back again. The comparison of the system profile with the user modalities raised the issue that the movement of the user turning from looking at the arm to the menu display, or vice versa, might be mis-interpreted by the system as a gesture command. The comparison of the user and system modalities with the environment profile again highlighted the issue of potentially poor lighting, with the user maybe not having enough light to see the menus.

When writing out the interaction sequence, the long and repetitive nature of the listing suggested that the structure of the interaction had too many steps. The similar nature in terms of modalities of the individual stages of the interaction suggested that it was made up

of a few key sequences repeated many times. Writing out the modalities involved in the **stop** command showed how complicated that section of the interaction was.

There were various properties and clashes noted. There was a physical clash at certain stages of the interaction resulting from the user having to potentially look in one direction to observe the arm, and in another to observe the menu options. There was a clash unless expert type clash involving the judging of the distance that the arm needed to be extended, which might cause problems. There were potential semantic clashes relating to the terms **move/movearm**, **arm** and **arm**, since the user may not understand the difference between what move and what movearm does, due to there being a joint called arm, in addition to the arm as a whole.



# 14. APPENDIX D: THE HEATING CONTROL

## PANEL WEBSITE MACHINE TASK

The heating control panel website is shown below.

The screenshot shows a web interface for 'ALHAMBRA Central Heating Controls'. At the top, the title 'ALHAMBRA' is centered, followed by 'Central Heating Controls'. Below this is a control panel with two rows of input fields. The first row has two empty boxes. The second row has two boxes labeled 'ON:' and 'OFF:'. Below the input fields is a button panel with four buttons: 'PLUS', 'ADVANCE', 'MINUS', and 'COPY'. Underneath the buttons, it says 'Press ADVANCE to start'. Below that is a section titled 'Instructions' with the following text: 'These controls determine three periods of heating for each weekday. By clicking the ADVANCE button you can move from day to day and within each day you can move through the ON and OFF times for each period. The PLUS and MINUS buttons alter whichever panel is currently highlighted. Don't try typing in the boxes - just use the PLUS and MINUS buttons.' Below the instructions is another note: 'When the day of the week is highlighted, PLUS moves on to the next day, and ADVANCE brings up the settings for the day, while you can use the COPY button to copy yesterday's settings to today.' At the bottom, a 'NOTE' states: 'NOTE - The original device has locks to prevent an Off time being set to earlier than an On time in the same period, etc. I haven't bothered to program those.'



# **15. APPENDIX E: TASK STEP SUMMARY FOR**

## **TUBE TICKET MACHINE TASK**

Task: buy an adult return ticket to St Pancras at 10 am on a week-day from Bounds Green tube station, using a ten pound note. You may assume that there is sufficient change in the machine.

This task required the following steps:

Machine displays “select ticket type”.

User selects Adult Return Ticket.

Machine displays “select destination”.

User selects Kings Cross/St.Pancras.

Machine displays “Not available”.

At this point some reports finished, since the task could not be completed in the form asked for. To continue the task, the following steps were required:

User selects Adult 4 Zone travelcard.

Machine displays “insert money”

User inserts money.

Machine prints ticket and deposits change, and displays “collect ticket”

User collects ticket and change.



# **16. APPENDIX F: STUDENT PACK**

## **Student Pack for COM 3210**

### **Modality Information**

#### **Contents**

Introduction

Modality Definition

Taxonomy

Modality Examples

Modality Types and Properties

Modality Clashes

Modality Methodology

Coursework Details

Appendices

Cover Sheets

#### **Introduction**

This pack is designed to give you all the information you need about modalities. It starts off by talking about what a modality is and how it is defined. It then talks about how to identify them, the various types and properties of modalities, and the kinds of clashes that can occur. Then the methodology is explained in full, with examples. The coursework instructions are included, and an appendix gives the information in an easily accessible format.

This work is part of my research into modality, and by doing the coursework using this methodology you will be directly contributing to my studies.

It may be that you require clarification and amplification of the material contained in this pack over the course of the next few weeks. If so, please email me on [j.k.hyde@mdx.ac.uk](mailto:j.k.hyde@mdx.ac.uk), and I will reply as soon as possible.

### **Modality Definition**

The word modality means different things to different disciplines. In psychology it refers to the human sensory channels, mainly the auditory (hearing), visual (seeing) and haptic (touching and feeling) senses. In computer science it refers to the input and output devices of a computer, such as a keyboard, mouse or monitor. Other disciplines have other definitions. When these areas converge in an area like human-computer interaction, it is difficult to gain a clear understanding of what is being discussed, because the definitions cannot be reconciled. Only when modalities can be usefully defined can their usability properties can be established, and potential usability problems identified.

I have defined modality as a **temporally based instance of information perceived by a particular sensory channel.**

This definition has three parts to it: time, information form, and sensory channel.

The temporal nature of the information is important in terms of processing by either the user or the system. Different temporal forms will require different resources; for example a repetitive or continuous form of information is going to require different resources to a dynamic stream of information which is constantly changing. Time is also important in the ordering of modalities.

By stating the form of the information, we are beginning to develop an understanding of what form the information takes. A word is processed differently to a picture. In addition to this, we have to allow for information which may be expressed but in a real-world rather than symbolic manner, for instance a picture or the sound of voices.

Lastly, the use of particular sensory channels makes the definition user-dependent. This is because the form of the user is relatively static in comparison to the form of the system. System input and output possibilities are constantly changing and developing as new technologies are implemented, whereas the sensory channels have been standardised for the past few thousand years and are less likely to change suddenly.



## Taxonomy

By defining modality in this way, it is now possible to create a way of identifying modalities, by means of a taxonomy, or classification system. This taxonomy has three dimensions.

### **Sensory dimension**

The first dimension takes the three human senses, audio, visual and haptic, through which humans interact with computers. They are the three main channels through which information is perceived and communicated, and thus any taxonomy of modality should include them. The three categories in this dimension are **acoustic**, **visual** and **haptic**.

An **acoustic** modality is one that utilises the acoustic sensory channel. An example of an acoustic modality could be speech or music.

A **visual** modality is one that utilises the visual sensory channel. An example of a visual modality is a display or a picture.

A **haptic** modality is one that involves touching or movement. An example of a haptic modality could be sign language or pressing buttons.

### **Temporal dimension**

The second dimension refers to the temporal nature of the modality, and whether a modality is **discrete**, **continuous**, or **dynamic**.

A **discrete** occurrence of a modality is one that is unchanging within its occurrence, which is brief, and thus needs only preliminary processing. An example of a discrete occurrence could be the press of a button, the single ring of a doorbell, or the appearance and quick disappearance of a message on a screen.

A **continuous** occurrence of a modality is one that is repeated exactly the same one or more times. It is distinguished from the discrete occurrence, because it allows for some degree of inspection. An example of a continuous occurrence could be the continuous pressing of a button, the continuous ringing of a telephone, or the continuous display of a photo on a screen.

A **dynamic** modality is one that changes in content within its occurrence, which may last for some time, and thus requires continuous processing. An example of a dynamic modality could be a song being played, or a video.

### **Information Form**

The third dimension of the taxonomy relates to how the information is presented. This can be divided into three: **lexical**, **concrete** and **symbolic**.

**Lexical** is where the information is in the form of text. An example of a lexical modality could be sign language, speech, or text.

**Concrete** is where the information is in the form of reproduction of a real life object. An example of a concrete modality could be actual noise made by a dog, photos or videos, or physical vibrations as picked up by a lie-detector.

**Symbolic** is where the information is a representation of something rather than an actual reproduction of it. An example of a symbolic modality could be a fire alarm, gestures with meaning other than sign language, icons on a windows display, or the use of colour to denote danger.

### **Modality Examples**

Any modality can now be identified in terms of its sensory form, information form, and temporal form. Here I will use the example of a telephone in order to show how modalities are identified.

Interaction with a telephone makes use of various modalities:

The bell ringing

The user lifting the receiver

The caller speaking

The bell ringing uses the acoustic sensory channel, is symbolic in its information form, in that the bell sound means “telephone”, and is continuous, in that it will keep making the same noise until someone answers the call. The modality can therefore be written as acoustic-symbolic-continuous (written aco-sym-cont for short).



The user lifting the receiver uses the haptic sensory channel, is symbolic in its information form, in that the user is lifting something and thereby signalling acceptance of the call, and is discrete, in that it is one action that is not prolonged or repeated. The modality can therefore be written as haptic-concrete-discrete (written hap-con-dis for short).

The caller speaking uses the acoustic sensory channel, is lexical in its information form, because the caller is (hopefully!) using words, and is dynamic, in that the words change and are not repeated. The modality can therefore be written as acoustic-lexical-dynamic (written aco-lex-dyn for short).

However, here I have not distinguished between what the system is doing and what the user is doing. It is important to be able to show both the modality that is expressed by either the system or user, and the modality which is received by either the system or user. For example, using this idea, the modalities used in the interaction with the telephone increases:

The bell ringing (expressed by the system)

The bell ringing (received by the user)

The user lifting the receiver (expressed by the user)

The user lifting the receiver (received by the system)

The caller speaking (expressed by the system)

The caller speaking (received by the user)

There is not always a direct pairing between expressive and receptive modalities in this way, but most often there is. For example, I might shout at my computer, thus using the modality UE aco-lex-dyn, but the computer would remain blissfully unaware of this. Unfortunately!

In order to denote user, the letter U is used. In order to denote system, the letter S is used. In order to denote if a modality is expressive, the letter E is used. In order to denote if a modality is receptive, the letter R is used.

So the modalities listed above could be written as:

SE aco-sym-cont (the bell ringing, expressed by the system)

UR aco-sym-cont (the bell ringing, received by the user)

UE hap-sym-dis (the user lifting the receiver, expressed by the user)

SR hap-sym-dis (the user lifting the receiver, received by the system)

SE aco-lex-dyn (the caller speaking expressed, by the system)

UR aco-lex-dyn (the caller speaking received, by the user)

### **Modality Types and Properties**

Having identified the modalities in an interaction, it is now possible to talk about the types and properties of those modalities.

Modalities are not always at the same level of **granularity**, or abstraction. For example, imagine a telephone conversation, where a person is talking. That modality would be written as **UE aco-lex-dyn**.

However, when a person is talking they can often add more information to what they are saying by varying their tone, the volume, and even the accent in which they are speaking. These are all extra modalities, because they are different layers of information on the original modality of speech. By taking it down to a lower level of granularity, we can add in further modalities. Obviously this depends on what level of analysis is appropriate: normally, when analysing a telephone conversation, the UR aco-lex-dyn modality would be sufficient.

Imagine a Scotsman talking on the telephone. He is irritated, so he has raised his voice. He is in a hurry, so he is talking fast. The modalities would be as follows:



UE aco-lex-dyn: the user talking. This would be the overall, **composite** modality. It is made up of other **atomic** modalities:

UE aco-lex-dyn: the words the Scotsman is using

UE aco-sym-dyn: the volume of his words

UE aco-sym-dyn: the speed of his words

UE aco-con-dyn: his accent

The volume and speed are symbolic, because they represent something in the interaction: they show that the user is irritated and in a hurry. The accent is not important in this case, it just is, therefore it is concrete.

The **atomic** modalities are those which make up a composite modality. They are at the lowest level of analysis at that stage. An atomic modality may in turn be made up of other modalities if the analysis was taken down to a lower level of analysis. This atomic modality would therefore become a composite modality, and those modalities which it was comprised of would then be known as atomic modalities.

The volume, speed and accent modalities are known as **dependent** modalities, because they are dependent on the aco-lex-dyn modality, or the actual words of the conversation. They are properties that can only exist if there is a modality originally.

Suppose it is you that the Scotsman is talking to. You would be receiving the modalities:

UR aco-lex-dyn: the words the Scotsman is using

UR aco-sym-dyn: the volume of his words

UR aco-sym-dyn: the speed of his words

UR aco-con-dyn: his accent

However, imagine that you don't realise that he is speaking fast because he is in a hurry. You might think that that is how he normally talks. To you, then, that modality would be UR aco-con-dyn, with the speed represented as concrete, because it does not represent anything of importance to you.

This is known as a **mis-match**, where information expressed in a modality is not correctly received. It most often occurs because the recipient is not aware that the modality exists in this context (if you knew that the Scotsman usually spoke slowly, you would probably realise that he was speaking faster than usual because he was in a rush).

Another type of mis-match is when information is received that was not intended. For example, imagine that the Scotsman usually spoke quickly, but that you were not aware of this. Hearing him speak, you might think that he was in a rush, when actually it was his normal speed. He would be giving the modality SE aco-con-dyn, and you would be receiving UR aco-sym-dyn.

Imagine that both volume and speed increased if you were speaking in a hurry. You would be expressing the following modalities: UE aco-sym-dyn and UE aco-sym-dyn. However, the system would only need to identify one of those modalities in order to realise that you were in a hurry. It would not matter if it was volume or speed, either would do. In this case we have a **redundant** modality. This is where exactly the same information is being given in an extra modality. In some cases this is important in order to emphasise a point. In other cases it can show that there is overload, and that the modality is not needed.

### **Modality Clashes**

Modalities can often interfere with each other, and cause processing problems both for humans and systems. Various types of clashes can be identified, and some examples are given below. It should be borne in mind that there can be more than one clash occurring at any given time. The clashes, with the exception of the first type, are user-dependent in the examples given; however, the same principles can be applied to a system once the system configuration and its limitations are known.

#### **Physical clashes**

This is where the modalities clash in terms of the physical characteristics of the user or system. For example, a computer may not be able to process two streams of voice input simultaneously. Similarly, a user cannot say two different things at the same time. It is not possible to detail exactly when system clashes will occur, because this will depend on the



configuration of each particular system. It is possible to identify human physical clashes, although users with special needs will have different problems.

One very important point is that a user cannot look in two different directions at the same time. Only if things are within a field of vision can they be visually processed. This is obviously important if a system is expressing two modalities, in two different spatial locations, where the user would not be able to see them both at once.

### **Lexical clashes**

There are various restrictions on what a user can and cannot process lexically. A user cannot process two different lexical inputs or outputs at the same time; for example, reading a book and listening to a sports commentary, writing an address and reciting the alphabet, or writing a letter and listening to the news on the radio. Only if the lexical information is the same would a user be able to do this.

### **Temporal Clashes**

The temporal nature of the modalities can have an effect on how well they are processed. It is difficult to combine two dynamic modalities unless the modalities relate closely to each other, and can be thought of as a composite modality. An example of this is reading someone's lips and listening to what they say. If the two modalities (lip-movement and speech) are co-ordinated, with no time delay or synchronisation problems, then these two modalities are combined and do not interfere with each other. However, if there is a time delay between the two, it becomes extremely difficult to process the modalities, and problems can occur.

The same problem can occur with combinations of dynamic and discrete modalities, and discrete and discrete modalities. This is due to the user focusing on one of the modalities, and thus losing the information given in the other modality. For example, imagine that I am listening to a radio commentary, and the doorbell rings. For a moment, my attention switches from the commentary to the doorbell, and I do not process what the commentary says for that brief instant.

There is no problem when combining continuous modalities with other continuous modalities, or with discrete or dynamic modalities. This is because continuous modalities allow a certain degree of freedom of inspection.

The following table clarifies this:

	Dynamic	Continuous	Discrete
Dynamic	possible interference	fine	possible interference
Continuous	fine	fine	fine
Discrete	possible interference	fine	possible interference

### **Semantic clashes**

These occur when information is provided in two modalities, which do not in themselves clash, but where the content of the modality causes problems. For example, having the word butterfly under the picture of a dog, or someone whispering conspiratorially about the weather.

### **Clash unless expert**

This is when the combination of modalities is such that only an experienced person would be able to either receive or express them properly. This type of clash is one that cannot be explicitly defined, and must be identified by the analyst using good judgement as to what the user may or may not be capable of. An example of such a clash is trying to change gears in a car whilst going round a bend and talking to a passenger. An expert driver would have no problem with this, whilst a learner driver might find that combination impossible.

### **Modality Methodology**

The methodology is quite straightforward and consists of several stages. Each stage will be examined in turn, using an example interaction in order to clarify the procedure. It is expected that the analyst will repeat some of the stages as their knowledge of the task and



interaction develops. It is not expected that the analysis will follow this methodology in a completely linear fashion. This methodology aims rather to support the analyst by making apparent the various stages necessary for a complete analysis of the modalities used in an interaction and the potential usability consequences of such modalities.

The example analysis will be of the modalities used in the following task:

**to answer a telephone and have a conversation with the person on the other end.**

The telephone in this case will have a flashing light as well as a ringing noise in order to attract the user's attention to the incoming call.

### **Stage 1. Define the task that is to be analysed**

Before the task can be analysed it has to be defined. In this case it is defined as:

To answer a telephone and have a conversation with the person on the other end.

The selection of this particular task as opposed to any other task now has to be justified and explicitly stated. For example, is this a standard task for this device? Is the task complicated enough to show a wide selection of potential problems?

Since the telephone is primarily a communication device, being able to receive incoming calls successfully is a standard task. Problems with this task would show that the device interface has been badly developed.

### **Stage 2. Modality lists**

At this stage all the modalities used by the user and the system should be listed, both descriptively, i.e. "phone ringing", and notationally, i.e. "SE aco-sym-cont". This list will probably grow as the analysis is iterated. The level of granularity should be discussed at this point, in terms of the level of detail appropriate to the analysis. Again, this may change as the analysis is iterated and further considerations become apparent.

User modalities:

(both expressive and receptive modalities should be listed here)

Hear phone ringing: U R aco-sym-cont

See phone ringing: U R vis-sym-cont

Pick up receiver: U E hap-sym-dis

Speak to voice on end of phone: U E aco-lex-dyn

Listen to voice on end of phone: U R aco-lex-dyn

System modalities:

(both expressive and receptive modalities should be listed here)

Phone ringing: S E aco-sym-cont

Phone ringing: S E vis-sym-cont

Voice on end of phone: S E aco-lex-dyn

Note that the buttons on the telephone have been ignored for the purposes of this analysis since they are not necessary for the completion of this task.

Granularity:

The analysis was done at a reasonably high level of granularity. For example, it was not considered necessary to break down the speech modalities (UE aco-lex-dyn, SE aco-lex-dyn) into further atomic modalities, since this would not have any bearing on the usability of the device.

**Stage 3. Define the user, system and environment profiles**



At this point, the analyst should explicitly state the characteristics of the user, system and environment that can reasonably be expected to have a bearing on the usability of the interaction.

#### User Profile:

The user is able-bodied adult human, who speaks the same language as the caller. There are no physical or cognitive impairments.

Various assumptions have been made about what the user knows:

The user has had experience of using a telephone before

The user knows the sound of a telephone

The user knows what the flashing light means

The user knows to pick up the receiver to answer it

The user knows to say something to alert the person on the other end that the phone has been answered

#### System Profile:

The telephone is a simple handset, which needs to be picked up to be answered. A call coming in is indicated by a continuous ringing noise and by a continuous flashing light. The phone will ring at a certain volume in a certain manner, the light will flash at a certain rate, the handset has to be held at a certain distance from the mouth of the user for effective communication, and the handset has to be held at a certain distance from the ear of the user for effective communication.

#### Environment Profile:

The environment is an office setting, with subdued noise levels, and good lighting. There is only one system in the immediate vicinity of the user.

#### Stage 4. Profiles compared to modality listings

The profiles of the user, system and environment are compared against the lists of modalities in order to check for obvious discrepancies.

User profile compared with System modalities: there are no obvious discrepancies.

System profile compared with User modalities: there are no obvious discrepancies.

Environment profile compared with User and System modalities: there are no obvious discrepancies.

#### Stage 5. Interaction modality listing

At this point the interaction should be written out in sequence and in terms of the modalities used. Check that for every expressive modality, there is a corresponding receptive modality. There will not always be a pairing, but in most cases there will be.

It would probably be easiest for the analyst to start by writing out the interaction sequence in terms of the actions, then breaking it down into a description of the modalities, and finally in terms of the notation. For this example, the writing has gone directly to the notational stage.

---

1. [SE aco-sym-cont] and [SE vis-sym-cont]

\*phone ringing\*

\*phone flashing\*

[UR aco-sym-cont]      or

[UR vis-sym-cont]

\*user hears phone ringing\*

\*user sees phone flashing\*

precon: [SE aco-sym-cont]

precon: [SE vis-sym-cont]

\*phone ringing\*

\*phone flashing\*

precon: user looking at phone



The interaction starts with the phone ringing and flashing, and the user becoming aware of this.

Each part of the interaction sequence is numbered to allow for easy discussion of the sequence.

The modalities are written both notationally, with square brackets round them, in order that their constituent parts can be examined, and descriptively, within the star symbols (\*) in order that the analyst can keep track of which modalities are being discussed.

Both of the system modalities occur at the same time, so they are written as joined with a logical **and**.

Only one of the user modalities needs to occur at this stage, so they are written as joined with a logical **or**.

There are ordering considerations at this point to take into account. The telephone has to ring or flash before the user can become aware of the ringing or flashing. Therefore the user modalities have the system modalities as **pre-conditions** (shortened to "precon" in the example given).

There is another pre-condition that the user should be looking at the phone in order to see if it is flashing. All visual modalities have this pre-condition.

---

2. [SE aco-sym-cont] **and** [SE vis-sym-cont]

\*phone ringing\*

\*phone flashing\*

[UR aco-sym-cont]            **or**

[UR vis-sym-cont]

\*user hears phone ringing\*

\*user sees phone flashing\*

precon: [SE aco-sym-cont]

precon: [SE vis-sym-cont]

\*phone ringing\*

\*phone flashing\*

precon: user sees phone flashing

[UE hap-con-dis]

\*user picks up phone\*

precon: [UR aco-sym-cont] or [UR vis-sym-cont]

\*user hears phone\* \*user sees phone flashing\*

The phone is ringing and flashing, the user is aware of this, and has picked up the receiver. The pre-condition to the user picking up the receiver is that the user is aware of the phone ringing and flashing.

-----

3. [UE hap-con-cont]

\*user holds phone\*

precon: [UE hap-con-dis]

\*user picks up phone\*

[SE aco-sym-cont]

\*line quiet\*

precon: [UE hap-con-dis]

\*user picks up phone\*

[UR aco-sym-cont]

\*user hears line quiet\*

precon: [UE hap-con-dis]

\*user picks up phone\*

The user is now holding the phone, which is different to picking up the phone. The line is giving no tone, indicating there is probably someone on the other end, and the user is aware of this. The line being quiet is dependent upon the user picking up the phone.



---

4. [UE aco-lex-dyn]

\*user talks\*

precon: [UE hap-con-cont]

\*user holds phone\*

[SR aco-lex-dyn]

\*user talks\*

precon: [UE aco-lex-dyn]

\*user talks\*

[UE hap-con-cont]

\*user holds phone\*

precon: [UE hap-con-dis]

\*user picks up phone\*

The user is still holding the phone, and is now talking to the person on the other end. In order for the user to start talking, they must be holding the phone. There is an ordering condition in that the person on the other end of the phone cannot hear the user until the user has started talking.

---

5. [SE aco-lex-dyn]

\*caller talks\*

precon: [UE aco-lex-dyn]

\*user talks\*

[UR aco-lex-dyn]

\*caller talks\*

precon: [SE aco-lex-dyn]

\*caller talks\*

precon: [UE hap-con-cont]

\*user holds phone\*

[UE hap-con-cont]

\*user holds phone\*

precon: [UE hap-con-dis]

\*user picks up phone\*

The user is now listening to the caller speaking. This is conditional upon the caller talking and the user holding the phone.

---

#### Stage 6. Add in clashes, etc.

Now that the modalities have been established, and the sequence and pre-conditions explicitly stated, it is possible to look for the properties of the modalities and the potential clashes.

1. [SE aco-sym-cons] and [SE vis-sym-cons]  
\*phone ringing\* \*phone flashing\*

[UR aco-sym-cons] or [UR vis-sym-cons]  
\*user hears phone ringing\* \*user sees phone flashing\*  
precon: [SE aco-sym-cons] precon: [SE vis-sym-cons]  
\*phone ringing\* \*phone flashing



precon: user looking at phone

At this point it can be discerned that the two system modalities are redundant, that is, exactly the same information (incoming call) is being given in two different ways. There are no clashes, and the only other property to be aware of is that the user will have to be looking at the telephone in order to see it flashing.

Because this was such a simple example, there were no further properties or clashes found in the rest of the sequence.

### **Stage 7. Assess the use of modalities**

The analyst should assess the modalities used at this point in order to see if any are under-utilised, and if any of the properties flag issues that should be addressed.

This interface makes use of all three senses. However, the use of a flashing light is redundant, since the user would hear the phone ringing (the environment is quiet and the user has no special physical constraints). Also, the user has to be looking at the phone in order to see the flashing, which makes it less effective than using sound.

### **Stage 8. Final report**

Finally, write up into a report, assessing the overall usability of the system as a result of analysing the interaction in this way. Include conclusions and recommendations.

The system seems very usable. No problems on a usability front were found, which is to be expected from such a simple system. However, the use of a flashing light seems unnecessary given the user and environment profiles.

### **Further Examples**

The previous example was simple and straightforward, in order to show how the methodology works. However, interactions can often be more complex, and involve many modalities. Here is an example of the methodology for a task using the wordprocessing

package Word 6 for a Mac. Only the most relevant sections of the methodology have been included.

### **Stage 1. Define the task that is to be analysed**

To change a paragraph of an open document from Times font to Courier font.

### **Stage 5. Interaction modality listing**

The Word 6 wordprocessing package for the Mac has a lot of information on the screen at a time. There is the text space, where text is entered. There are the scroll bars which border this space. There is the menu at the top of the window giving options such as File, Edit, View; and there is a row of icons and text information at both the top and bottom of the text space. There are thus many visual modalities being expressed by the computer at once. This can make it quite complicated to write out the interaction modality listing.

Effectively what will happen is that user will only be looking at certain of the modalities displayed, but the system will be expressing all of them. Therefore, the ones that user is not looking at can be expressed at a very high level of abstraction, in order to show that they exist, but without unnecessary detail.

A portion of the interaction sequence below explains this:

-----

1.

[SE vis-sym-cont]

**\*scroll bars\***

**and**

[SE vis-sym-cont]

**\*menu options\***

**and**

[SE vis-sym-cont]

**\*icons at top of text space\***



**and**

[SE vis-sym-cont]

\*information strip at bottom\*

**and**

[SE vis-sym-dyn]

\*mouse cursor moving\*

precon: [UE hap-con-dyn]

\*user moving mouse\*

**and**

[SE vis-lex-cont]

\*text\*

[UE hap-con-dyn]

\*moves mouse\*

**and**

[UR vis-lex-cont]

\*scanning text\*

**and**

[UR vis-sym-dyn]

\*watching mouse cursor\*

precon: [UE hap-con-dyn]

\*user moving mouse\*

The system is displaying information on the screen. This information is represented at a very high level of abstraction. The user is only looking at two of these modalities, and is moving the mouse.

The list of modalities used here might seem lengthy, but imagine if all the composite modalities were written in terms of their atomic modalities. The list would be incredibly long, and it would be almost impossible to effectively analyse.

As the user switches attention, the high-level modalities can be described at a lower level of detail. For example, the section of the screen giving the menu options [SE vis-sym-cont] would then be written as follows:

[SE vis-sym-cont] \*area containing menu options\*

[SE vis-lex-cont] \*File\*

[SE vis-lex-cont] \*Edit\*

[SE vis-lex-cont] \*View\*

[SE vis-lex-cont] \*Insert\*

[SE vis-lex-cont] \*Format\*

[SE vis-lex-cont] \*Tools\*

[SE vis-lex-cont] \*Table\*

[SE vis-lex-cont] \*Window\*

[SE vis-lex-cont] \*time display\*

[SE vis-sym-cont] \*system option\*

[SE vis-sym-cont] \*help option\*

[SE vis-sym-cont] \*current packages option\*

and the rest of the screen could be described as:

[SE vis-sym-cont] \*scroll bars\*

[SE vis-sym-cont] \*icons at top of text space\*

[SE vis-sym-cont] \*information strip at bottom\*



[SE vis-sym-cont] \*mouse cursor\*

[SE vis-lex-cont] \*text\*

### Coursework Instructions

This coursework is worth 40% of your final coursework mark (i.e. 20% of your overall mark for this module).

The coursework is to be handed in on Thursday 12th November before close of Student Office. Marks will be deducted for late submission. Please use the cover sheet supplied at the back of this pack.

The report is to take the form of an analysis of two interactive devices, as detailed below. You are to work in pairs on the analysis of the devices, but the final written report is to be completed individually.

You are also required to hand in any notes made whilst working on the coursework at the same time as you hand in the report. These notes will **not** be assessed, and will be used only as information towards my research. Please use the cover sheet supplied at the back of this pack.

The report is to consist of:

An analysis of the modalities used in the following task:

**to buy an adult return ticket to St Pancras at 10am on a week-day from Bounds Green tube station, using a ten pound note. You may assume that there is sufficient change in the machine.**

An analysis of the modalities used in the following task:

**using the Alhambra heating control panel, located at <http://www.ndirect.co.uk/~thomas.green/devices/>, set the heating controls for the whole week (Monday through to Sunday), with the heat coming on at 7am each day, and going off at 10pm each day.**

The report should be clearly written, with each stage of the methodology detailed, and final usability considerations discussed, for each analysis.

## **Appendices**

A short summary of the information detailed in the pack is given here.

### **Taxonomy**

Modalities are identified according to three dimensions:

Three dimensions: sense, information form, time.

Sense: acoustic (aco), visual (vis) and haptic (hap).

Information form: lexical (lex), concrete (con), symbolic (sym).

Time: discrete (dis), continuous (cont), dynamic (dyn).

They are further distinguished as being user (U) or system (S), and expressive (E) or receptive (R).

### **Modality Types and Properties**

Modalities can be composite, atomic, dependent, mis-matched and redundant. Analyses can be done at differing levels of granularity.

### **Modality Clashes**

There are several types of clashes: physical, lexical, temporal, semantic, and clash unless expert.

### **Methodology**

The methodology is quite straightforward and consists of several stages. It is expected that the analyst will repeat some of the stages as their knowledge of the task and interaction develops. It is not expected that the analysis will follow this methodology in a completely linear fashion. This methodology aims rather to support the analyst by making apparent the various stages necessary for a complete analysis of the modalities used in an interaction and the potential usability consequences of such modalities.

#### **Stage 1. Define the task that is to be analysed**

Before the task can be analysed it has to be defined.



The selection of this particular task as opposed to any other task has to be justified and explicitly stated. For example, is this a standard task for this device? Is the task complicated enough to show a wide selection of potential problems?

## **Stage 2. Modality lists**

At this stage all the modalities used by the user and the system should be listed, both descriptively, ie “phone ringing”, and notationally, ie “SE aco-sym-cont”. This list will probably grow as the analysis is iterated. The level of granularity should be discussed at this point, in terms of the level of detail appropriate to the analysis. Again, this may change as the analysis is iterated and further considerations become apparent.

User modalities:

(both expressive and receptive modalities should be listed here)

System modalities:

(both expressive and receptive modalities should be listed here)

Granularity:

## **Stage 3. Define the user, system and environment profiles**

At this point, the analyst should explicitly state the characteristics of the user, system and environment that can reasonably be expected to have a bearing on the usability of the interaction.

User Profile:

System Profile:

Environment Profile:

## **Stage 4. Profiles compared to modality listings**

The profiles of the user, system and environment are compared against the lists of modalities in order to check for obvious discrepancies.

User profile compared with System modalities:

System profile compared with User modalities:

Environment profile compared with User and System modalities:

### **Stage 5. Interaction modality listing**

At this point the interaction should be written out in sequence and in terms of the modalities used. Check that for every expressive modality, there is a corresponding receptive modality. There will not always be a pairing, but in most cases there will be.

It would probably be easiest for the analyst to start by writing out the interaction sequence in terms of the actions, then breaking it down into a description of the modalities, and finally in terms of the notation.

Number each part of the interaction sequence to allow for easy discussion of the sequence.

Write the modalities both notationally, with square brackets round them, in order that their constituent parts can be examined, and descriptively, within the star symbols (\*) in order that the analyst can keep track of which modalities are being discussed.

If modalities happen at the same time they should be connected with the logical **and**.

If there is a choice between modalities, they should be connected with the logical **or**.

Take into account any pre-conditions relating to ordering or the use of the visual sense.

### **Stage 6. Add in clashes, etc.**

Now that the modalities have been established, and the sequence and pre-conditions explicitly stated, it is possible to look for the properties of the modalities and the potential clashes.

### **Stage 7. Assess the use of modalities**

The analyst should assess the modalities used at this point in order to see if any are under-utilised, and if any of the properties flag issues that should be addressed.

### **Stage 8. Final report**

Finally, write up into a report, assessing the overall usability of the system as a result of analysing the interaction in this way. Include conclusions and recommendations.



# **COM 3210 Coursework**

## **Modality analyses of two interactive devices**

**Module Leader: Ann Blandford**

**Name:**

**Student Number:**

# **COM 3210 Coursework Notes**

**Notes made whilst doing modality analyses of two interactive devices**

**Module Leader: Ann Blandford**

**Name:**

**Student Number:**



# **17. APPENDIX G: QUESTIONNAIRES**

## **17.1 First Questionnaire Questions**

Student Number:

Age:

Gender (male/female):

1. How clear did you find the lecture? not clear at all /very clear

2. How confident are you that you would be able to identify the modalities used in an interaction? no confidence/complete confidence

3. How confident are you that you can identify modality properties?

no confidence/complete confidence

4. How confident are you that you can identify modality types?

no confidence/complete confidence

5. How confident are you that you can identify modality and clashes?

no confidence/complete confidence

6. How confident are you that you can use the methodology as outlined in the lecture?

no confidence/complete confidence

7. Which previous HCI modules have you done? (please list here)

8. How experienced do you consider yourself to be in HCI?

no confidence/complete confidence

9. What previous experience have you had of structured HCI reports? (please list here)

10. How experienced do you consider yourself to be in structured HCI reports?

no experience/very experienced

11. What previous experience have you had of HCI analysis?

no experience/very experienced

12. What previous experience have you had of applying HCI and software engineering methodologies? (please list here)

13. How experienced do you consider yourself to be in applying HCI and software engineering methodologies? no experience/very experienced

14. How much experience have you had of using the world wide web?

no experience/very experienced

15. How much experience have you had of using London Underground tube ticket machines? no experience/very experienced

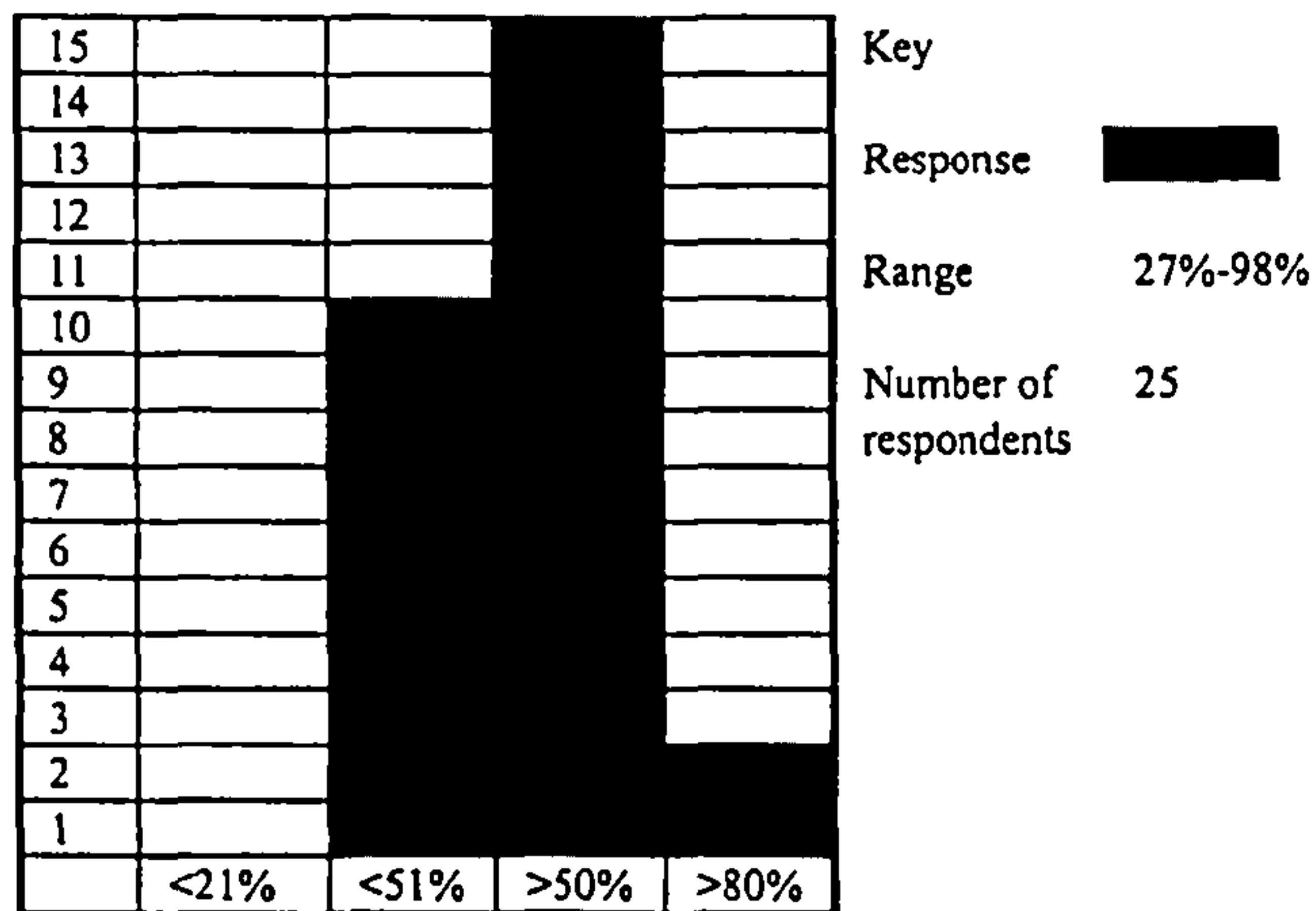
16. Is there any other information that you consider relevant? (please list here)

## **17.2 First Questionnaire Results Tables**

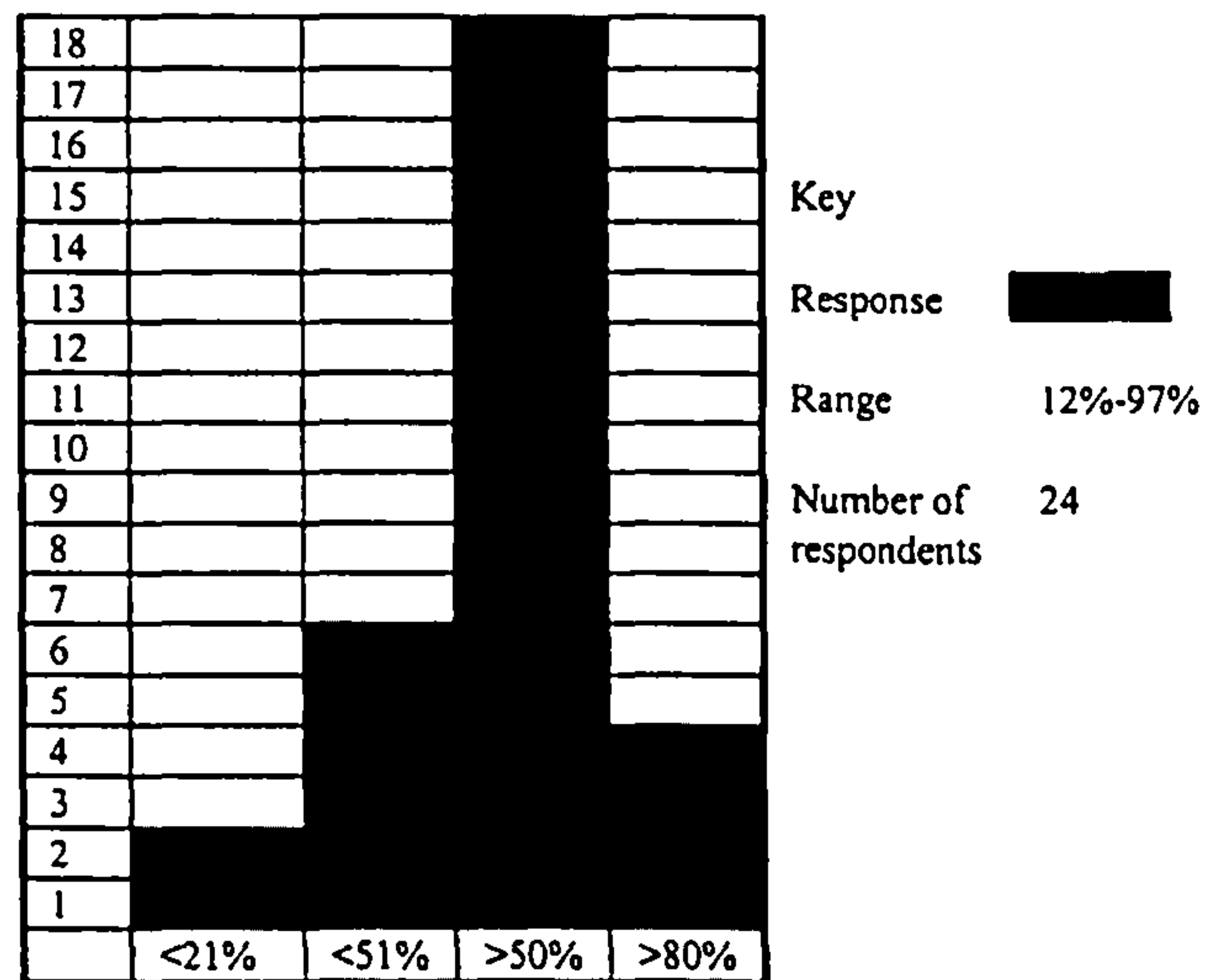
The results are presented in tables which show the number of responses above and below 50%, with additional columns giving responses also 20% or below, and above 80%, in order to provide a further breakdown of the results. Thus a response of 87% would be given in two columns; the above 50% column and the above 80% column.

The responses presented here relate to confidence in applying the methodology and identifying clashes and properties, and previous experience in HCI and analysis. The seven questions relate to: confidence in using the methodology; confidence in identifying modality properties; confidence in identifying modality clashes; experience in HCI; experience in structured HCI reports; previous experience of HCI analysis; and experience in applying HCI and software engineering methodologies.

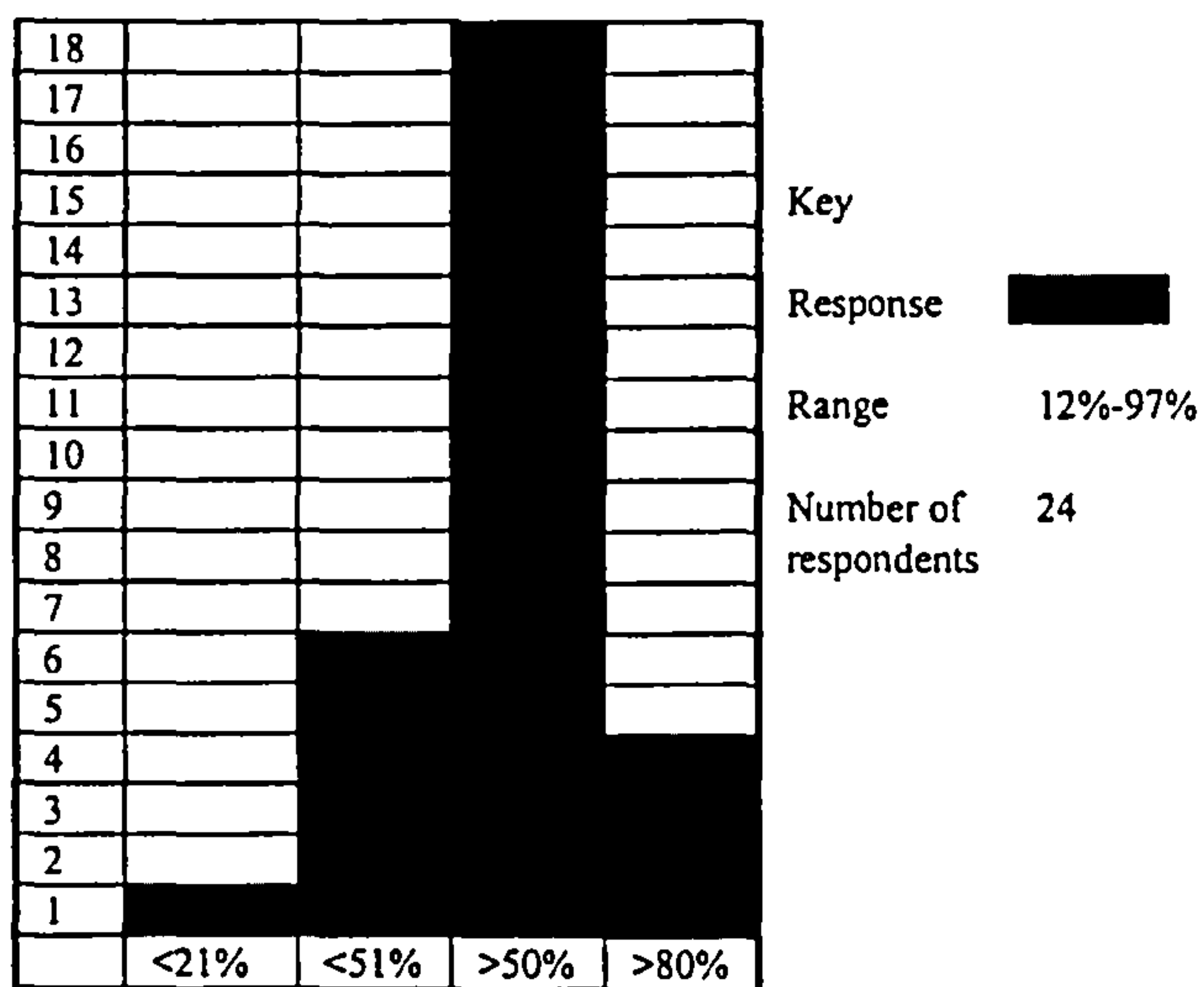




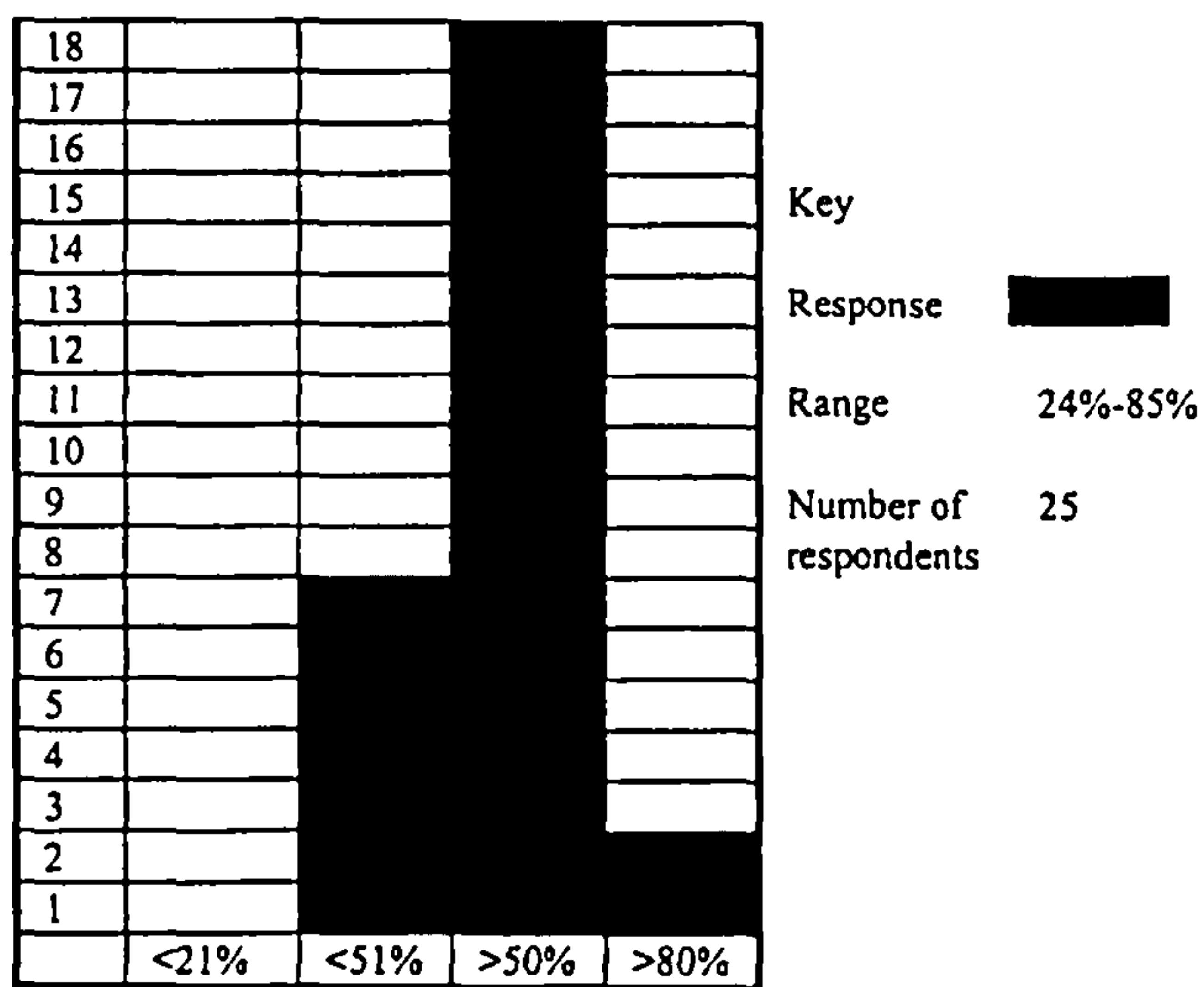
**Table 17-1: Responses to Question 6: How confident are you that you can use the methodology as outlined in the lecture?**



**Table 17-2: Responses to Question 3: How confident are you that you can identify modality properties?**

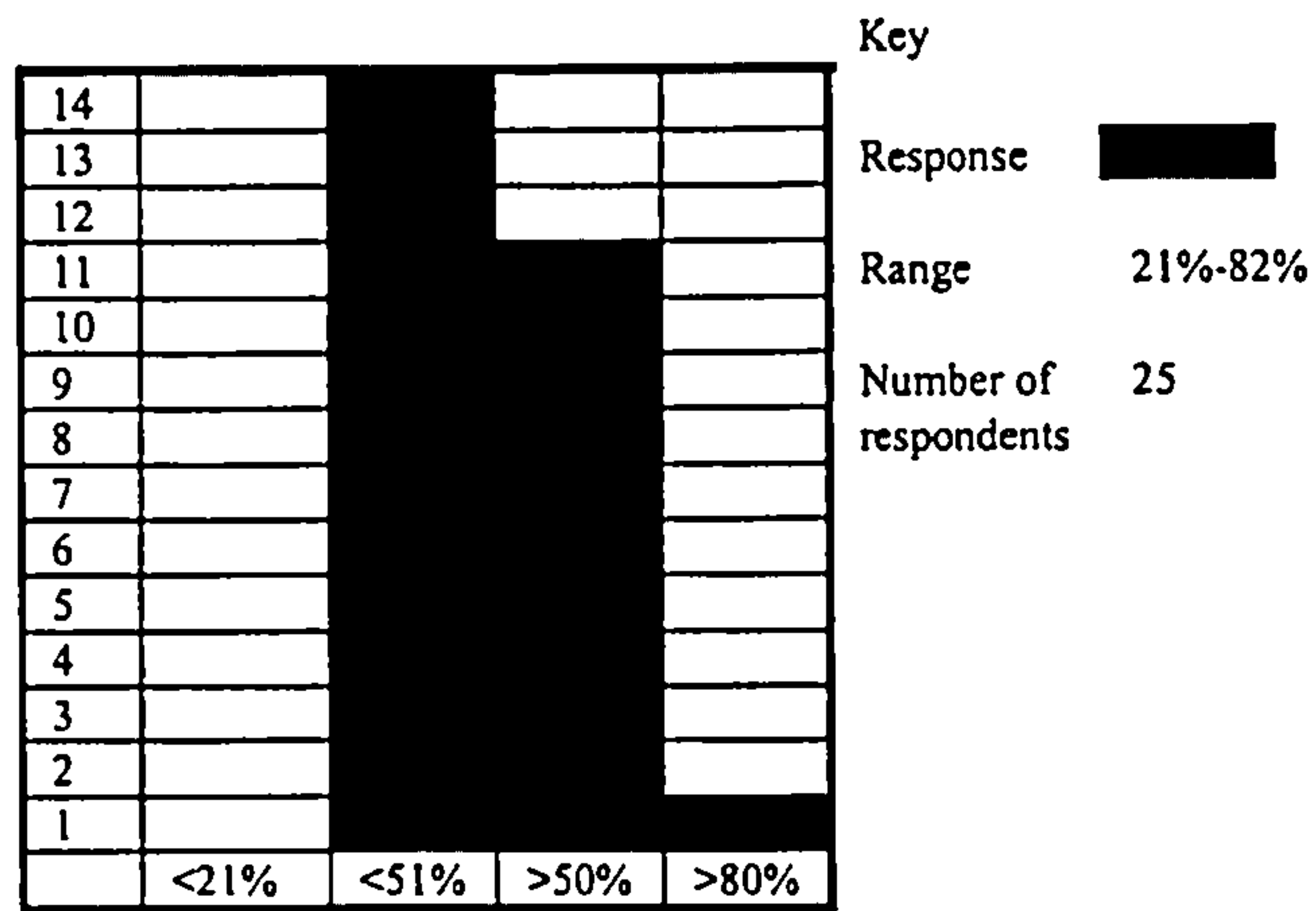


**Table 17-3: Responses to Question 5: How confident are you that you can identify modality clashes?**

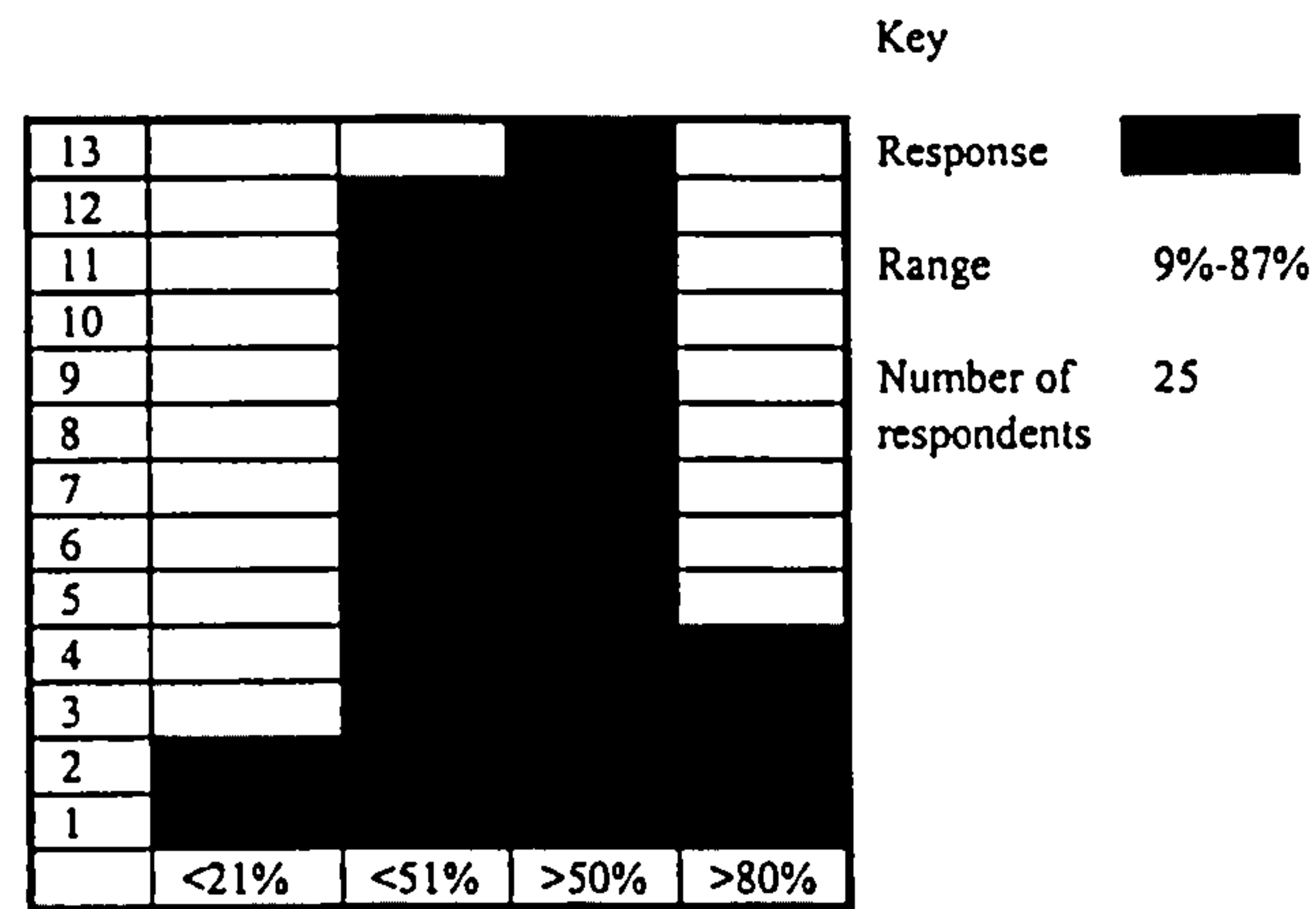


**Table 17-4: Responses to Question 8: How experienced do you consider yourself to be in HCI?**

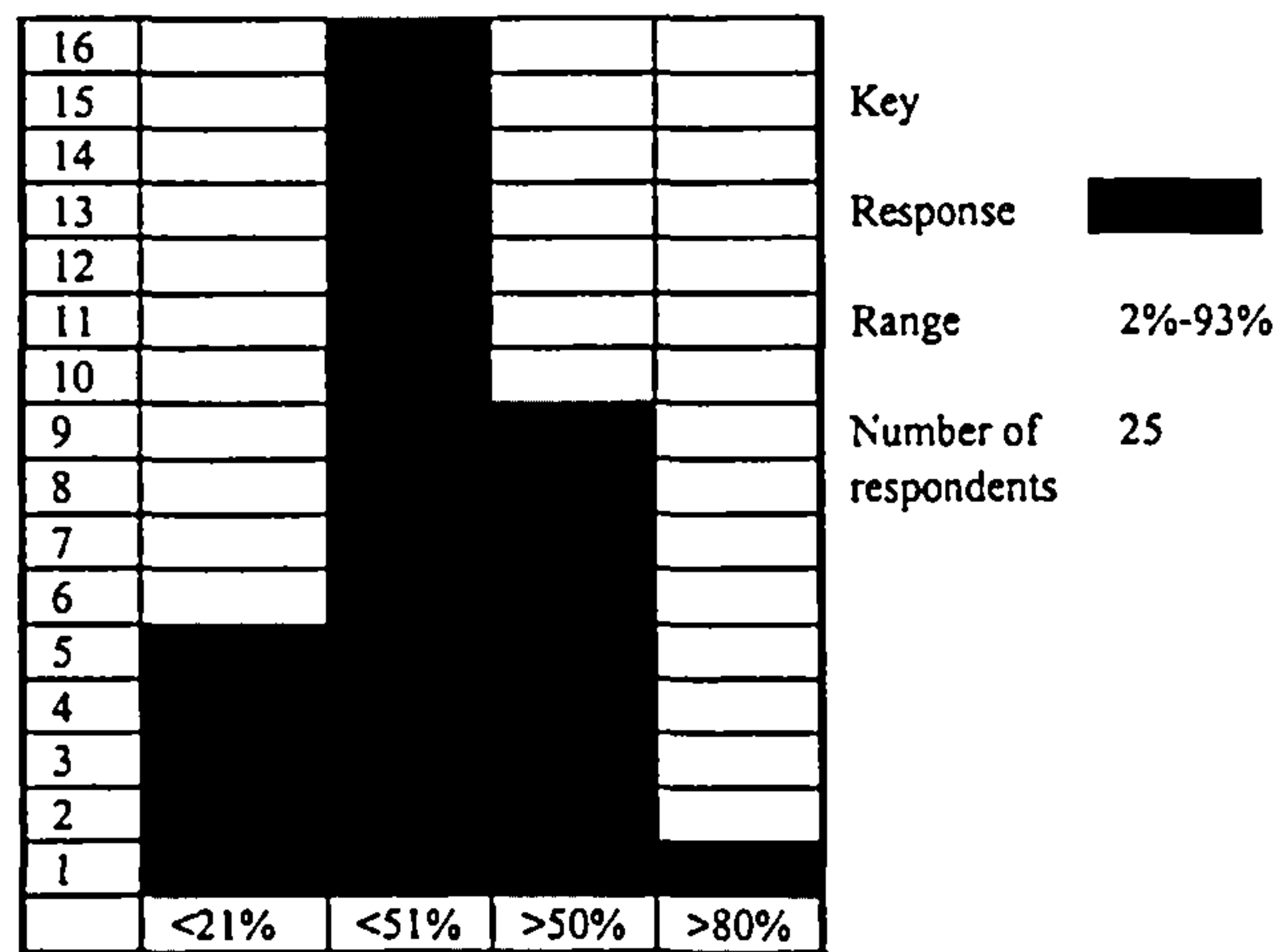




**Table 17-5: Responses to Question 10: How experienced do you consider yourself to be in structured HCI reports?**



**Table 17-6: Responses to Question 11: What previous experience have you had of HCI analysis?**



**Table 17-7: Responses to Question 13: How experienced do you consider yourself to be in applying HCI and software engineering methodologies?**

### **17.3 Second Questionnaire Questions**

Student Number:

Age:

Gender (male/female):

1. How straightforward did you find the methodology to apply?

not straightforward/very straightforward

2. How confident are you that you correctly identified the **modalities** used in the two interactions? no confidence/complete confidence

3. How confident are you that you correctly identified the modality **properties**?

no confidence/complete confidence

4. How confident are you that you correctly identified the modality **types**?

no confidence/complete confidence

5. How confident are you that you correctly identified modality **clashes**?

no confidence/complete confidence

6. How confident are you that you would be able to use this methodology again?



no confidence/complete confidence

7. Did you find that you were given enough information in the pack provided to complete the two reports successfully? not enough/enough information

8. How easy did you find the methodology as outlined in the pack to apply?

difficult/easy

9. How well did using the methodology help you to identify general usability issues?

not well/very well

10. For each of the methodology stages please indicate how easy they were to apply:

**Stage One:** Define the task that is to be analysed and justify its selection:

difficult/easy

**Stage Two:** List all modalities used by system and user, and talk about the granularity of the analysis: difficult/easy

**Stage Three:** Define the user, system and environmental profiles: difficult/easy

**Stage Four:** Compare profiles to modality listings: difficult/easy

**Stage Five:** Interaction listing of all modalities: difficult/easy

**Stage Six:** Identify clashes and modality properties: difficult/easy

**Stage Seven:** Assess the use of modalities: difficult/easy

**Stage Eight:** Assess overall usability of system and write up into a report:

difficult/easy

11. For each of the methodology stages please indicate how useful it was in uncovering relevant usability issues:

**Stage One:** define task that it to be analysed and justify its selection: not useful/useful

**Stage Two:** List all modalities used by system and user, and talk about the granularity of the analysis: not useful/useful

**Stage Three:** Define the user, system and environmental profiles: not useful/useful

**Stage Four:** Compare profiles to modality listings: not useful/useful

**Stage Five:** Interaction listing with all modalities: not useful

**Stage Six:** Identify clashes and modality properties: not useful/useful

**Stage Seven:** Assess the use of modalities: not useful/useful

**Stage Eight:** Assess overall usability of system and write up into a report:

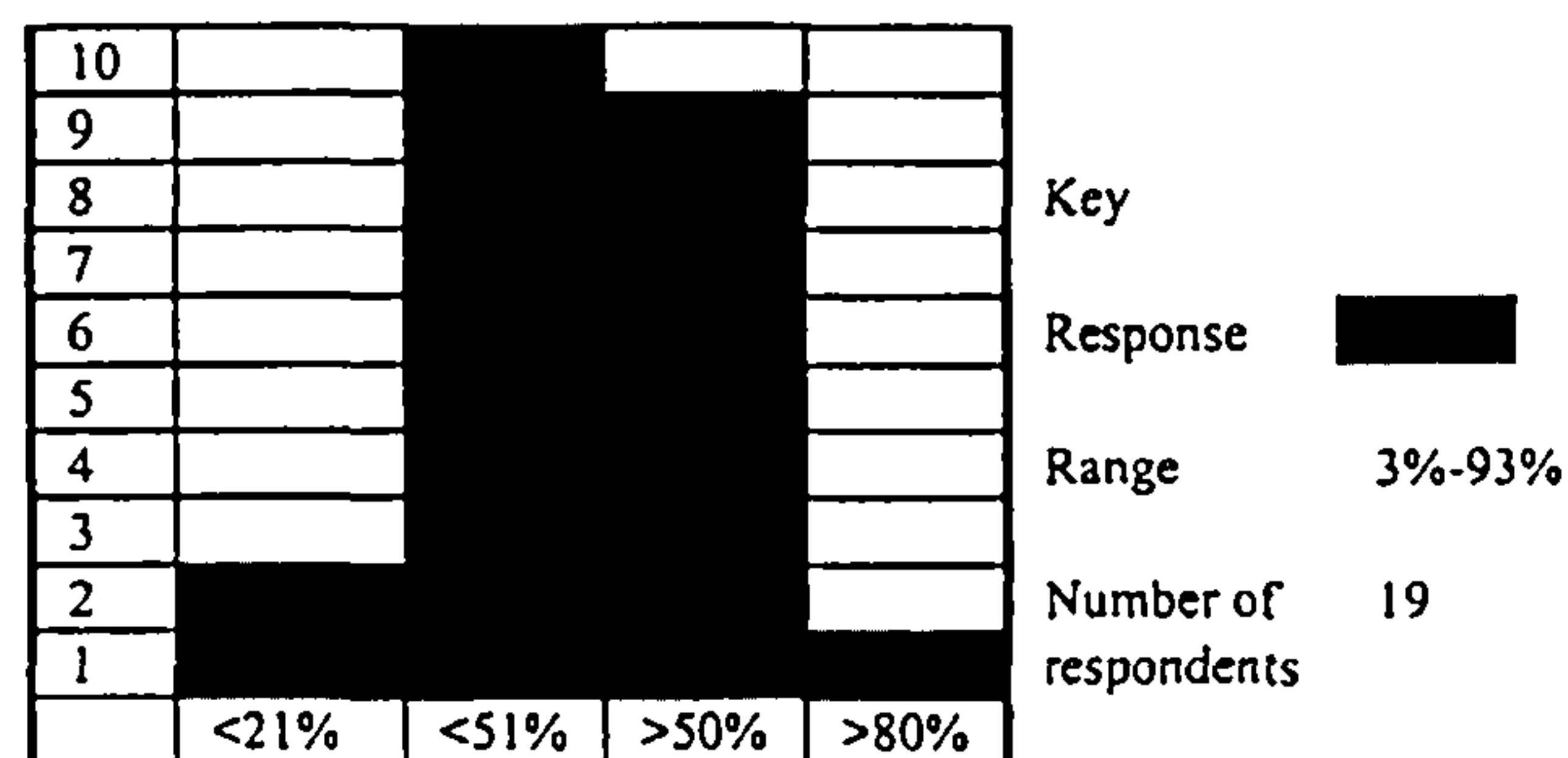
not useful/useful

12. Any other comments (list):

## 17.4 Second Questionnaire Results Tables

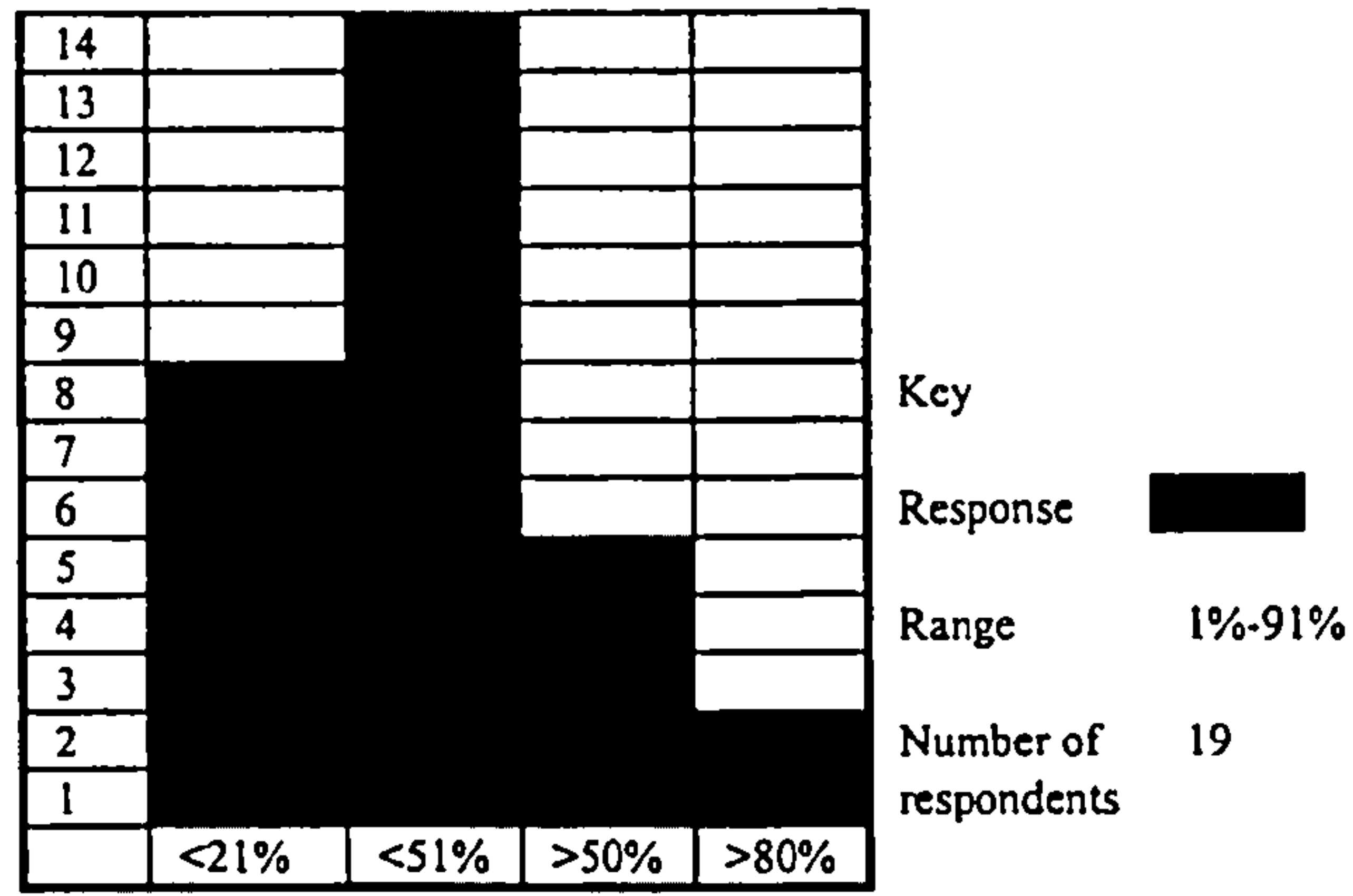
The responses presented here relate to subjects confidence in identifying modality properties and clashes, and their use of the methodology as a whole, and with regards to the individual stages.

The responses to seven questions are presented here, relating to: subjects confidence as to the correct identification of modality properties; subjects confidence as to the correct identification of modality clashes; how straightforward the subjects found the methodology to apply; subjects confidence in using the methodology again; how easy the subjects found each of the stages of the methodology to apply; and how useful in uncovering usability issues the subjects found each of the stages of the methodology.

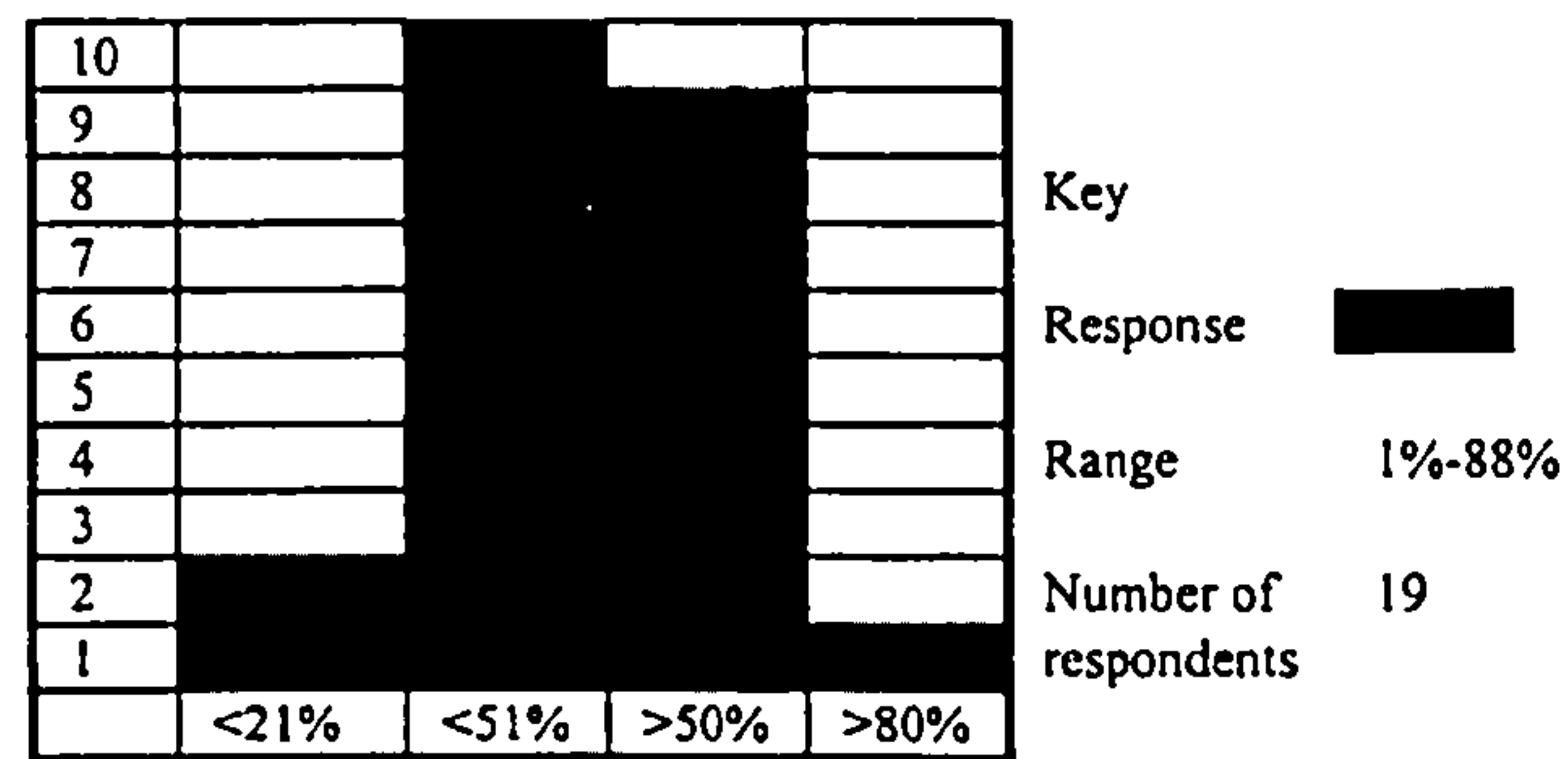


**Table 17-8: Responses to Question 3: How confident are you that you correctly identified the modality properties?**

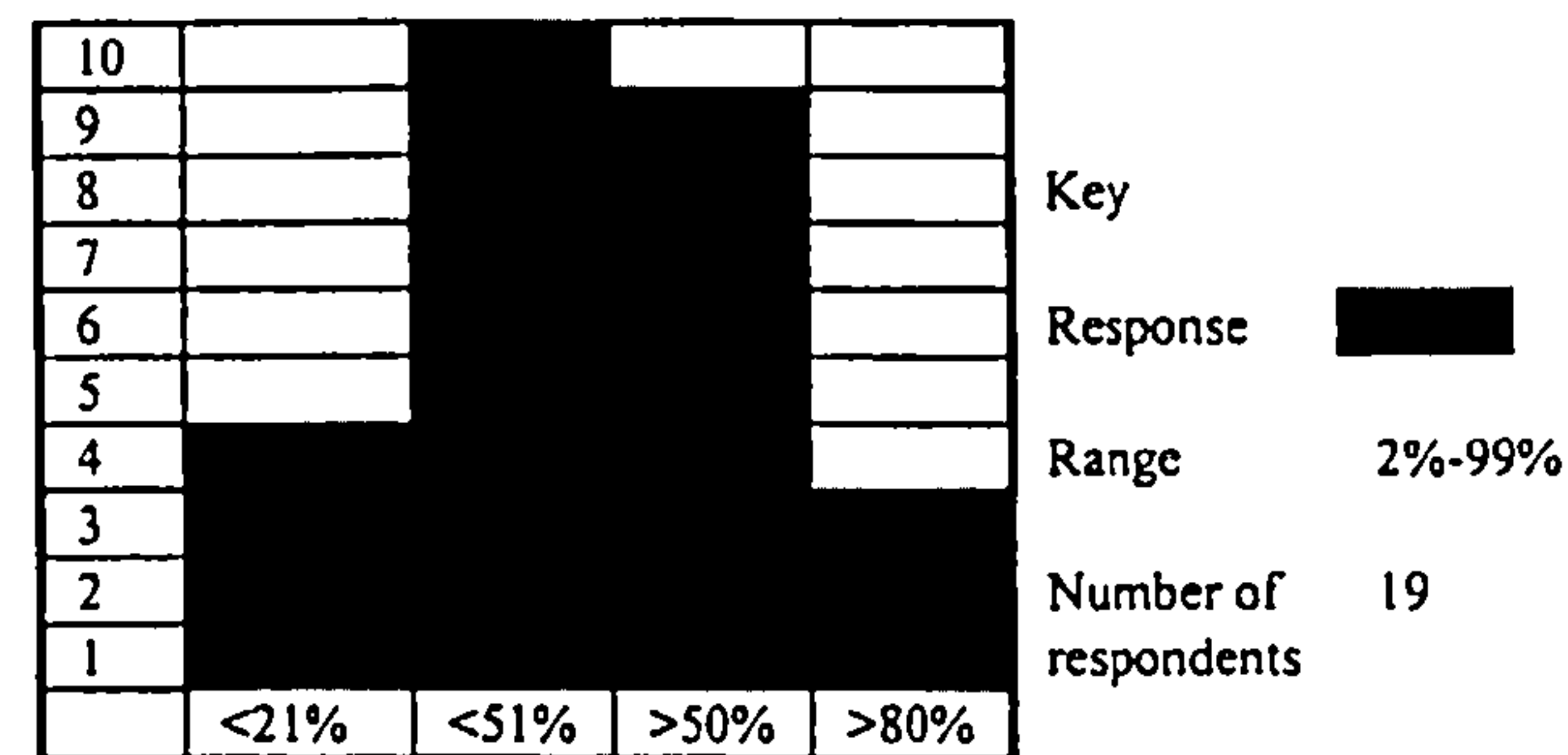




**Table 17-9: Responses to Question 5: How confident are you that you correctly identified modality clashes?**




**Table 17-10: Responses to Question 1: How straightforward did you find the methodology to apply?**




**Table 17-11: Responses to Question 6: How confident are you that you would be able to use this methodology again?**

13				
12				
11				
10				
9				
8				
7				
6				
5				
4				
3				
2				
1				
	<21%	<51%	>50%	>80%

Key  
Response   
Range 15%-100%  
Number of respondents 19

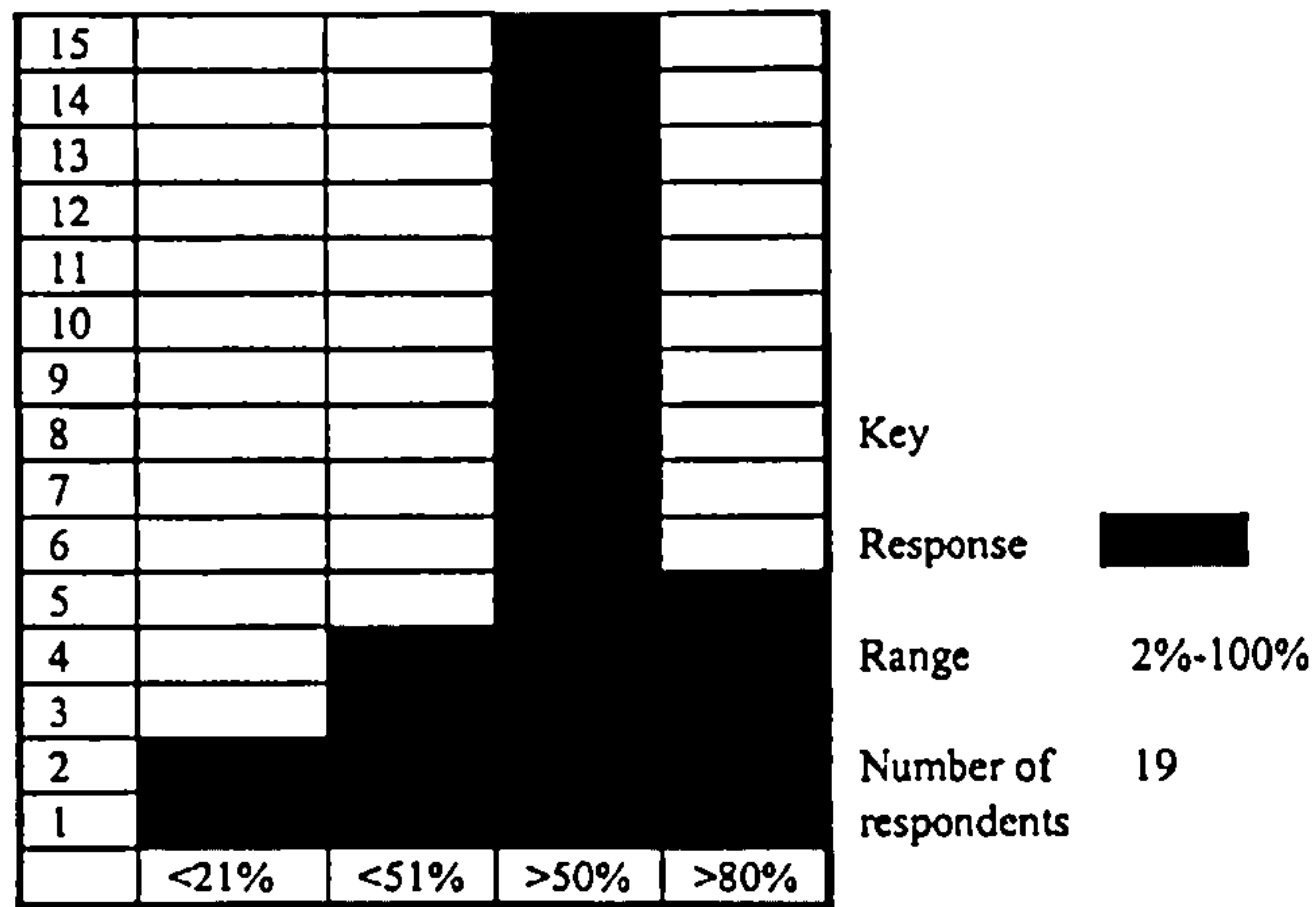
**Table 17-12: Responses to Question 8: How easy did you find the methodology as outlined in the pack to apply?**

17				
16				
15				
14				
13				
12				
11				
10				
9				
8				
7				
6				
5				
4				
3				
2				
1				
	<21%	<51%	>50%	>80%

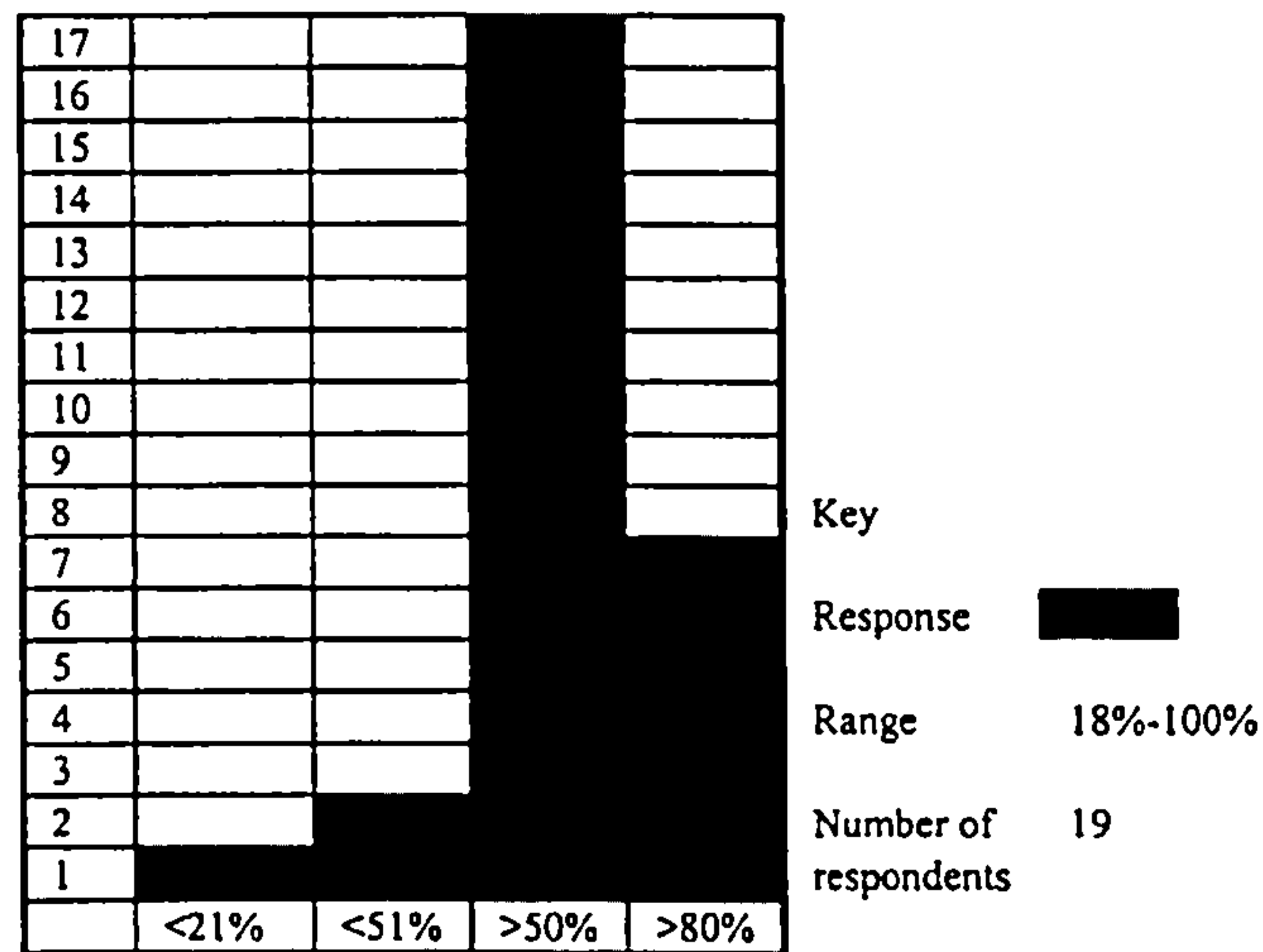
Key  
Response   
Range 15%-99%  
Number of respondents 19

**Table 17-13: Responses for Stage One of Question 10: Define the task that is to be analysed and justify its selection**

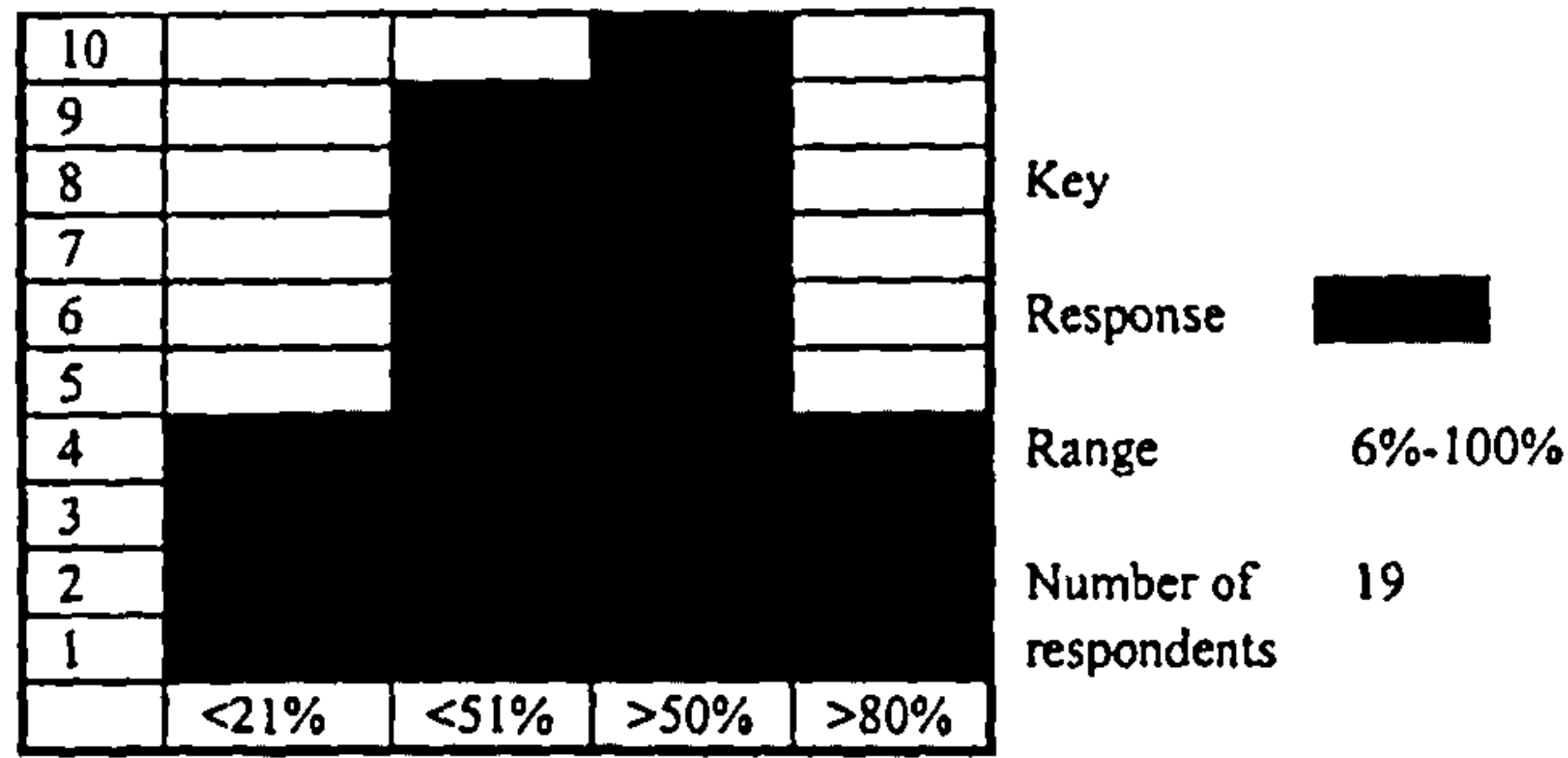




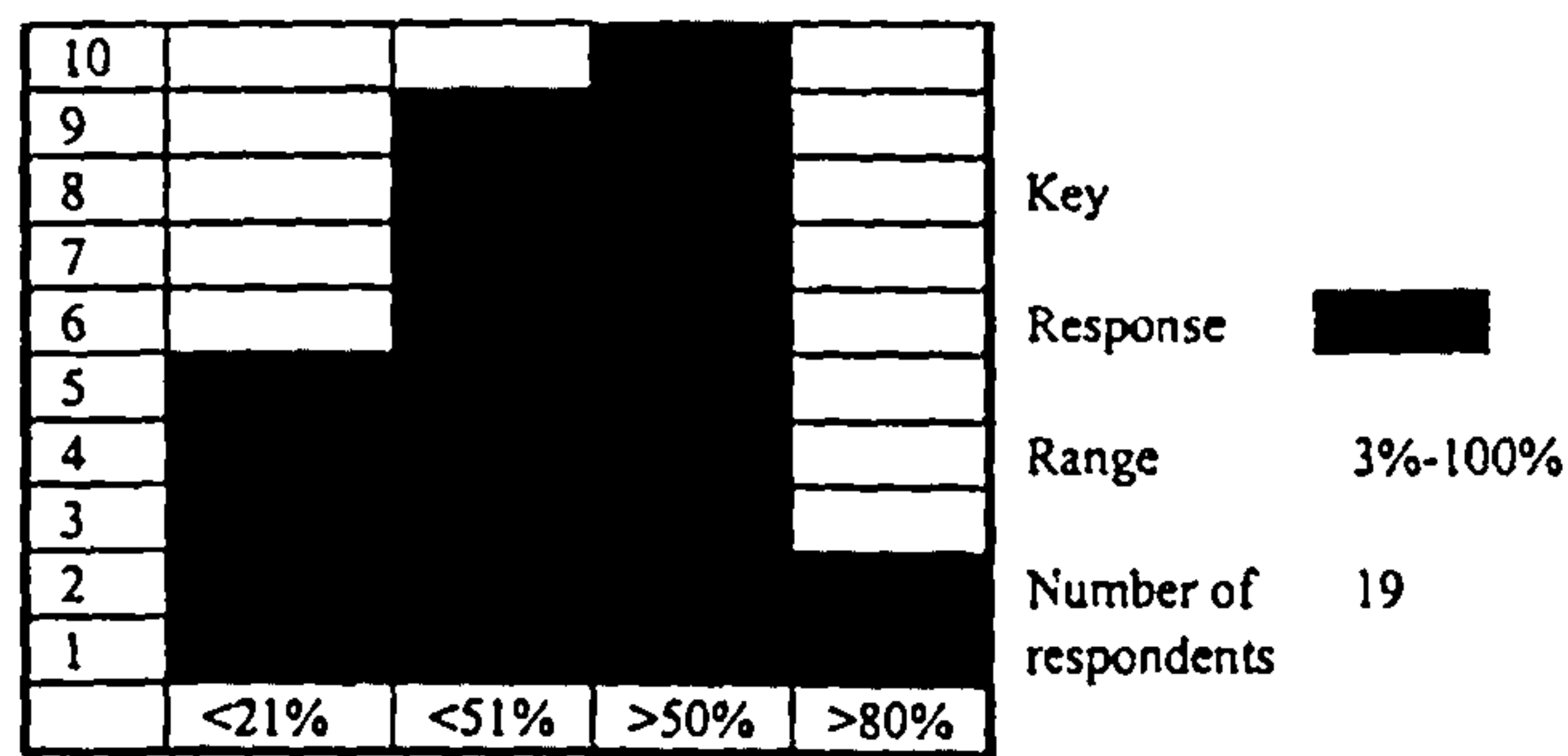
**Table 17-14: Responses to Stage Two of Question 10: List all the modalities used by system and user, and talk about the granularity of the analysis**



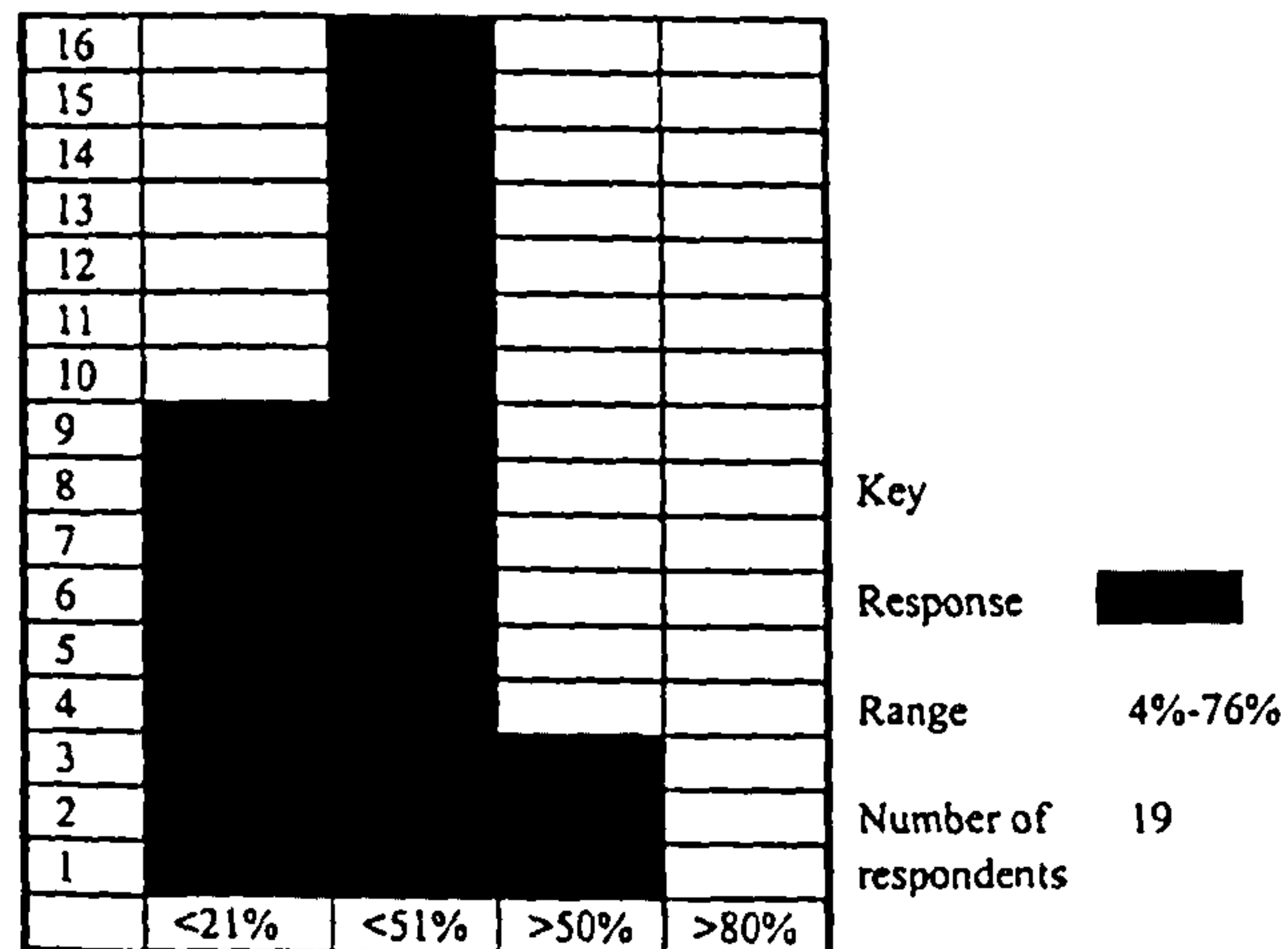
**Table 17-15: Responses to Stage Three of Question 10: Define the user, system and environmental profiles**



**Table 17-16: Responses to Stage Four of Question 10: Compare profiles to modality listings**

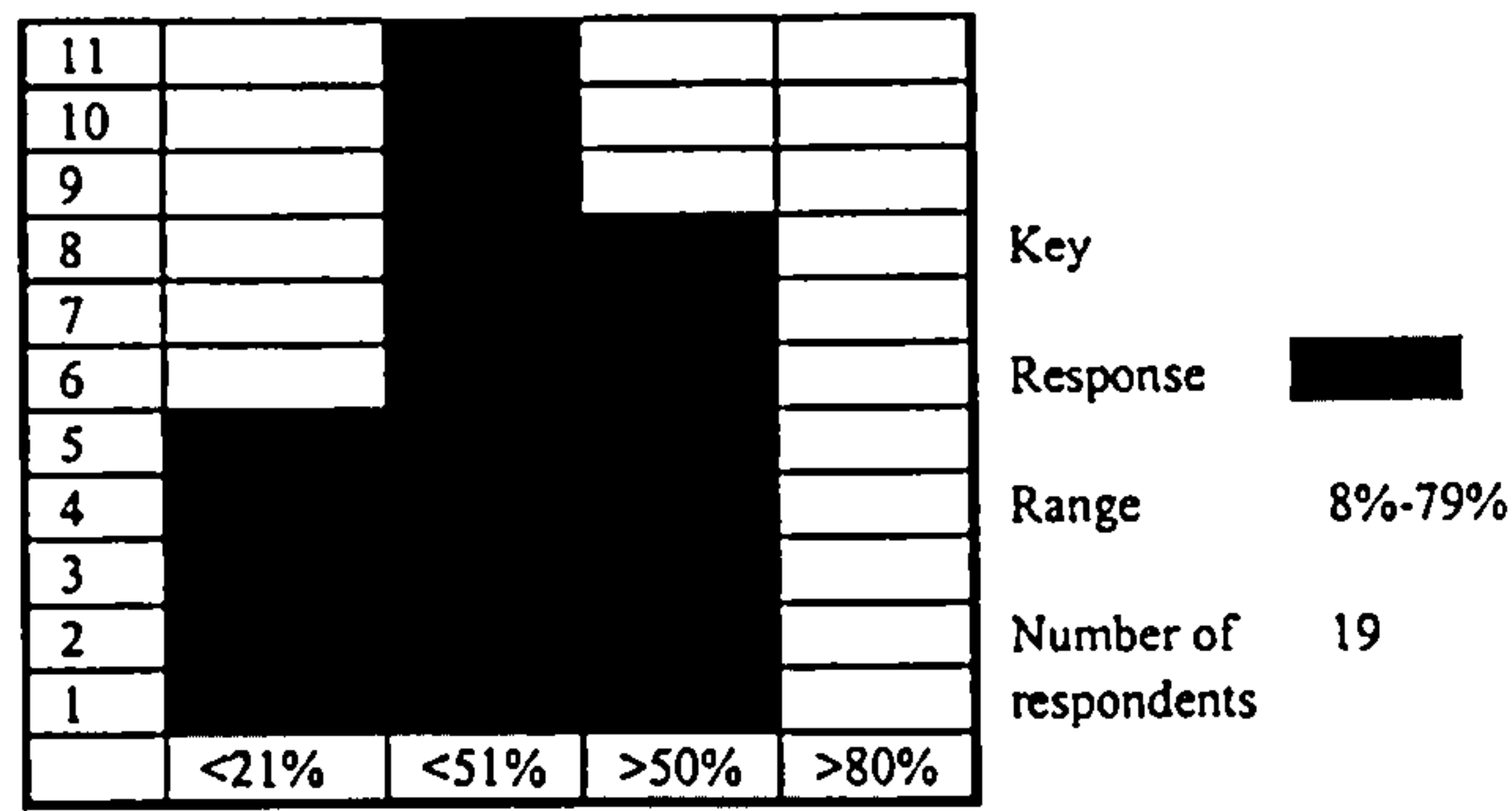


**Table 17-17: Responses to Stage Five of Question 10: Interaction listing of all modalities**

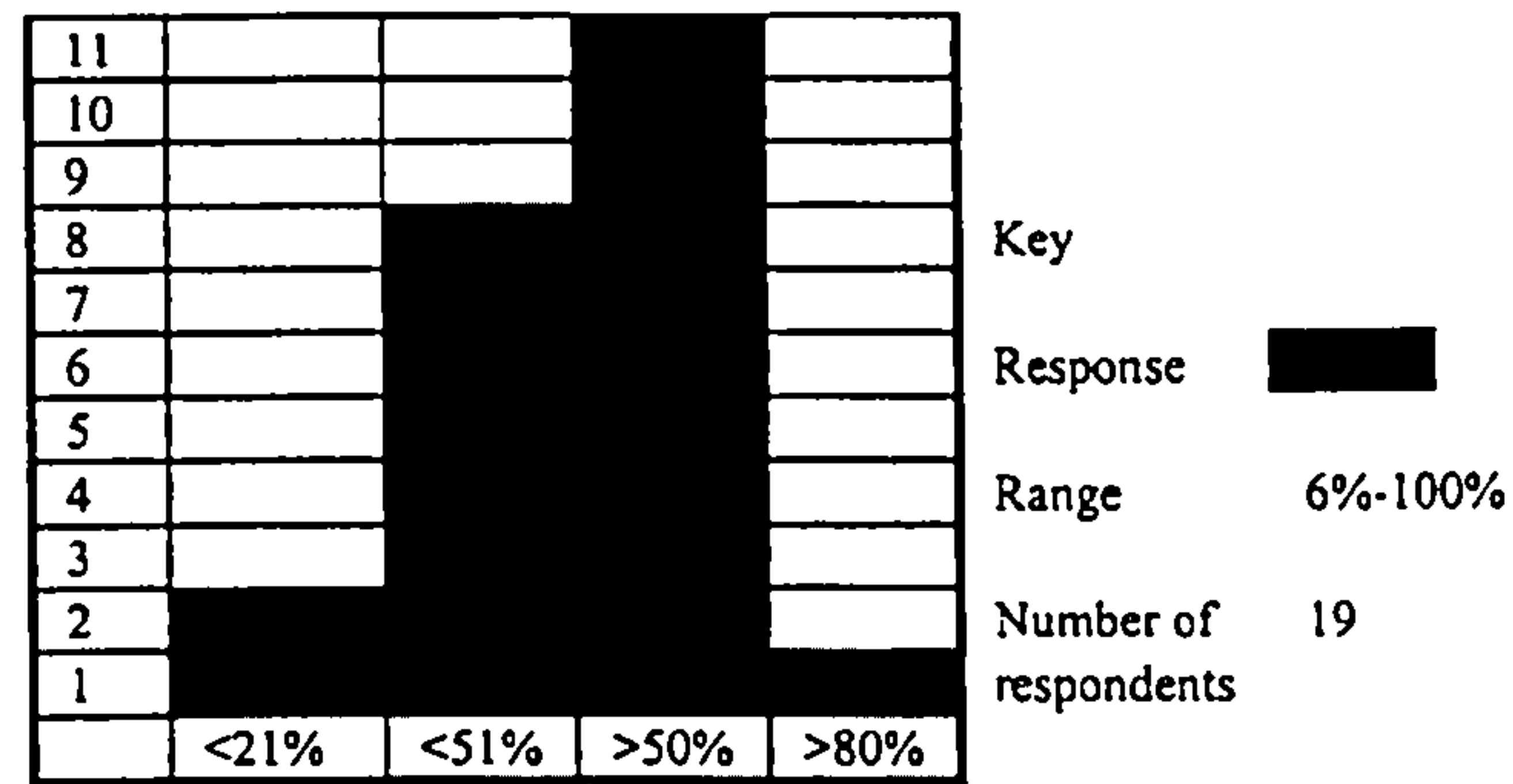


**Table 17-18: Responses to Stage Six of Question 10: Identify clashes and modality properties**

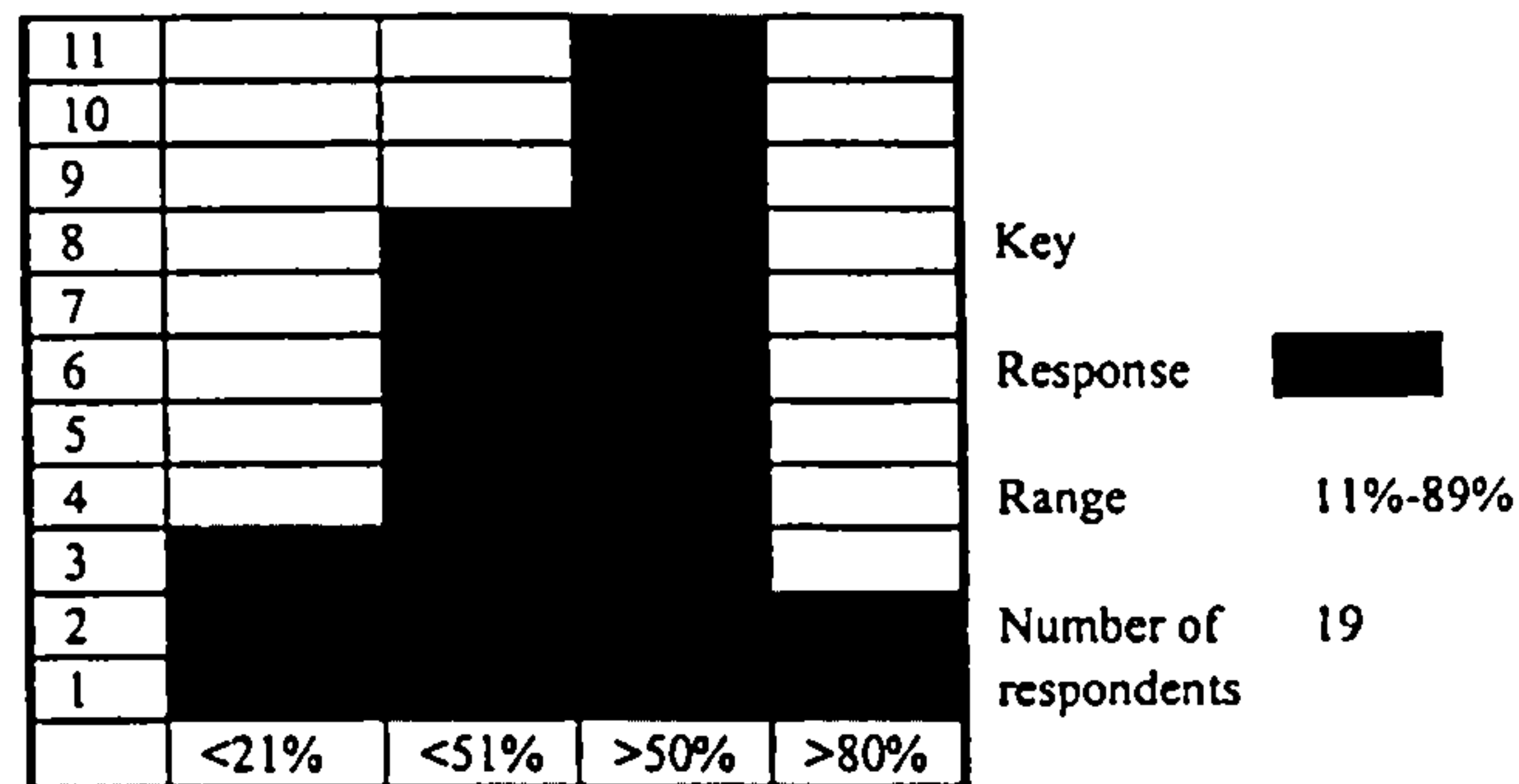




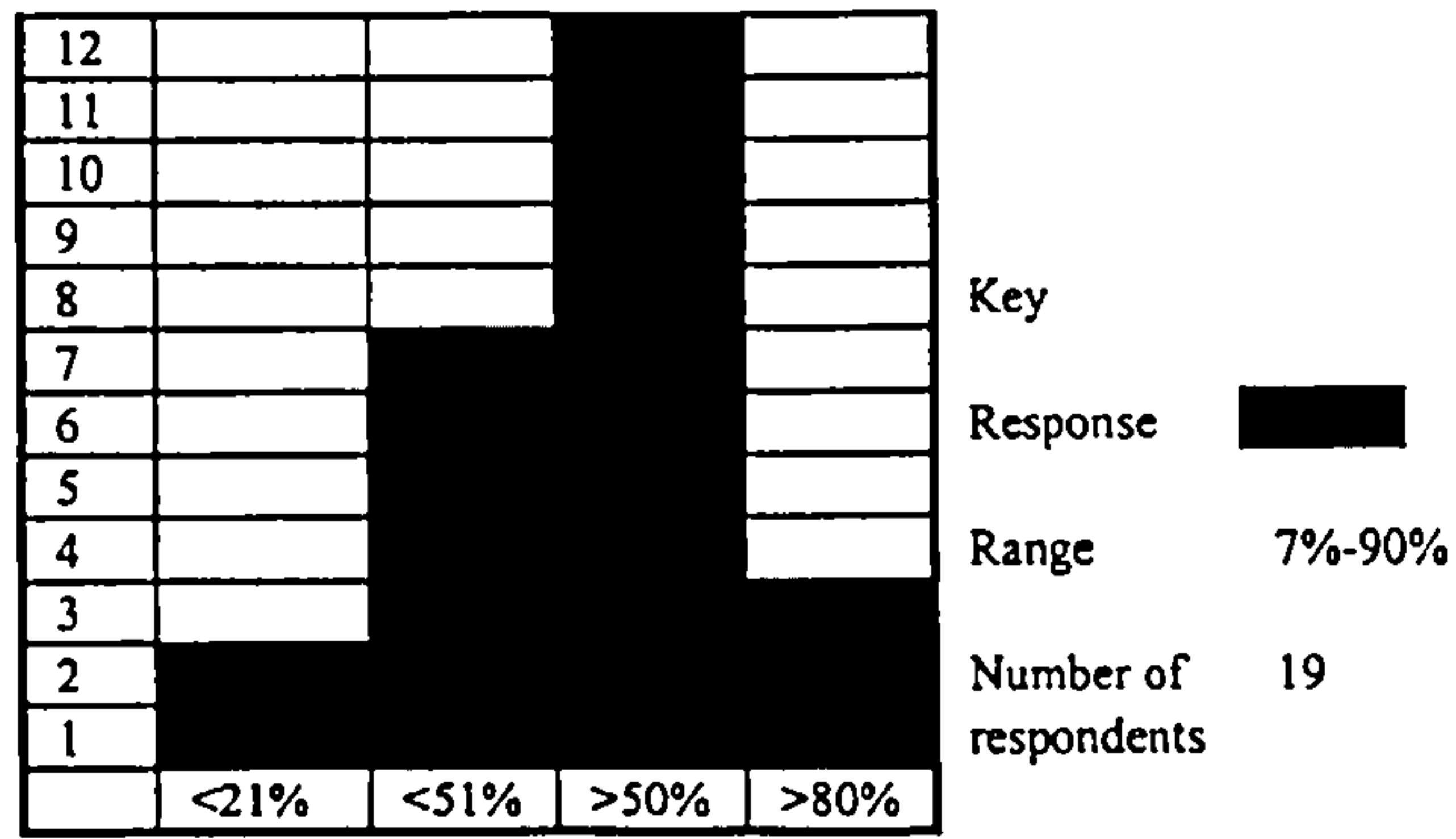
**Table 17-19: Responses to Stage Seven of Question 10: Assess the use of modalities**



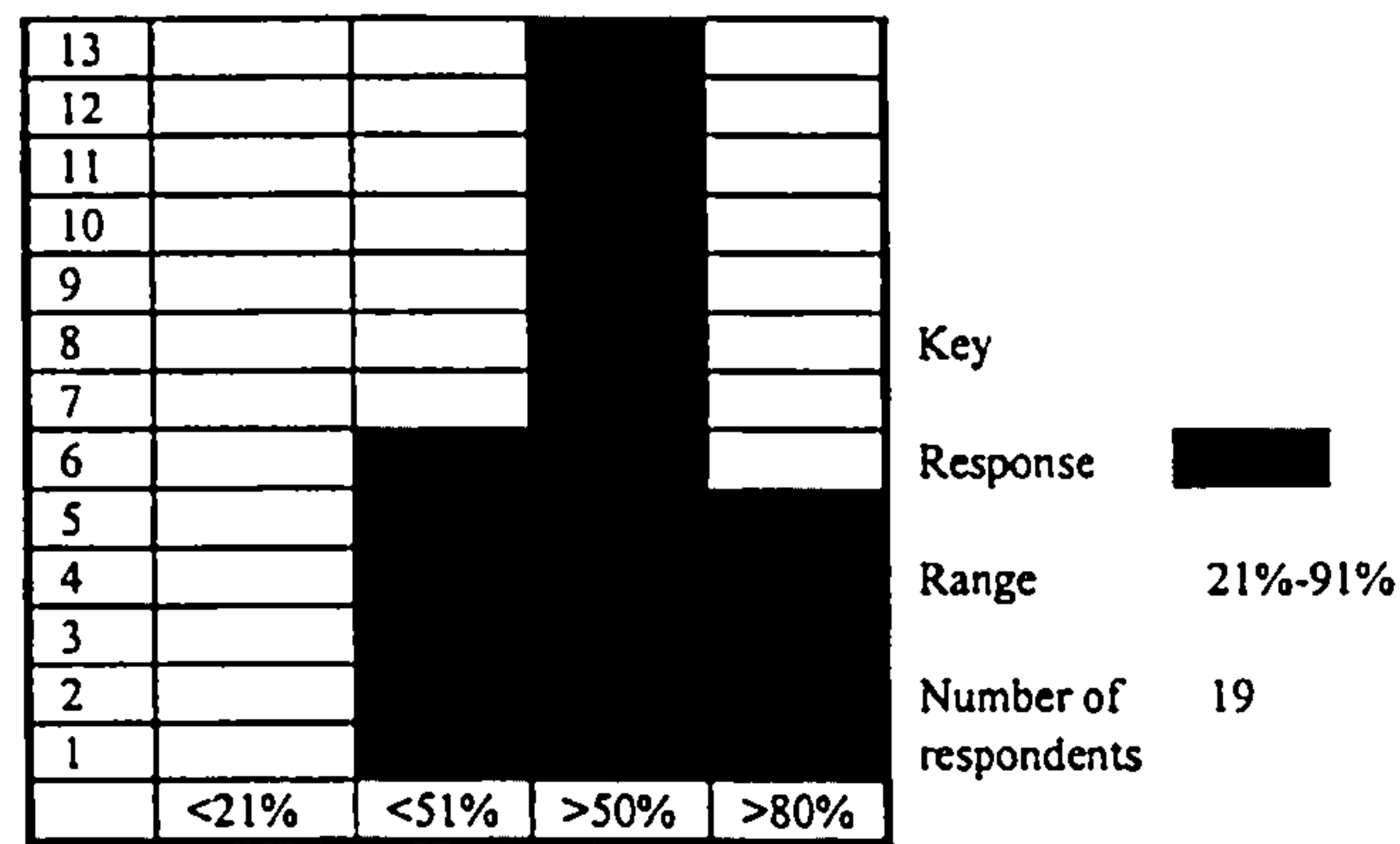
**Table 17-20: Responses to Stage Eight of Question 10: Assess overall usability of the system and write up into a report**



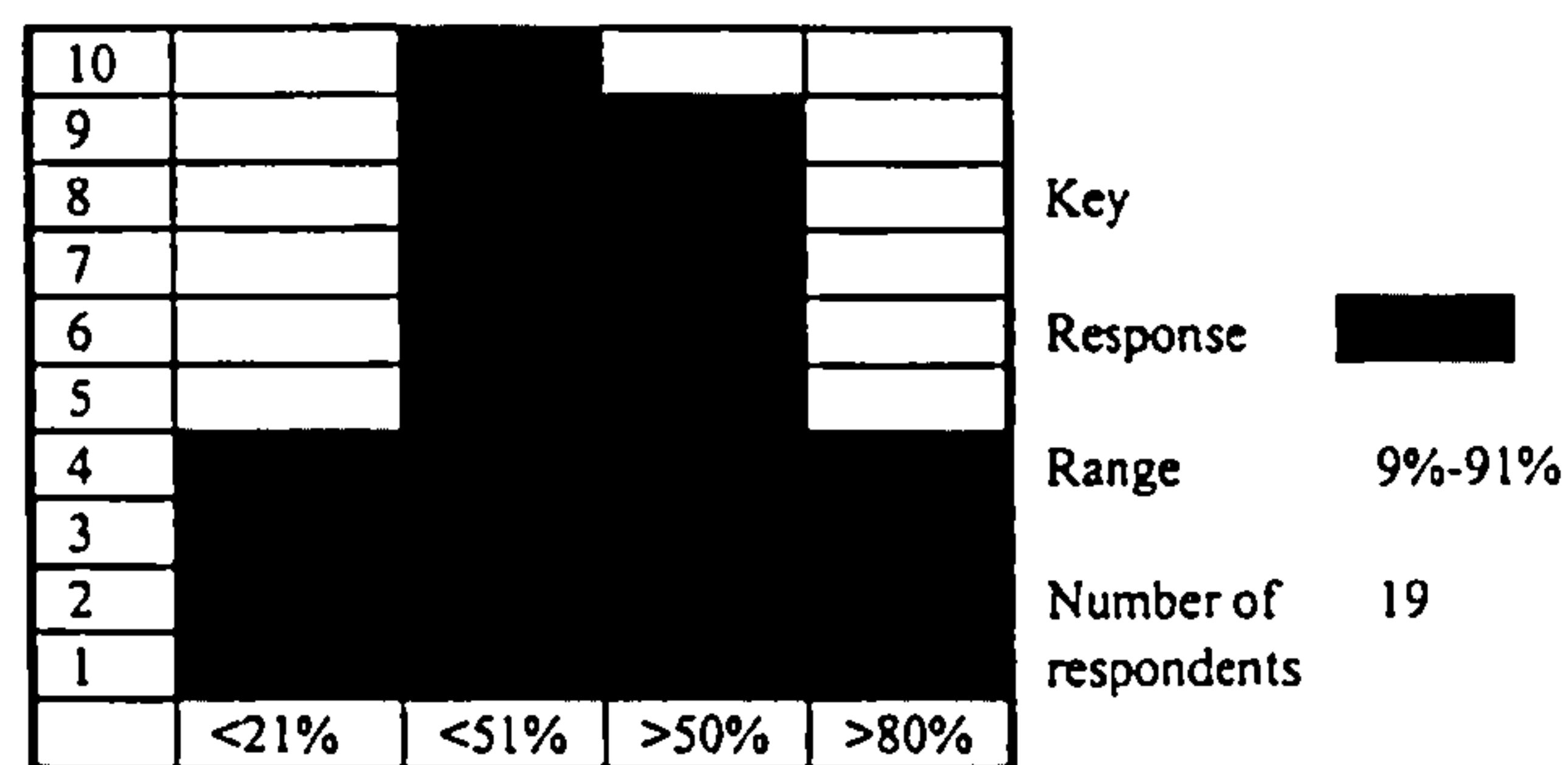
**Table 17-21: Responses to Stage One of Question 11: Define the task that is to be analysed and justify its selection**



**Table 17-22: Responses to Stage Two of Question 11: List all the modalities used by system and user, and talk about the granularity of the analysis**

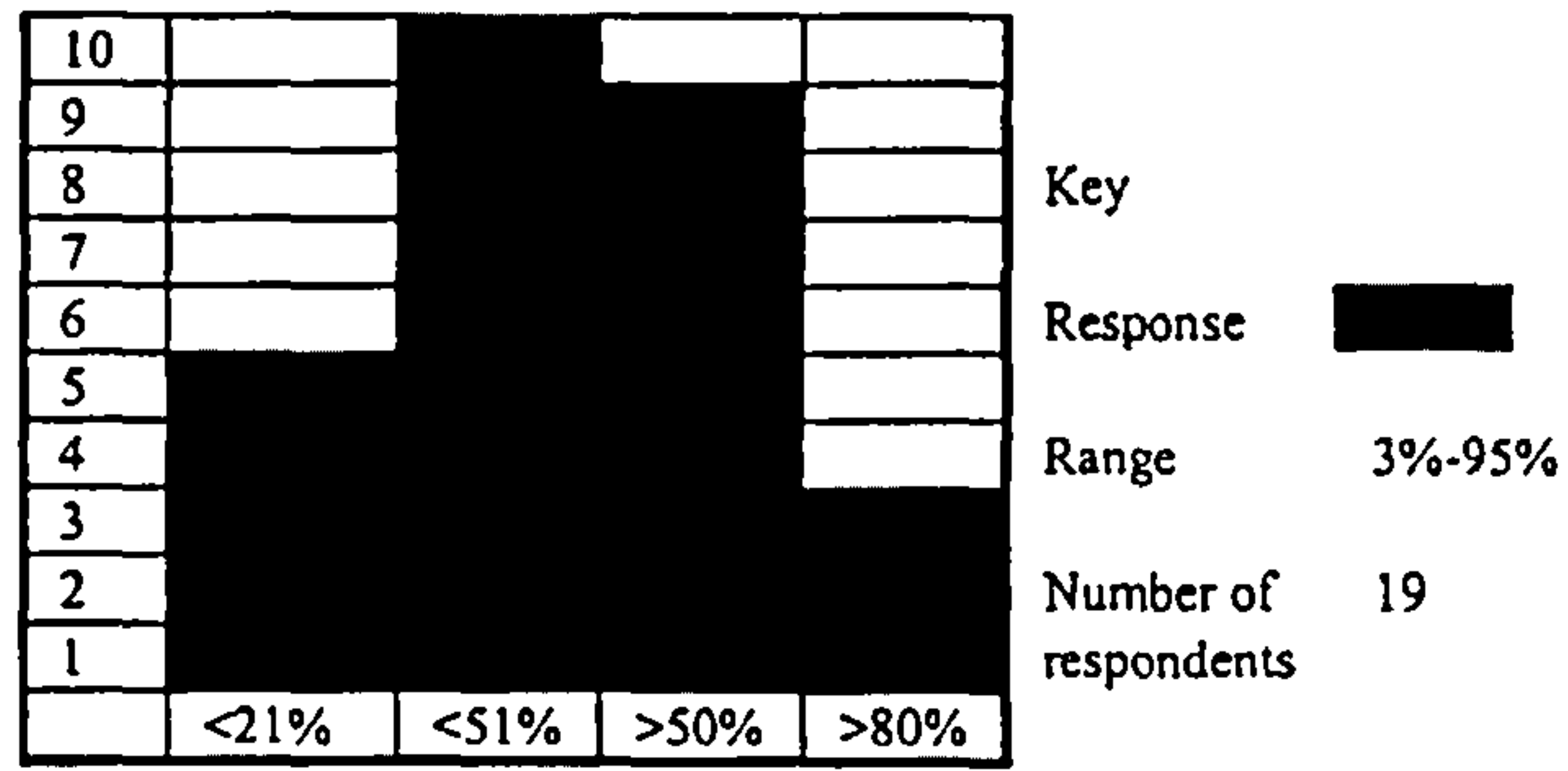


**Table 17-23: Responses to Stage Three of Question 11: Define the user, system and environmental profiles**

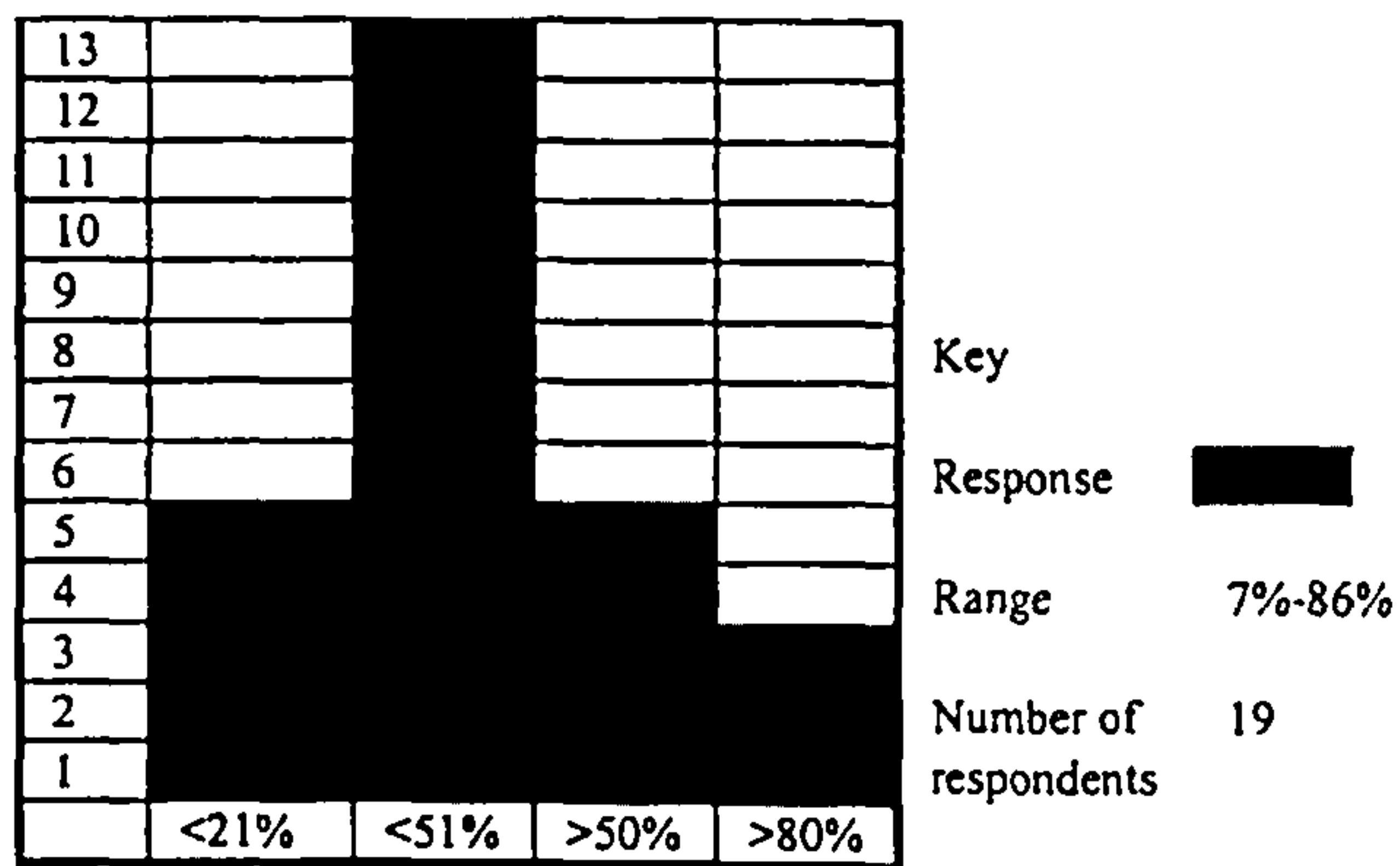


**Table 17-24: Responses to Stage Four of Question 11: Compare profiles to modality listings**

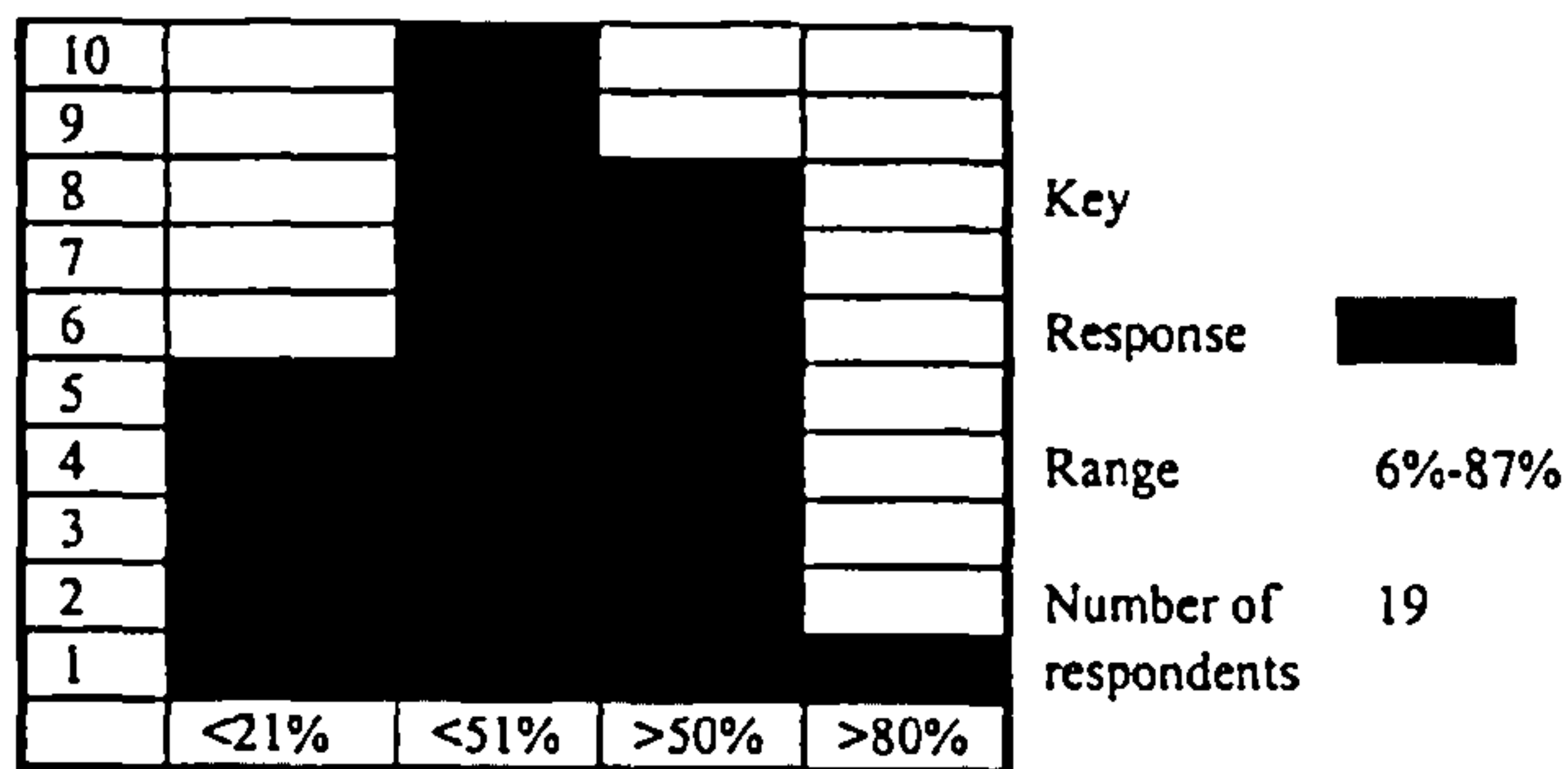




**Table 17-25: Responses to Stage Five of Question 11: Interaction listing of all modalities**




**Table 17-26: Responses to Stage Six of Question 11: Identify clashes and modality properties**



**Table 17-27: Responses to Stage Seven of Question 11: Assess the use of modalities**

10				
9				
8				
7				
6				
5				
4				
3				
2				
1				
	<21%	<51%	>50%	>80%

Key

Response 

Range 6%-95%

Number of respondents 19

**Table 17-28: Responses to Stage Eight of Question 11: Assess overall usability of the system and write up into a report**



# 18. APPENDIX H: VIDEO DATA ANALYSIS

The video data examined comprised six excerpts, each one showing a user performing a specific task and using a particular means of input, as summarised in the table below. The users and excerpts are described below.

EXCERPT:	One	Two	Three	Four	Five	Six
SECONDS:	123	83	89	50	89	525
INPUT:	Mouse	Voice	Gesture	Gesture	Voice	Mouse
TASK:	Feeding	Feeding	Feeding	Feeding	Feeding	Drinking
USER:	Expert	Expert	Expert	Novice	Novice	Novice
POSITIONS:	Pre-taught	Pre-taught	Pre-taught	Pre-taught	Pre-taught	Mixture

**Table 18-1: Summary table of video data excerpts**

## 18.1 The users

In excerpts one, two and three the arm user is the developer of the arm, Bernard Parsons. These excerpts are demonstrations of the arm's capabilities for movement, using options selected directly from the display and the speech and gesture input devices. Only pre-taught positions are used. Thus, the user is an expert user, who is not having to make adjustments to the arm. The user is the only person present on these three occasions.

In excerpts four, five and six the arm user is Paul Rocca, who works for the Spinal Injuries Association. Mr Rocca has suffered a previous spinal injury, and is classed as C4 Incomplete, and lacks full movement, strength and dexterity in his limbs. His role was to use the arm and provide feedback from the perspective of the intended user-group. Excerpts four and five were filmed during Mr Rocca's first session using the arm, and show the gesture and voice input devices being used to control the arm for the feeding exercise demonstrated in excerpts two and three using pre-taught positions. Excerpt six was filmed

during Mr Rocca's second session using the arm, by which point he had approximately three hours experience of its use, and uses a combination of pre-taught positions and manual adjustments, by means of mouse input, to fill a glass with water from a dispenser and position it so that the user can drink from it via a straw. Excerpts four and five were filmed with another person present, and conversations were held. Excerpt six was filmed with no one else present.

## **18.2 Excerpt One**

In excerpt one Bernard Parsons selects commands from a menu using a mouse in order to demonstrate the capability of the arm to move. The options are directly selected, and are a set of pre-taught positions. The arm is being used to scoop yoghurt out of a plate using a spoon, and bring the spoon round to the user's head. One sequence of movements is shown on the tape, starting from the arm lowering the spoon to the plate, and finishing with the arm swinging away from the user back towards the plate.

### **18.2.1 Length of interaction examined**

2.03 minutes

### **18.2.2 Listing of what happened**

0.00 - 0.06: Arm moves down towards plate and stops.

0.06 - 0.16: four clicks made by mouse.

0.16 - 0.24: arm extends into tray and stops.

0.24 - 0.33: five clicks made by mouse.

0.33 - 0.41: arm rises up and stops.

0.41 - 0.51: five clicks made by mouse.

0.51 - 1.14: arm swivels round to face user and stops.

1.14 - 1.20: five clicks made by mouse.

1.20 - 1.37: arm extends to mouth.



1.37 - 1.40: user eats from spoon. Has to raise head slightly to do so.

1.40 - 1.50: four clicks made by mouse.

1.50 - 2.03: arm swings back towards table and tape ends.

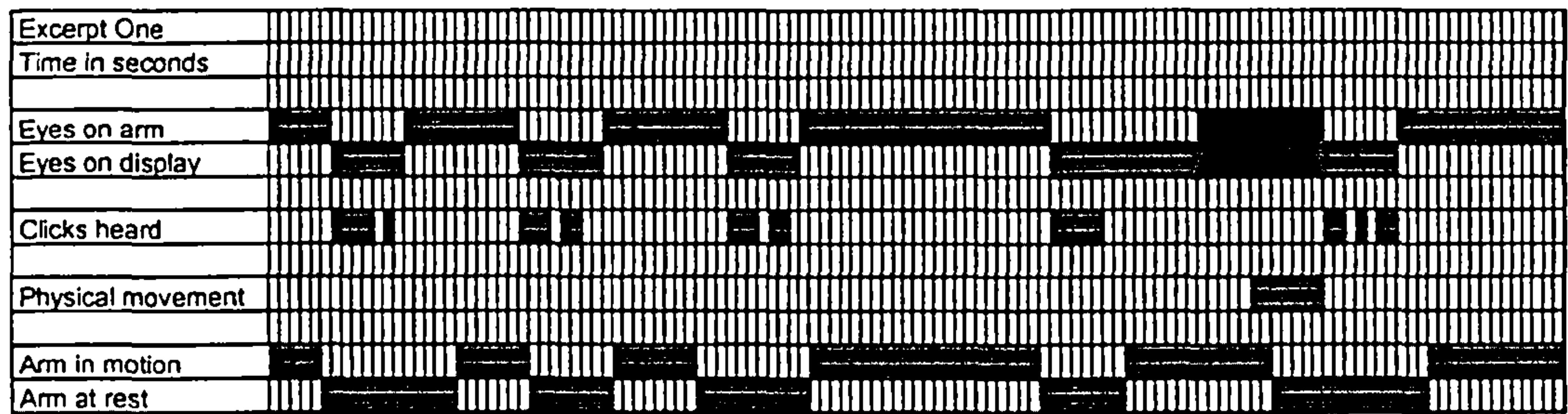


Figure 18-1: Excerpt One

The light grey area on the graph represents where the eyes are on both the robotic arm and the display simultaneously.

### 18.3 Excerpt two

This is similar to excerpt one, in that Bernard Parsons is using the arm to eat yoghurt. The only difference is that the voice input mechanism is being used. There are five possible commands used here: side, front, one, two, three. Side and front refer to the plate and by the user respectively, and the numbers refer to pre-recorded positions. One sequence of movements is shown, starting with the arm lowering the spoon to the plate, and finishing with the arm swinging round to face the user.

#### 18.3.1 Length of interaction examined

1.22 minutes

#### 18.3.2 Listing of what happened

click = sound of computer when choice received.

2.04 - 2.05: "Side"

2.07 - 2.11: "Side" click "Two" click

2.11 - 2.20: arm moves down towards plate on table and stops

2.22: "Side"

2.25 - 2.28: "Side" click "Three" click

2.28 - 2.41: arm extends into plate and stops

2.41: "Side"

2.43 - 2.46: "Side" click "One" click

2.46 - 3.01: arm lifts up and stops

3.02: "Front"

3.05 - 3.08: "Front" click "One" click

3.08 - 3.26: arm swivels round to face user and tape ends.

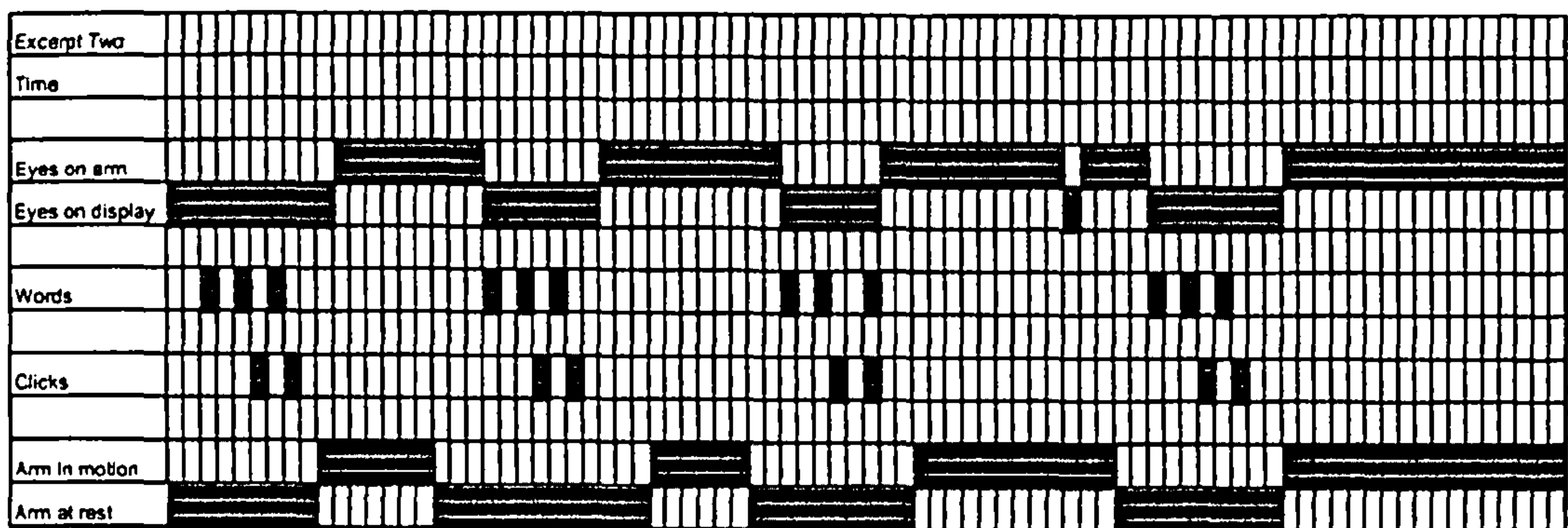


Figure 18-2: Excerpt Two

## 18.4 Excerpt three

This excerpt is similar to the previous two excerpts in format, with Mr Parsons using the gesture input to select the options for the arm. On the display is a sequence of menu options with a scanning bar beneath. When the bar is beneath the right option, the gesture input is used to select that option. One sequence of movements is shown, starting with the arm moving towards the plate, and finishing with the arm swinging round to face the user.



### **18.4.1 Length of interaction examined**

1.25 minutes

### **18.4.2 Listing of what happened**

Bing = noise made by computer when input received

3.33: bing

3.33 - 3.38: arm swings from horizontal position above table round towards user, stopping after a few inches.

3.39: bing

3.39 - 3.50: arm swings down to plate and stops

3.55 - 4.00: bing bing

4.00: arm starts to extend

4.07: bing. Arm still extending

4.09: arm stops extending

4.13: bing

4.17: arm starts to rise

4.27: bing. Arm still rising

4.29: arm stops rising

4.31: bing

4.31 - 4.58: arm rises and swings round to front of user and stops. End of tape.

At the end, the robotic arm is in front of the user's face. Since both the display and the arm are in the user's field of vision, it is impossible to determine which is being looked at. This is represented by the light grey on the 'Eyes on arm' and 'Eyes on display' lines.

The arm in motion has two movements joined together for last section of motion, as represented by lighter grey section on the graph.

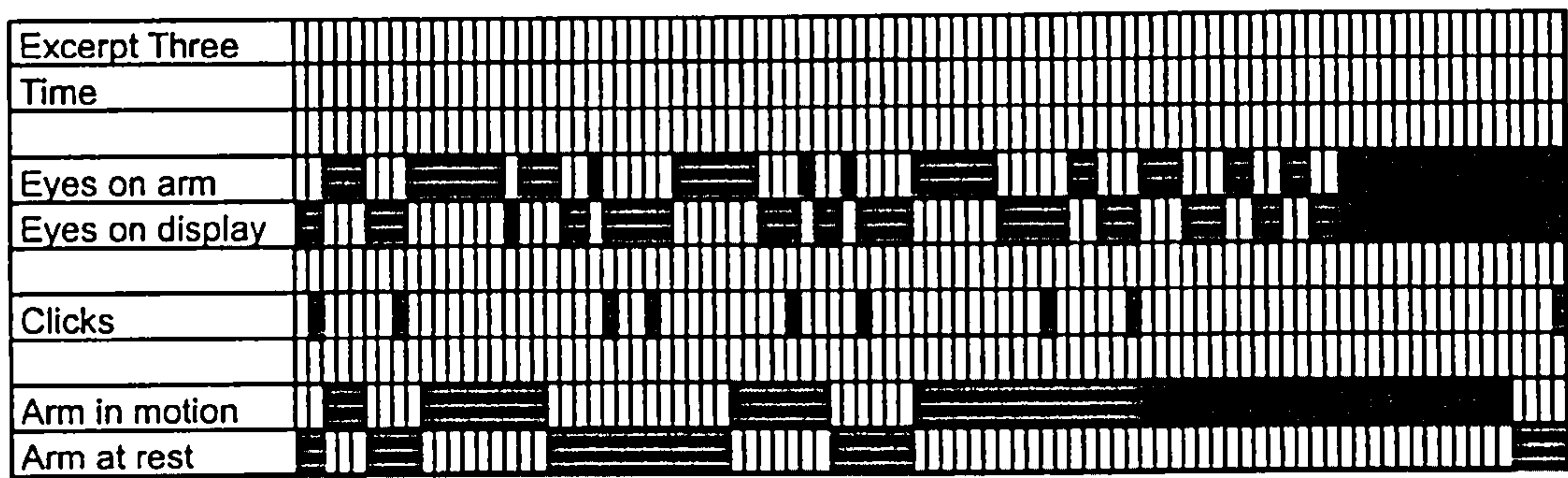


Figure 18-3: Excerpt Three

## 18.5 Excerpt four

Mr Rocca uses the gesture input device to select pre-taught positions to complete the same feeding exercise from excerpts one, two and three.

### 18.5.1 Length of interaction examined

1.32 minutes

### 18.5.2 Listing of what happened

5.56: arm moving away from user towards plate.

6.10: "Plate"

6.11: arm stops moving

6.11 - 6.14: "Plate" bing "Two" bing

6.17: arm starts to move down towards plate

6.24: "Plate", arm still moving

6.25: arm stops

6.26 - 6.31: "Plate" bing "Three" "Three" bing

6.35: arm starts to move into plate

6.40 - 6.44: "Front" bing "Two" bing, arm still moving



6.45: arm stops

6.46 - 7.06: arm moves up, round to user and stops

7.06 - 7.10: user eats yogurt

7.10 - 7.14: "Plate" bing "One" bing

7.18: arm starts moving back to table once more

7.20: jumbled user comments, finishing with "speaking with your mouth full", arm still moving

7.28: arm still moving, end of tape

### 18.5.3 Speech from excerpt

1. That's just a matter of having the right angle innit, that's all, that's not a problem.
2. Yeah.
3. [Indistinguishable]
4. I think it's on slow, innit?
5. Looks like it, yeah.
6. Ah, look at that.
7. Perfect.
8. Excellent.

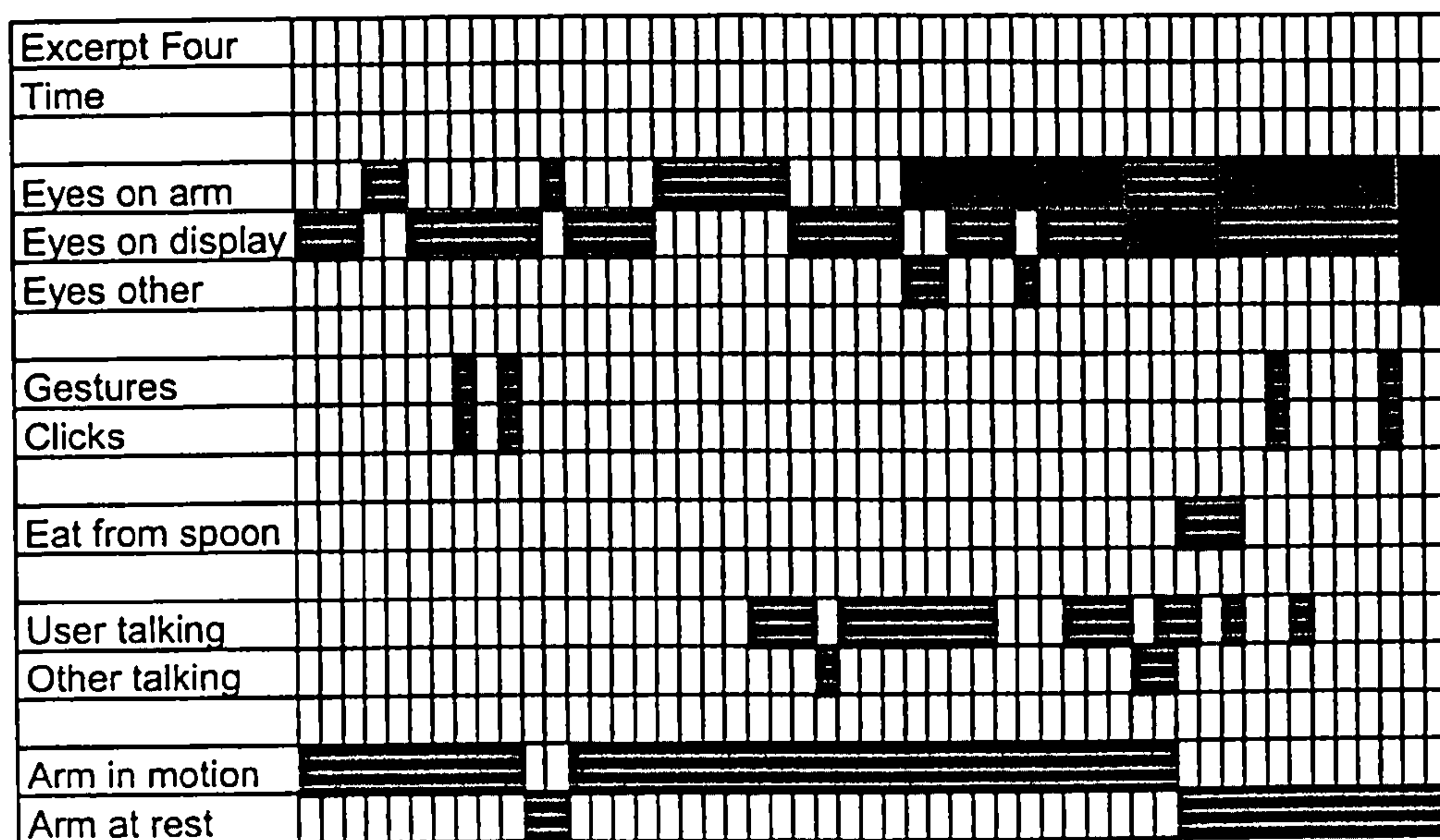


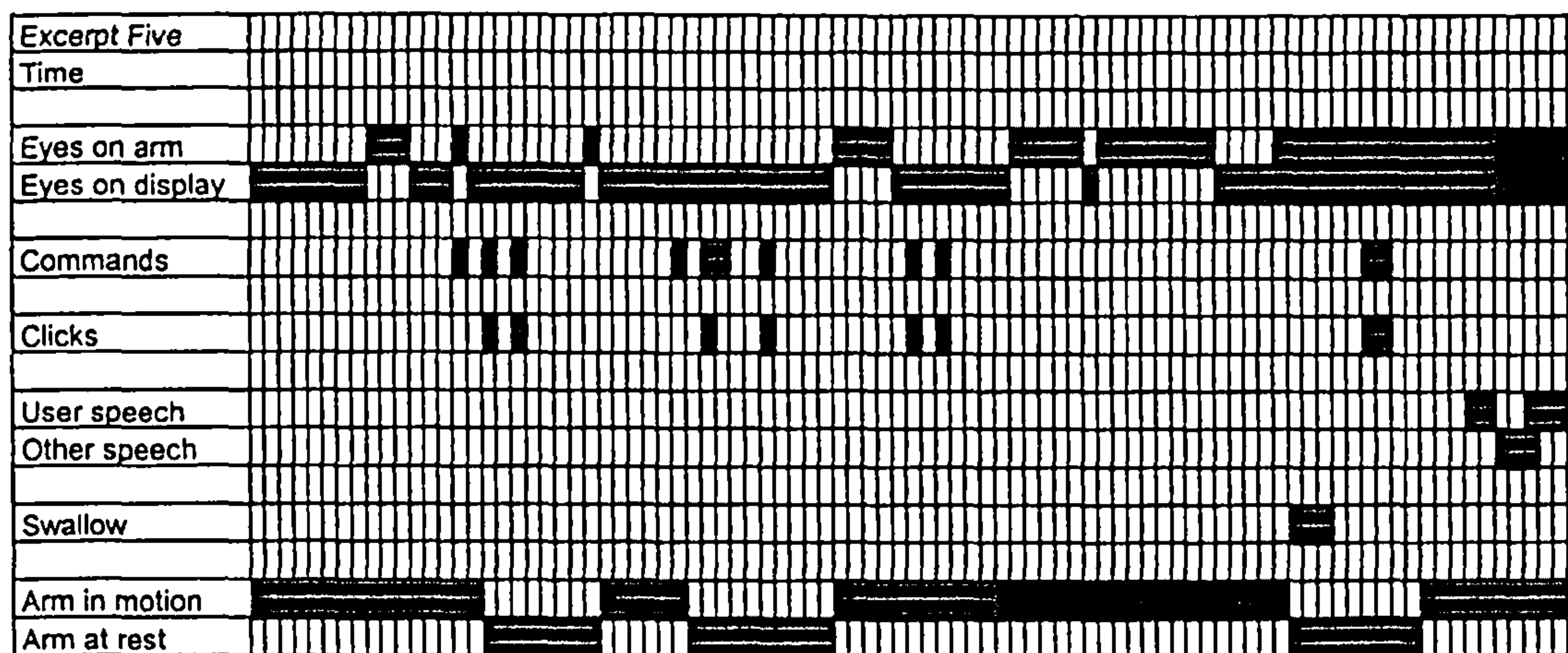
Figure 18-4: Excerpt Four

Note that for a long section, the arm is in the field of vision which includes the display. This is represented by the light grey. In the last part of the tape, visibility is blocked by Bernard coming to turn off the video camera. This is represented by black on the graph.

## 18.6 Excerpt five

In excerpt five Mr Rocca uses the voice input device to select pre-taught positions to complete the same feeding exercise from excerpts one, two and three. The word “plate” is used in this exercise to denote “side” from the previous excerpts. Three additional comments were made by the user:

1. You’ve got to wait until your mouths full then teach it speech
2. Well this is what I was wondering its a serious thing
3. Speaking with your mouth full





## 18.7 Excerpt six

This shows Mr Rocca performing a more demanding task using the robotic arm. This was to gain feedback on the effectiveness of a new gripper, and to determine whether the efforts to decrease noise from the arm had been successful (personal communication from B. Parsons). A combination of pre-taught and manually selected options was used, selected using a mouse. This sequence uses the shoulder joint in addition to other joints. The shoulder joint was not used in the previous excerpts due to its shaking, which has now been corrected. The sequence shows a straw being lowered into a cup, the cup filled from a dispenser, and the cup lifted and positioned in front of the user so that the user can drink from the straw. The drink is then returned to the table.

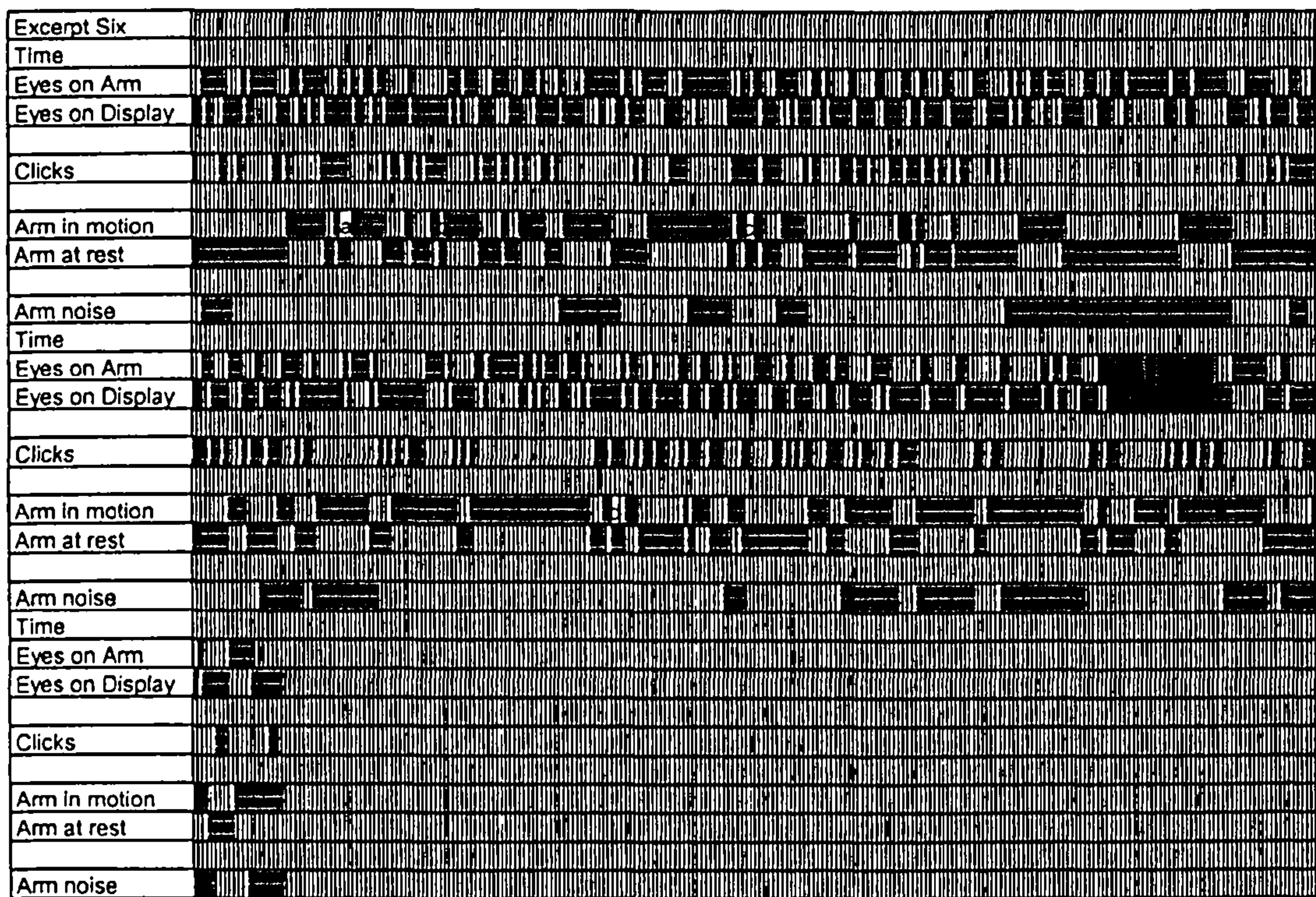


Figure 18-6: Excerpt Six

The light grey section represents where the arm and the display were both in the field of vision.

# **19. APPENDIX I: THE RE-ANALYSES**

## **19.1 STN re-analysed**

### **1. Long sequence of operators to move arm**

Since the STN shows the number of states that the user has to navigate through before the robotic arm can be moved, this issue should have been identified in the original analysis. That it was not identified shows the extent to which the analysis was dependent on the craft skill (or lack thereof) of the analyst.

### **2. Inability to backtrack**

This issue is apparent from the STN, and was identified as a problem. However, the identification of this issue was possibly influenced by the explicit mention of this kind of problem in a discussion on “undo” in the source materials (Dix et al, 1993, p.291). This shows the effect that the source materials of a technique has on the application of a technique.

### **3. Difficulty of choosing between Move Arm or Move**

The STN concentrates on the actual choice between the system states, rather than on the difficulties the user has in choosing between them. It is therefore not an issue that the STN on its own would be expected to identify, but might have been identified through craft skill.

### **4. Lack of short cuts**

Since the STN explicitly shows the possible path of the interaction through the various states, the lack of short-cuts was an issue that might have become apparent if the analyst had been looking for it. This is therefore an issue that is a combination of craft skill and representation. That it was not noticed was possibly because the analyst’s attention was more on obtaining the correct representation of the system states.



### **5. Continue versus Go: Continue seen as redundant**

This was an issue found through drawing the STN, since the use of Continue creates more states for a user to navigate through, and adds little to the functionality of the interface. This is the kind of issue that the use of the STN should make apparent, and did.

### **6. Confusion over joint called Arm**

The STN did not go into the detail of the individual options, so this issue did not arise. If the STN had been done at a different level of abstraction, this issue might have been identified through the craft skill of the analyst. It is not something that the STN would identify however, since it is concerned more with the user understanding of what a particular option choice means rather than with the option choice itself. Thus this issue highlights questions associated with both craft skill and appropriate levels of abstraction.

### **7. Gesture input with twice as many operations as voice because dependent on cursor movement**

The STN was not written at the level of abstraction which would identify this issue. If it had been, this issue would probably have been identified, since it would be concerned with the number of states and transitions. For the gesture input, there are a series of states and transitions between them as opposed to the voice input which has one state with multiple transitions coming from it. This raises questions concerning the appropriate level of abstraction of an analysis.

### **8. Problem if head moved to look at arm while gesture system operational may be interpreted as a command**

This is not an issue that could be identified from the STN since the STN is concerned with the choices available for the user in moving from one state to another, rather than how that choice is made or the problems the user might have with that choice.

### **9. If user pauses in middle of saying "Move arm"...**

This is not an issue that could be identified from the STN since the STN is concerned with the choices available for the user in moving from one state to another, rather than how that choice is made or the problems the user might have with that choice.

#### **10. If user engaged in conversation...**

This is not an issue that could be identified from the STN since the STN is concerned with the choices available for the user in moving from one state to another, rather than how that choice is made or the problems the user might have with that choice.

#### **11. Lack of feedback after Move Arm selected, no indication that whole arm is to be moved**

The feedback for the interface had not been implemented at this stage, so could not have been represented. Even if the feedback had been implemented this was not an issue that could be identified from the STN since the STN is concerned with the choices available for the user in moving from one state to another, rather than in providing confirmation that a particular choice has been made.

#### **12. Problems of determining left and right, especially when arm contorted**

This is not an issue that could be identified from the STN, since it relates to how the user views the arm rather than the states and transitions of the arm.

#### **13. User cannot check direction choice until arm starts to move**

This is not an issue that could be identified from the STN since the STN is concerned with the choices available for the user in moving from one state to another, rather than in providing confirmation that a particular choice has been made.

#### **14. Time taken to interact with system to stop arm**

This was not a problem identified by the STN, since this is a problem related to the user making the choice rather than the actual choice. Vocalising the word or making the gesture may take too long for the arm to stop in exactly the correct place. This is therefore a matter of user judgement rather than states and transitions, and as such the STN would not be applicable.

#### **15. Similarity between moving joint and moving whole arm**

This was an issue found through drawing the STN, since both moving the joint and moving the arm follow similar pattern of states and transitions. This is the kind of issue that the use of the STN should make apparent and did.



## **16. Illegal options**

This issue was not represented on the STN. There was no state showing that the arm had reached its limit of movement, nor was there an end option leading from the travel until stop state which might also represent it. This shows how difficult it is to draw STNs correctly, and relates to the level of skill of the analyst in determining how the system states should be represented. If the STN diagram had been correctly drawn, it is still unlikely that this issue would have been identified without explicitly checking for illegal options. This is similar to issues discussed in the action properties section of the source materials (Dix et al, 1993, p.288) which acknowledge how difficult it is to identify these issues. On this occasion, although the STN allowed for this issue to be identified, the ease of identifying this issue was dependent upon the craft skill of the analyst.

## **17. Mismatch between way that arm works and way that user would move arm**

This is not an issue that could be identified from the STN since the STN is concerned with the choices available for the user in moving from one state to another, rather than how that choice is made or the problems the user might have with that choice. However, the difference between state and transitions represented by the STN and how a user might naturally go about moving an arm might have been identified through the craft skill of an analyst.

## **18. Not clear that End returns user to main menu**

This is not an issue that could be identified from the STN since the STN is concerned with the choices available for the user in moving from one state to another, rather than how that choice is made or the problems the user might have with that choice. In the STN it is clear that End returns the user to the main menu.

## **19. End having two meanings**

This is not an issue that could be identified from the STN since the STN is concerned with the choices available for the user in moving from one state to another, rather than how that choice is made or the problems the user might have with that choice. From the STN it is clear that End returns the user to the main menu.

## **20. Lighting conditions**

This is an issue relating to the robotic arm's environment rather than the states and transitions of the system, and would therefore not have been identified by the STN analysis. It is unlikely that the craft skill of the analysis would have identified this.

## **21. Difficulty for user to move field of vision**

This is a user concern, therefore the STN analysis, which concentrates on the system functionality, would not be able to identify this issue.

## **22. User looking one way, menu options in other direction**

The STN is concerned with the choices available for the user in moving from one state to another, rather than how that choice is made or the problems the user might have with that choice, so this is not an issue that could be identified.

## **23. Difficulty of judging arm movements**

The STN is concerned with the choices available for the user in moving from one state to another, rather than how that choice is made or the problems the user might have with that choice, so this is not an issue that could be identified. However, the difference between the states and transitions represented by the STN, compared with how a user might go about moving an arm, might have been identified by the craft skill of an analyst.

## **19.2 Discussion of STN Re-analysis**

When the STN analysis was re-examined, based on (Dix et al, 1993), two of the three issues originally identified were found to be identified through the use of the method rather than the craft skill of the analyst. STN was able to identify issue 5 (Continue versus Go: Continue seen as redundant) since the use of the option Continue creates more states for a user to navigate through, and adds little to the functionality of the interface. Similarly issue 15 (similarity between moving joint and moving whole arm) was identified since both operations follow similar pattern of states and transitions. STN is specifically concerned with the states of the interface and the functional movement between those states.

The identification of issue 2 (inability to backtrack) was outside the direct scope of STN and was influenced by the explicit mention of this kind of problem in a discussion on



“undo” in the source materials (Dix et al, 1993, p.291), demonstrating the effect that the source materials can have on the application of a particular technique.

Issue 1 (long sequence of operators to move arm) was not identified by the original analysis, despite being within the scope of STN with the diagram clearly showing the large number of states that the user has to navigate through before the robotic arm can be moved. This demonstrates the extent to which the analysis was dependent on the craft skill of the analyst in being able to identify the usability issues.

On re-examination of issue16 (illegal options) it was found that the STN diagram was incomplete. There was no state showing that the arm had reached its limit of movement, nor was there an end option leading from the travel until stop state. This demonstrates how difficult it can be to correctly represent an interaction and how dependent it is on the skill of the particular analyst. However, even if the STN diagram had been correctly drawn, it is unlikely that this issue would have been identified without explicitly checking for illegal options.

The STN was not written at a level of detail which could identify issue 7 (gesture input with twice as many operations as voice because dependent on cursor movement). A more detailed diagram would have shown the gesture input as a series of states and transitions between them, and the voice input as one state with multiple transitions coming from it. This raises questions concerning the appropriate level of detail that an analysis requires. The STN would have identified this issue if it had been written at the appropriate level of detail.

Seventeen remaining issues were outside the direct scope of STN and are examined in detail. Issue 4 (lack of short cuts) was an issue that might have been identified through the craft skill of the analyst since the STN explicitly shows the possible path of the interaction through the various states. However, although this representation might lend itself to such issues being identified, it is the skill of the analyst that determines if the issue is identified as a cause for concern.

STN was unable to identify issues 11 (lack of feedback after Move Arm selected, no indication that whole arm is to be moved) and13 (user cannot check direction choice until arm starts to move) since the STN is concerned with the choices available for the user in

moving from one state to another, rather than in providing confirmation that a particular choice has been made.

The STN concentrates on the actual choice between the system states, rather than on the difficulties the user has in choosing between or user understanding of what a particular option choice means, and was therefore unable to identify issues 3 (difficulty of choosing between Move Arm or Move), 6 (Confusion over joint called Arm), 18 (not clear that End returns user to main menu), and 19 (End having two meanings).

Similarly, issues 12 (problems of determining left and right, especially when arm contorted), 14 (time taken to interact with system to stop arm), 17 (mismatch between way that arm works and way that user would move arm), and 23 (difficulty of judging arm movements) could not be identified since they relate to user expertise rather than states and transitions. STN is also unable to identify physical constraints on the user as shown in issues 21 (difficulty for user to move field of vision), and 22 (user looking one way, menu options in other direction).

Issues originally identified	2,5,15
Issues identified due to method	5,15
Issues identified due to source material	2
Issues that should have been identified by method	1
Issues not applicable to method	3,4,6,8,9,10,11,12,13,14,17,18,19,20,21,22,23
Level of abstraction issues	7
Other	16

**Table 19-1: Table showing nature of issues after STN re-analysis**

The STN diagram does not help with issues relating to misinterpretation by the system of a user's actions, as shown in issues 8 (problem if head moved to look at arm while gesture system operational may be interpreted as a command), 9 (if user pauses in middle of saying "Move arm"...), and 10 (if user engaged in conversation...).



Finally, the STN is not concerned with environmental issues or other contextual issues, and so was unable to identify issue 20 (lighting conditions).

### **19.3 CW Re-Analysis**

#### **1. Long sequence of operators to move arm**

This is something that would not be found strictly by the method, but by using craft skill on the material gathered.

#### **2. Inability to backtrack**

CW does not deal with error in terms of its implications, therefore would not find this issue, although it might come out from the craft skill of the analyst through thinking about rectifying errors.

#### **3. Difficulty of choosing between Move Arm or Move**

This is the kind of issue of misleading option labels that CW is designed to uncover, and did so.

#### **4. Lack of short cuts**

This is something that would not be found strictly by the method, but by using craft skill on the material gathered.

#### **5. Continue versus Go: Continue seen as redundant**

This would not be uncovered by the method, since it has no adverse effects on the use of the robotic arm. It might come out through craft skill.

#### **6. Confusion over joint called Arm**

This is the kind of issue of misleading option labels that CW is designed to uncover, and did so.

#### **7. Gesture input with twice as many operations as voice because dependent on cursor movement**

The CW does not examine the interface at this level of detail, and would not uncover this issue because it is not an issue that can be identified from the questions.

**8. Problem if head moved to look at arm while gesture system operational may be interpreted as a command**

This issue is one that could not be derived from the failure stories, and would depend on the skill of the analyst, therefore dependent upon craft skill for identification.

**9. If user pauses in middle of saying “Move arm”...**

This issue is one that could not be derived from the failure stories, and would depend on the skill of the analyst, therefore dependent upon craft skill for identification.

**10. If user engaged in conversation...**

This issue is one that could not be derived from the failure stories, and would depend on the skill of the analyst, therefore dependent upon craft skill for identification.

**11. Lack of feedback after Move Arm selected, no indication that whole arm is to be moved**

For the purposes of the original analysis, this was not relevant, since the feedback had not been implemented, but it was an important issue raised by the method that would have to be addressed once feedback had been implemented.

**12. Problems of determining left and right, especially when arm contorted**

This issue is one that could not be derived from the failure stories, and would depend on the skill of the analyst, therefore dependent upon craft skill for identification.

**13. User cannot check direction choice until arm starts to move**

This issue is one that could not be derived from the failure stories, and would depend on the skill of the analyst, therefore dependent upon craft skill for identification.

**14. Time taken to interact with system to stop arm**

This issue is one that could not be derived from the failure stories, and would depend on the skill of the analyst, therefore dependent upon craft skill for identification.

**15. Similarity between moving joint and moving whole arm**

This is not the kind of issue that CW looks for, so would not have been addressed. It might have come up through craft skill recognition of the similarities of the action sequences.



## **16. Illegal options**

CW would not have uncovered this issue since the task did not call for any of these illegal states to be explored. It would depend on the task as to whether this issue would be uncovered by CW.

## **17. Mismatch between way that arm works and way that user would move arm**

One of the aims of the method is to uncover this kind of issue, however there is not much support within the questions for this to be identified at a high level, because of the method's concentration on the step-by-step nature of the task. This is more likely to be uncovered by craft skill therefore.

## **18. Not clear that End returns user to main menu**

This is the kind of issue of misleading option labels that CW is designed to uncover, and did so.

## **19. End having two meanings**

This is the kind of issue of misleading option labels that CW is designed to uncover, and did so.

## **20. Lighting conditions**

This issue is one that could not be derived from the failure stories, and would depend on the skill of the analyst, therefore dependent upon craft skill for identification.

## **21. Difficulty for user to move field of vision**

This issue is one that could not be derived from the failure stories, and would depend on the skill of the analyst, therefore dependent upon craft skill for identification.

## **22. User looking one way, menu options in other direction**

This issue is one that could not be derived from the failure stories, and would depend on the skill of the analyst, therefore dependent upon craft skill for identification.

## **23. Difficulty of judging arm movements**

This issue is one that could not be derived from the failure stories, and would depend on the skill of the analyst, therefore dependent upon craft skill for identification.

## **19.4 Discussion of CW Re-analysis**

On re-examination of the twelve original issues, based on the description of CW given in (Wharton et al, 1994), five were identified strictly by the use of CW rather than by the craft skill of the analyst. Four of these issues (3,6,18,19) concerned misleading option names, and one (11) concerned the lack of feedback. CW explicitly supports the “label-following strategy” used by most users and progress towards user goals through option availability (Wharton et al, 1994).

The other seven issues originally identified were found to be beyond the boundaries of the method and were identified through craft skill. These related to the long sequence of operators needed to use the arm (1), the time taken to interact (14), the possible wrong interpretation of commands by the system (8,9,10), and the effects of an option choice in the particular context of use (12,13).

That these issues were not identified from the method is due both to the explicit limitations of CW, in that it only handles user exploration, and also the way in which the method is defined, with loose criteria, and failure stories defined as being the absence of a credible story rather than the interference of other issues. Although CW does not examine the long sequences of operators since it is task-based and examines each step one at a time, rather than as a whole, the method does require a list of the task sequence needed to successfully accomplish the task, from which issues related to the length of the interaction can be identified through the craft skill of the analyst. Craft skill can also be used to examine issues relating to the system correctly understanding what the user wishes to accomplish, although CW places its emphasis on the correct interpretation of system actions and options by the user. The failure stories can also be expanded by the analyst to include all conceivable problems rather than those concerning the lack of a credible success story. CW thus is able to support the inherent craft skill of the analyst through the representations and processes used, despite the limitations of the method in specifically identifying particular issues.

The method itself does not support all that it aims to do. For example, issue 17 concerns the mismatch between the way that the arm works and the way that a user would move the arm. One of the explicit aims of CW is to find “mismatches between users’ and designers’



conceptualization of a task” (Wharton et al, 1994). However, despite there being a distinct conceptual gap between how the user would move the arm and how the design of the robotic arm constrains movement, this issue cannot be identified by CW because of the method’s concentration on the step-by-step nature of the task rather than considering the task overall.

The ten issues remaining which were not identified by the original analysis were found to be beyond the boundaries of CW. Issue 2 (inability to backtrack) was not identified since CW can identify errors but does not examine how to recover from them. Issues 4 (lack of shortcuts),5 (continue seen as redundant), and 15 (similarity between moving joint and moving whole arm) were not identified because CW examines the success of the interaction rather than the efficiency of the interaction. Therefore the number of steps taken to correctly interact is less important than that the interaction was successful. CW was unable to identify issue 7 (gesture input with twice as many operations as voice) because the method does not examine the inputs to that depth and is concerned with user knowledge rather than physical actions. Issue 16 (illegal options) was outside the scope of the particular task sequence examined by CW. CW does not consider environmental issues such as those raised by issue 20 (lighting conditions), and is interested in user knowledge rather than in the physical constraints that might affect users as exemplified in issues 21 (difficulty for user to move field of vision) and 22 (user looking one way, menu options in other direction). Lastly, with regards to issue 23 (difficulty of judging arm movements), CW investigates user knowledge rather than user expertise.

Issues originally identified	1,3,6,8,9,10,11,12,13,14,18,19
Issues identified due to method	3,6,11,18,19
Issues identified due to craft skill	1,8,9,10,12,13,14
Issues not applicable to method	2,4,5,7,15,16,20,21,22,23
Other	17

**Table 19-2: Table showing nature of issues after CW re-analysis**

## **19.5 CMN and CPM GOMS Re-Analyses**

The CPM GOMS analysis was unable to identify any issues other than the difference between the use of voice and gesture operators. Thus only issue seven was able to be identified, and this was the only issue that can be considered to be within the bounds of the method. This re-analysis consequently focused on the use of CMN GOMS.

### **1. Long sequence of operators to move arm**

By writing out the methods, even though the operators were not examined, the long sequence showed that this would take a long time. This is something that the CMN GOMS should explicitly identify, and did.

### **2. inability to backtrack**

This was an issue that was identified using craft skill from the CMN GOMS sequence of goals and methods.

### **3. difficulty of choosing between Move Arm or Move**

CMN GOMS does not support the identification of issues relating to problems choosing between commands. This is therefore an issue identified by craft skill.

### **4. Lack of short cuts**

By writing out the methods, the long sequence showed that this would take a long time and that there were no short cuts. This is something that the CMN GOMS should identify, and did.

### **5. Continue versus Go: Continue seen as redundant**

The use of CMN GOMS in writing the goals and the sequence of methods and operators allowed this issue to become apparent. Therefore this issue was identified by the method.

### **6. Confusion over joint called Arm**

CMN GOMS does not support issues relating to the correct identification of options. This issue was therefore identified through craft skill.



**7. Gesture input with twice as many operations as voice because dependent on cursor movement**

CMN-GOMS was able to identify the individual operations. This issue was therefore identified through the use of the method.

CPM GOMS was also able to identify this issue, which was within the scope of the technique.

**8. problem if head moved to look at arm while gesture system operational may be interpreted as a command**

This issue is outside the scope of CMN GOMS and was not identified.

**9. if user pauses in middle of saying "Move arm"...**

This issue is outside the scope of CMN GOMS and was not identified.

**10. if user engaged in conversation...**

This issue is outside the scope of CMN GOMS and was not identified.

**11. lack of feedback after Move Arm selected, no indication that whole arm is to be moved**

At the time of writing the CMN GOMS analysis the feedback was not implemented. However, CMN GOMS would not identify this issue since it assumes correct user action.

**12. problems of determining left and right, especially when arm contorted**

This issue is outside the scope of CMN GOMS and was not identified.

**13. user cannot check direction choice until arm starts to move**

This issue is outside the scope of CMN GOMS and was not identified.

**14. time taken to interact with system to stop arm**

This issue was clearly identified by CMN GOMS which identified the number of operators for both voice and gesture, and found that gesture had twice as many as voice for this.

### **15. similarity between moving joint and moving whole arm**

The use of CMN GOMS in writing the goals and the sequence of methods and operators allowed this issue to become apparent. Therefore this issue was identified by the method.

### **16. illegal options**

CMN GOMS was unable to identify this issue because it is a procedural and task-based technique, and the task as given did not explore those states.

### **17. mismatch between way that arm works and way that user would move arm**

This issue is outside the scope of CMN GOMS and was not identified.

### **18. Not clear that End returns user to main menu**

This issue is outside the scope of CMN GOMS and was not identified.

### **19. End having two meanings**

This issue is outside the scope of CMN GOMS and was not identified.

### **20. Lighting conditions**

This issue is outside the scope of CMN GOMS and was not identified.

### **21. difficulty for user to move field of vision**

This issue is outside the scope of CMN GOMS and was not identified.

### **22. user looking one way, menu options in other direction**

This issue is outside the scope of CMN GOMS and was not identified.

### **23. difficulty of judging arm movements**

This issue is outside the scope of CMN GOMS and was not identified.

## **19.6 Discussion of CMN and CPM GOMS Re-analysis**

The re-analysis of the CMN GOMS and CPM GOMS was based on (John & Kieras, 1994). The original CPM GOMS analysis was unable to identify any issues other than the difference between the use of voice and gesture operators (issue 7: gesture input with twice as many operations as voice because dependent on cursor movement). On re-examination



this issue was found to be within the bounds of the method, since CPM GOMS supports the identification of the critical path of interaction.

The re-analysis consequently focused on the use of CMN GOMS. Of the eight issues originally identified, five of these (1,4,5,7,15) were found to be within the strict scope of the method. One issue (7) concerned the number of operators, two issues (1,4) concerned the sequence of operators and lack of short cuts, and two other issues (5,15) concerned the similarity of options. CMN supports the identification of operators and their sequences.

The other three issues originally identified were found to be beyond the boundaries of CMN GOMS and were identified by craft skill. These related to the inability to backtrack (2), and confusion regarding option names (3,6). The inability to backtrack was not identified by CMN GOMS because although it considers the task sequence in some detail, it only considers limited error recovery. Issues 3 and 6 were not identified by CMN GOMS because the method does not support the identification of issues relating to problems of choosing between options.

One issue which should have been identified by CMN GOMS and was not was issue 14 (time taken to interact with system to stop arm). Although CMN GOMS identified the number of operators for both voice and gesture, and found that gesture had twice as many as voice, the implications for this with regards to the stopping of the arm were not considered.

The remaining fourteen issues were found to be outside the scope of CMN GOMS, and are examined in detail.

CMN GOMS was unable to identify issue 16 (illegal options) because it is a procedural and task-based technique, and the task as given did not explore those states.

CMN GOMS does not help with issues relating to misinterpretation by the system of a user's actions, as shown in issues 8 (problem if head moved to look at arm while gesture system operational may be interpreted as a command), 9 (if user pauses in middle of saying "Move arm"...), and 10 (if user engaged in conversation...).

CMN GOMS was unable to identify issues 11 (lack of feedback after Move Arm selected, no indication that whole arm is to be moved) and 13 (user cannot check direction choice

until arm starts to move) since the feedback was not implemented at that time. Even if the feedback had been implemented, CMN GOMS is concerned with the sequence of operations for moving through the task rather than in checking that the user has the necessary information to do so. Therefore CMN GOMS was also unable to identify issues relating to the difficulties the user has in choosing between or user understanding of what a particular option choice means (issues 18: not clear that End returns user to main menu; and 19: End having two meanings). Similarly, issues 12 (problems of determining left and right, especially when arm contorted), 17 (mismatch between way that arm works and way that user would move arm), and 23 (difficulty of judging arm movements) could not be identified. CMN GOMS is also unable to identify physical constraints on the user as shown in issues 21 (difficulty for user to move field of vision), and 22 (user looking one way, menu options in other direction).

Finally, CMN GOMS is not concerned with environmental issues or other contextual issues, and so was unable to identify issue 20 (lighting conditions).

Issues originally identified	1,2,3,4,5,6,7,15
Issues identified due to method	1,4,5,7,15
Issues identified due to craft skill	2,3,6
Issues that should have been identified by method	14
Issues not applicable to method	8,9,10,11,12,13,16,17,18,19,20,21,22,23

**Table 19-3: Table showing nature of issues after CMN GOMS re-analysis**

## **19.7 PUM Re-Analysis**

### **1. long sequence of operators to move arm**

This issue was mentioned in the original analysis, but not in a strong enough way for it to be apparent as an issue of consequence. It was identified from looking at the heavy ordering identified by the analysis, and was therefore dependent upon the craft skill of the analyst.



## **2. inability to backtrack**

The original analysis found a heavy ordering, which is within the bounds of the PUM technique. However, from this was derived the lack of backtracking provision, which is therefore identified by the craft skill of the analyst, based on the representation provided by the method.

## **3. difficulty of choosing between Move Arm or Move**

This was identified by the original analysis from modelling the user knowledge required.

## **4. Lack of short cuts**

This is not an issue that PUM would identify, however it is one that might be expected to be identified through craft skill, since it is closely connected to the heavy ordering of the task, and the lack of mapping between the device and the domain.

## **5. Continue versus Go: Continue seen as redundant**

This is not an issue that PUM would identify, since it has little relevance to user knowledge. It is unlikely that it would have been identified through craft skill during the analysis due to the way the analysis was conducted.

## **6. Confusion over joint called Arm**

This was identified by the original analysis from modelling the user knowledge required.

## **7. Gesture input with twice as many operations as voice because dependent on cursor movement**

This did not come out in the original PUM analysis, because the analysis was not written at a low enough level of abstraction for this to be apparent. If the precondition of the cursor under the correct option had been written then the differing number of preconditions for the two input devices should have been recognised, and this issue identified. This is therefore a matter concerning the appropriate level of abstraction of analysis.

**8. problem if head moved to look at arm while gesture system operational may be interpreted as a command**

PUM does not consider the misinterpretation of user actions by the system, since it concentrates more on user interpretation of the system. Therefore, PUM was unable to identify this issue.

**9. if user pauses in middle of saying "Move arm"...**

PUM does not consider the misinterpretation of user actions by the system, since it concentrates more on user interpretation of the system. Therefore, PUM was unable to identify this issue.

**10. if user engaged in conversation...**

PUM does not consider the misinterpretation of user actions by the system, since it concentrates more on user interpretation of the system. Therefore, PUM was unable to identify this issue.

**11. lack of feedback after Move Arm selected, no indication that whole arm is to be moved**

The output was not included in the original analysis. If it had been then the PUM analysis might have picked up on this issue, in the modelling of the user knowledge, because the user would not know that the option had been selected. As it was, PUM could not identify this issue because of the boundaries of the original analysis.

**12. problems of determining left and right, especially when arm contorted**

PUM assumes that the user has certain knowledge, and is therefore unlikely to make this issue explicit from the analysis. However, the craft-skill of the analyst might identify this issue.

**13. user cannot check direction choice until arm starts to move**

The output was not included in the original analysis. If it had been then the PUM analysis might have identified this issue through the modelling of the user knowledge, because the user would not know that the option had been selected.



#### **14. time taken to interact with system to stop arm**

PUM only considers mental operations rather than physical constraints, so would be unable to identify this issue.

#### **15. similarity between moving joint and moving whole arm**

The way that the PUM analysis was conducted meant that this issue was not identified, although it would probably have been recognised if the analyst was looking for it. Therefore, although the PUM analysis represented the operations, it would take the craft skill of the analyst to identify their similarity.

#### **16. illegal options**

The PUM analysis did not find this issue since it was concerned with the user performing a correct task, rather than all possible states and choices.

#### **17. mismatch between way that arm works and way that user would move arm**

This was an issue that was identified by the PUM analysis, and is the kind of issue that the technique should identify, because of PUMs consideration of the device and domain issues.

#### **18. Not clear that End returns user to main menu**

This was identified by the original analysis from modelling the user knowledge required.

#### **19. End having two meanings**

This was identified by the original analysis from modelling the user knowledge required.

#### **20. Lighting conditions**

PUM can not identify issues related to the context of the task rather than the task itself.

#### **21. difficulty for user to move field of vision**

PUM does not consider user physical constraints, and therefore would be unable to identify this issue.

#### **22. user looking one way, menu options in other direction**

PUM does not consider user physical constraints, and therefore would be unable to identify this issue.

### **23. difficulty of judging arm movements**

PUM does not consider tactical goal choice, and therefore would be unlikely to identify this issue.

## **19.8 Discussion of PUM Re-analysis**

The original issues were re-analysed, based on (Blandford et al, 1998b). Of the six issues previously identified by the PUM analysis, five were identified through the use of the method as opposed to craft skill. Four issues (3, 6, 18, 19) concerned misleading option names, and one (17) concerned the mismatch between the way that the robotic arm works, and how a user would move an arm. PUM supports the identification of knowledge needed by a user, and the difference between user and device tasks.

The remaining issue (2) was found to be beyond PUM's scope, in that it related to the lack of backtracking opportunities. It was derived from the representation of the interaction, with its emphasis on the heavy ordering of operations.

Issue 1 (long sequence of operators to move arm) was mentioned in passing in the original analysis, but not strongly enough for it to be identified as a usability issue and included in the table of issues. However, it was derived in a similar manner to issue 2 discussed above, and thus falls outside the scope of a PUM analysis.

Issue 7 (gesture input with twice as many operations as voice because dependent on cursor movement) was not identified in the original analysis since it was not written at a low enough level of detail. If the precondition of the cursor under the correct option had been written then the differing number of preconditions for the two input devices would have been recognised and this issue identified. This highlights the problems involved in ensuring that an analysis is written at the correct level of detail.

Two issues: 11 (lack of feedback after Move Arm selected) and 13 (user cannot check direction choice until arm starts to move); were not identified in the original analysis. Although there was no output implemented for the robotic arm at this point, the method should still have identified these issues since PUM traces what information users need and where it comes from. That these issues were not identified suggests that either the analysis



was done at the wrong level of detail, or that the preconceptions of the analyst, namely that since the feedback was not implemented those kind of issues were not to be identified, interfered with the analysis.

Issue 16 (illegal options) was not identified since the task did not involve those options being chosen. These issue would have been identified if the task had included the illegal options.

The remaining issues were considered to be beyond the scope of a PUM analysis, and are considered in more detail.

Issue 4 (lack of short cuts) is not an issue that PUM could identify, however it is one that might be expected to be identified through the craft skill of the analyst, since it is closely connected to the heavy ordering of the task, and the lack of mapping between the device and the domain.

Issue 5 (Continue versus Go: Continue seen as redundant) and 15 (similarity between moving joint and moving whole arm) would not have been recognised due to the way in which the PUM analysis was conducted and represented, although they would probably have been recognised if the analyst was looking for them.

PUM does not consider the misinterpretation of user actions by the system, since it concentrates more on user interpretation of the system. Therefore, PUM was unable to identify issues 8 (problem if head moved to look at arm while gesture system operational may be interpreted as a command), 9 ( if user pauses in middle of saying "Move arm"...), and 10 (if user engaged in conversation...).

Two issues relating to user judgement were not identified by PUM: 12 (problems of determining left and right, especially when arm contorted) and 23 (difficulty of judging arm movements). PUM assumes that the user has the capacity for judgement, and does not consider tactical goal choice, and therefore would not be able to identify these issues.

PUM is concerned with mental operations rather than physical constraints or contextual issues, so would be unable to identify issue 14 (time taken to interact with system to stop arm), 21 (difficulty for user to move field of vision), 22 (user looking one way, menu options in other direction), and issue 20 (lighting conditions).

Issues originally identified	2,3,6,17,18,19
Issues identified due to method	3,6,17,18,19
Issues identified due to craft skill	2
Issues that should have been identified by method	11,13
Issues not applicable to method	4,5,8,9,10,12,14,15,16,20,21,22,23
Level of abstraction issues	7
Other	1

**Table 19-4: Table showing nature of issues after PUM re-analysis**

## **19.9 Z re-analysis**

### **1. long sequence of operators to move arm**

This issue was not apparent because of the way that the specification was constructed, although the specification did represent it. This issue therefore highlights the important difference between an issue being represented and identified. It would take a certain amount of craft skill on the part of the analyst to identify this issue.

### **2. inability to backtrack**

This was an issue identified by the Z analysis, and one that would be expected to be identified due to the strongly ordered nature of Z specifications.

### **3. difficulty of choosing between Move Arm or Move**

The Z specification was concerned more with representing the choice than with how the user would make that choice, so this is not an issue that the Z analysis would be expected to uncover. It would take a certain amount of craft skill for this issue to be identified.

### **4. Lack of short cuts**

The Z specification represented the lack of backtracking opportunities, due to its concentration on the ordering of the interaction. The lack of short-cuts was therefore also represented. However, the issue, although represented, was not identified, which again



illustrates the difference between an issue being represented and identified, and the importance of the craft skill of the analyst in identifying significant issues.

#### **5. Continue versus Go: Continue seen as redundant**

This issue was identified by the Z specification since both options share the same functionality, and were represented by different schemas with identical contents.

#### **6. Confusion over joint called Arm**

This issue should have been identified when the type ARMPART was declared, since the whole arm had to be called “wholearm” rather than “arm”, due to there already being a joint called arm. Therefore, although this issue was represented by the specification it was not identified by the analyst, which indicates the amount of craft skill necessary to identify important issues.

#### **7. Gesture input with twice as many operations as voice because dependent on cursor movement**

This issue was not identified by the Z specification because the specification was not written at a low enough level of detail to represent the cursor movement. This illustrates the need for the appropriate level of abstraction of the representation.

#### **8. problem if head moved to look at arm while gesture system operational may be interpreted as a command**

The Z specification was concerned with the possible states of the robotic arm and movement between them rather than the specifics of how the user interacted with the robotic arm, and therefore was unable to identify this issue.

#### **9. if user pauses in middle of saying “Move arm”...**

The Z specification was concerned with the possible states of the robotic arm and movement between them rather than the specifics of how the user interacted with the robotic arm, and therefore was unable to identify this issue.

**10. if user engaged in conversation...**

The Z specification was concerned with the possible states of the robotic arm and movement between them rather than the specifics of how the user interacted with the robotic arm, and therefore was unable to identify this issue.

**11. lack of feedback after Move Arm selected, no indication that whole arm is to be moved**

This Z specification did not cover the interface output other than the motion of the robotic arm, since this was not implemented at the time of the analysis, and therefore did not identify this issue.

**12. problems of determining left and right, especially when arm contorted**

The Z specification was concerned with the possible states of the robotic arm and movement between them rather than the specifics of how the user interacted with the robotic arm, and therefore was unable to identify this issue.

**13. user cannot check direction choice until arm starts to move**

This Z specification did not cover the interface output other than the motion of the robotic arm, since this was not implemented at the time of the analysis, and therefore did not identify this issue.

**14. time taken to interact with system to stop arm**

The Z specification was concerned with the possible states of the robotic arm and movement between them rather than the specifics of how the user would interact with the robotic arm, and therefore did not identify this issue.

**15. similarity between moving joint and moving whole arm**

This issue was identified by the Z specification because of the similar functionality, and represented in almost exactly the same manner, except that the choice of directions available changed.



## **16. illegal options**

This issue was identified since the Z specification involves examining in detail each operation and what is allowable for any given state of the interface. The interface offers options to the user that cannot in fact be carried out. When the arm reaches its limit of movement, the interface displays the options **Continue**, **Speed Level**, **End**. However, **Continue** and **Speed Level** are irrelevant: **Continue** because the arm cannot move any further; and **Speed Level** because any speed selected now will not be kept when **End** is chosen and the interface returns to the initial menu.

## **17. mismatch between way that arm works and way that user would move arm**

The Z specification concentrates on representing the states of the robotic arm, and is less concerned with problems that the user has in interacting with the robotic arm. Therefore the specification would not identify this issue as such. However, by writing the specification, the difference between the way that the robotic arm works and how a user would move their arm would be apparent. Identifying this issue, although represented, would take a certain amount of craft skill.

## **18. Not clear that End returns user to main menu**

The Z specification was concerned with the states of the arm, rather than the user interpretation of available options, and therefore was not able to identify this issue.

## **19. End having two meanings**

The Z specification was concerned with the states of the arm, rather than the user interpretation of available options, and therefore was not able to identify this issue.

## **20. Lighting conditions**

The Z specification was concerned only with the states of the robotic arm and movement between those states, and therefore could not identify this issue.

## **21. difficulty for user to move field of vision**

The Z specification was concerned with the possible states of the robotic arm and movement between them rather than the specifics of how the user interacted with the robotic arm, and therefore was unable to identify this issue.

## **22. user looking one way, menu options in other direction**

The Z specification was concerned with the possible states of the robotic arm and movement between them rather than the specifics of how the user interacted with the robotic arm, and therefore was unable to identify this issue.

## **23. difficulty of judging arm movements**

The Z specification was concerned with the possible states of the robotic arm and movement between them rather than the specifics of how the user interacted with the robotic arm, and therefore was unable to identify this issue.

## **19.10 Discussion of Z Re-analysis**

On re-examination of the Z analysis, based on the description of Z in (Spivey, 1989), three of the four issues originally identified (issues 5,15,16) were not identified due to the craft skill of the analyst.

Two of these issues (issue 5: Continue versus Go: Continue seen as redundant; and issue 15: similarity between moving joint and moving whole arm) related to similarity of operations, and were identified by the schemas having identical or nearly identical contents. The remaining issue 16 (illegal options) was identified since the Z specification involves examining in detail each operation and what is allowable for any given state of the interface. When the arm reaches its limit of movement, the interface offers options to the user that cannot be carried out, displaying the options **Continue**, **Speed Level**, **End**. However, **Continue** and **Speed Level** are irrelevant: **Continue** because the arm cannot move any further; and **Speed Level** because any speed selected now will not be kept when **End** is chosen and the interface returns to the initial menu.

Issue 2 (inability to backtrack) was an issue identified by craft skill from the Z specification due to the strongly ordered nature of the specification.

Issue 6 (confusion over joint called Arm) should have been identified when the type ARMPART was declared, since the whole arm had to be called “wholearm” rather than “arm”, because there was already a joint called “arm”.



Issue 7 (gesture input with twice as many operations as voice because dependent on cursor movement) was not identified by the Z analysis because the specification was not written at a low enough level of detail to represent the cursor movement, which illustrates the need for the appropriate level of abstraction in a representation.

The remaining issues were outside the scope of the Z analysis, and are examined in detail.

Although the Z specification concentrates on ordering issues, it did not represent the long sequence of operators to move arm (issue 1) or the lack of short cuts (issue 4) and was therefore unable to identify these issues.

The Z specification did not cover the interface output other than the motion of the robotic arm, since this was not implemented at the time of the analysis, and therefore did not identify issues 11 (lack of feedback after Move Arm selected, no indication that whole arm is to be moved) and 13 (user cannot check direction choice until arm starts to move).

The Z specification is concerned with the states of the arm, rather than the user interpretation of available options, and therefore was not able to identify issues 3 (difficulty of choosing between Move Arm or Move), 18 (Not clear that End returns user to main menu) and 19 (End having two meanings).

Z is unable to identify issues relating to user expertise, such as issues 12 (problems of determining left and right, especially when arm contorted), 14 (time taken to interact with system to stop arm), 17 (mismatch between way that arm works and way that user would move arm), and 23 (difficulty of judging arm movements). Z is also unable to identify physical constraints relating to the user as exemplified by issues 21 (difficulty for user to move field of vision), and 22 (user looking one way, menu options in other direction).

The Z specification does not help with issues relating to misinterpretation by the system of a user's actions, as shown in issues 8 (problem if head moved to look at arm while gesture system operational may be interpreted as a command), 9 (if user pauses in middle of saying "Move arm"...), and 10 (if user engaged in conversation...), nor is it able to identify environmental or other contextual issues such as issue 20 (lighting conditions).

Issues originally identified	2,5,15,16
Issues identified due to method	5,15,16
Issues identified due to craft skill	2
Issues that should have been identified by method	6
Issues not applicable to method	1,3,4,8,9,10,11,12,13,14,17,18,19,20,21,22,23
Level of abstraction issues	7

**Table 19-5: Table showing nature of issues after Z re-analysis**

### **19.11 EMU re-analysis**

#### **1. long sequence of operators to move arm**

This issue was identified from the interaction sequence using craft skill.

#### **2. inability to backtrack**

EMU does not look at the implications of error and therefore cannot identify this kind of issue. The representation of the interaction does not allow for the identification of this issue.

#### **3. difficulty of choosing between Move Arm or Move**

This is a potential semantic clash, and is the kind of issue that EMU can identify, and did.

#### **4. Lack of short cuts**

The notation used by EMU does not allow for the identification of this issue. It is not an issue that would be identified by either the method, or craft skill based on the materials provided as a result of following the method.

#### **5. Continue versus Go: Continue seen as redundant**

This particular task did not involve the use of the continue option, and was therefore not identified as an issue.



## **6. Confusion over joint called Arm**

This is a potential semantic clash, and is the kind of issue that EMU can identify, and did.

## **7. Gesture input with twice as many operations as voice because dependent on cursor movement**

The interaction sequence of the modalities represented the cursor, but the extra modalities relating to the cursor were not identified as significant. This kind of issue could not be identified by EMU.

## **8. problem if head moved to look at arm while gesture system operational may be interpreted as a command**

The comparison of the system profile with the user modalities raised this issue. The method therefore supports the identification of this kind of issue.

## **9. if user pauses in middle of saying "Move arm"...**

This was not an issue identified by the method because EMU does not consider timings and the start and finish of modalities.

## **10. if user engaged in conversation...**

This was not an issue identified by the method since it was not indicated in the initial scenario that the user would be engaged in conversation. If that information had been included in the environment profile, then this issue would have been identified by EMU.

## **11. lack of feedback after Move Arm selected, no indication that whole arm is to be moved**

Feedback was not considered due to it not being implemented.

## **12. problems of determining left and right, especially when arm contorted**

EMU is not able to identify issues of this nature, except possibly in terms of clash unless expert, which would take a large amount of craft skill to identify.

## **13. user cannot check direction choice until arm starts to move**

This issue relates to the feedback available. Since the feedback was not implemented at the time of this analysis, this issue was not identified.

#### **14. time taken to interact with system to stop arm**

The interaction sequence was able to represent the modalities used for this command in detail, and was therefore able to show that both the gesture and voice commands would only use the same number of modalities. Although EMU does not represent time as such, it is able to examine sequences of modalities, whose timing properties can then be investigated. This issue was identified through craft skill using the representation from the EMU analysis.

#### **15. similarity between moving joint and moving whole arm**

This issue was not identified by EMU since EMU does not require this to be examined.

#### **16. illegal options**

This was not an issue which EMU could identify since the analysis was task-based, and the task as given did not explore those options.

#### **17. mismatch between way that arm works and way that user would move arm**

Stage one of EMU concentrates attention on the task, so at this point this kind of issue would be expected to be noted, through the craft skill of the analyst.

#### **18. Not clear that End returns user to main menu**

This issue was not identified in the original analysis and should have been, since it is a potential mismatch. This demonstrates how the identification of any issue is dependent upon the analyst, and that mistakes and omissions can occur.

#### **19. End having two meanings**

This issue was not identified in the original analysis and should have been, since it is a potential mismatch. This demonstrates how the identification of any issue is dependent upon the analyst, and that mistakes and omissions can occur.

#### **20. Lighting conditions**

EMU calls for the explicit examination of the environment, and comparison with the modalities used. Therefore, the identification of this issue is supported by the method.



### **21. difficulty for user to move field of vision**

The method in stage four compared the user profile with the system modalities and was able to identify this issue.

### **22. user looking one way, menu options in other direction**

This is the kind of issue that is explicitly found by the method with regards to the field of vision of the user and the resulting potential physical clash.

### **23. difficulty of judging arm movements**

This is a clash unless expert issue, and the method instructs the analyst to look for these clashes.

## **19.12 Discussion of EMU Re-analysis**

The EMU analysis was re-examined based on the description of the methodology given to the students (see Appendix D). Of the ten issues originally identified, seven were identified through EMU rather than through the craft skill of the analyst. Four of these issues concerned clashes, with issues 3 and 6 potential semantic clashes, issue 22 a potential physical clash, and issue 23 a potential clash unless expert. Issues 8, 20 and 21 were identified through the comparison of user, system and environmental profiles with the modality listings. EMU explicitly supports the identification of clashes, and the identification of issues arising from incompatibilities of profiles and listings.

The three remaining issues were found to be beyond the boundaries of the method and were identified through the craft skill of the analyst. These related to the long sequence of operators needed to use the arm (1), the time taken to interact with the system to stop the arm (14), and the mismatch between the way that the arm works and how the user might use the arm (17). EMU does not specifically examine the length of the interaction, although the interaction sequence produced through following the method is amenable to this kind of analysis. This interaction sequence can also be used to examine issues such as 14. The interaction sequence was able to represent the modalities used for this command in detail, and was therefore able to show that both the gesture and voice commands would use the same number of modalities. Although EMU does not represent time as such, it is able to

examine sequences of modalities, whose timing properties can then be investigated. With regards to the mismatch between the arm and the user, stage one of the EMU methodology concentrates attention on the task selected, which would encourage this kind of issue to be identified by the craft skill of the analyst.

Issues 18 (not clear that End returns user to main menu) and 19 (End having two meanings), both potential mismatches, should have been identified by EMU but were omitted. This demonstrates how the identification of any issue is dependent upon the analyst, and that mistakes and omissions can occur.

There were eleven issues which were not applicable to the analysis using EMU. These are examined in more detail.

Issues 2 and 4 relate to how EMU does not identify errors or long-winded procedures specifically, and how the representation, although it allows for the length of the interaction to be observed, does not support the identification of issues relating to error recovery. Issue 9 (if user pauses in middle of saying "Move arm"... ) was not identified by the method because EMU does not consider timings and the start and finish of modalities.

Issues 5 and 15 relate to the similarity of particular operations. EMU does not specifically look for similarities in methods between operations, although this might be identified through the craft skill of an analyst from the interaction modality listing. Issue 7 (gesture input with twice as many operations as voice because dependent on cursor movement) was not identified by EMU since although the interaction sequence of the modalities showed the movement of the cursor, these extra modalities were not identified as significant. This would have taken craft skill to identify, and is therefore outside the scope of EMU. The method was unable to identify issue 12 (problems of determining left and right, especially when arm contorted) except possibly in terms of a clash unless expert, which would take a large amount of craft skill to identify.

At the time of analysis the feedback was not implemented, so EMU was unable to identify issues 11 (lack of feedback after Move Arm selected, no indication that whole arm is to be moved) and 13 (user cannot check direction choice until arm starts to move). Since EMU is a task-based methodology it was unable to identify issue 16 (illegal options) as the task did not involve their use. Issue 10 (if user engaged in conversation...) was not identified by



EMU since it was not indicated in the initial scenario that the user would be engaged in conversation. If that information had been included in the environment profile, then this issue would have been identified by the methodology.

Issues originally identified	1,3,6,8,14,17,20,21,22,23
Issues identified due to method	3,6,8,20,21,22,23
Issues identified due to craft skill	1,14,17
Issues that should have been identified by method	18,19
Issues not applicable to method	2,4,5,7,9,10,11,12,13,15,16

**Table 19-6: Table showing nature of issues after EMU re-analysis**

# **20. APPENDIX J: A COMPARISON OF FIVE TECHNIQUES FOR INVESTIGATING GENERAL AND MULTI-MODAL USABILITY ISSUES**

J.K.Hyde  
A.E.Blandford  
H.S.Goodman

## **1. ABSTRACT**

Five different modelling techniques are used to describe the interface of a multi-modal robotic arm. The techniques are examined in order to determine both what they can and cannot reason about, and what they can explicitly represent in terms of the multi-modal usability aspects of the interface. Analysis of the techniques show that a wide selection of usability issues were uncovered, that craft skill is a strong component of usability assessment, and that techniques not normally associated with usability evaluation can provide relevant insights. However, all the techniques had difficulties in dealing with the multi-modal aspects of the arm, emphasising the lack of appropriate techniques in this field.

## **2. INTRODUCTION**

This report examines five different ways of representing user interaction with a robotic arm, and shows how existing techniques, both user and system oriented, can be applied to an interface under development in order to identify general and multi-modal usability problems. There is a lack of techniques that explicitly deal with multi-modal issues, and one of the aims of this work was to analyse the capability of some existing techniques in dealing with this area. The five approaches cover a wide spectrum of representation techniques, and include a hierarchical goal-based one (GOMS), a natural language goal-based method (Cognitive Walkthrough), a means-end planning based one (Programmable User Model), a diagrammatic representation (State Transition Diagram), and one based on set theory and first-order predicate logic. The techniques were chosen in order to show a wide range of formal system and user approaches. The GOMS and Cognitive Walkthrough (CW) are highly user based, the Programmable User Model (PUM) strives to analyse both the user domain tasks and the device tasks, and the Z and State Transition Network (STN) representations are more system-oriented. Z and STN are often used in software engineering to describe the specification and functionality of a system. It was therefore decided to see what leverage such well-known techniques could give in the area of usability evaluation. If Z and STN were shown to have significant utility, it would mean that usability issues could be discussed from specifications and diagrams without



the need for special usability studies. It would also potentially involve less training, since both are commonly used in system design.

The differences in emphasis between the techniques means that different areas of concern are raised, but all show that worthwhile issues can be raised before the interface reaches an implementation stage, and without the need for empirical user testing. This is not to say that empirical testing is not beneficial, but that design problems can be identified earlier in the development process when the appropriate methods are used. The purpose of this report is to determine the scope of the five techniques with regards to general usability issues, and to demonstrate how well they are able to identify potential multi-modal usability issues.

The development of computers has been followed by a recognition of the need to ensure that they are usable by the people for whom they are designed. Effective interface design is seen as the key to effective use of computer resources, and therefore applicable ways of ensuring efficient interfaces is very important. One way in which attempts have been made to improve system usability has been through the use of user modelling techniques. These involve making models of user behaviour based on established cognitive principles, and applying them in a structured way to the analysis of the interface. The resulting "engineering models" of human performance [B.E.John, 1996] allow user actions to be predicted and the interface changed without the need for extensive empirical testing or prototyping. In this way, not only can changes occur earlier in the design process where they can be more easily introduced, they also save on the associated costs. User models also allow different levels of abstraction (granularity), allowing for different levels of design fixing, and are ideally designed to be used by software engineers rather than HCI specialists.

The past twenty years have seen such modelling techniques develop in sophistication along with the types of interfaces now available. However, despite their advantages, there are significant problems associated with their use. One problem with the use of user-modelling is that of scale. Models can typically only represent certain aspects of user behaviour if the model is to be small and tractable enough to be of practical use to the designer. Thus only certain actions or usability situations may be perceived through using such a model. Such aspects may not be as important as other aspects which are not brought out by user modelling, and which may be better examined using empirical methods. An associated problem is that user modelling is often able to deal with small-scale scenarios, but unable to deal with more complex interfaces, such as graphical user interfaces, and novel methods of interaction, such as multi-modal interaction.

Another drawback is that user models tend to be better at evaluating a given design, rather than putting forward suggestions for design. They can analyse where, and provide an explanation as to why, problems are likely to occur, but they are often unable to provide alternatives or to avoid the problem in the first place [Barnard and Harrison, 1989]. In this way they are seen as reactive rather than proactive.



Many techniques are difficult and time consuming to apply, and therefore are seen as not worth the effort, since the information can often be too general to usefully guide the design process, which can lead to empirical testing being considered as offering better value. Unfortunately, empirical testing is expensive and often too late for useful changes to be made to an interface. When modelling methods are established, formalisation is often a barrier to straightforward application [Blandford and Young, 1996] and can lead to craft skill becoming more important. Also, the stage at which formalisation is used is important, since early on in the design process the emphasis is often on rapid prototyping rather than in-depth usability analysis. At this stage a very rigorous formalisation may be inappropriate [Blandford and Young, 1996].

This first section of the report has examined the role of and justification for user modelling techniques as opposed to other methods of representation and evaluation. The robotic arm and its interface is described, and the report goes on to scope each of the techniques in turn, by applying it to the interface. The conclusions analyse how well these five approaches were able to identify and justify general usability issues arising from the examination of the interface, and to what extent they were able to uncover potential multi-modal usability issues.

### **3. THE FIVE TECHNIQUES**

#### **3.1 GOMS**

GOMS is a cognitive modelling technique developed by Card, Moran and Newell in the 1970s, and is based on the idea of the human as an information processor. It is one of the earliest cognitive task analysis theories, and was ground-breaking in its day. Since then it has evolved into a method of considerable depth and sophistication, and is probably the oldest and most widely known cognitive method presently used.

GOMS stands for Goals, Operators, Methods, and Selection and is based on the premise that a user achieves goals by breaking them down into sub-goals which can then be separately achieved. The user goals are identified at different levels of granularity. The Operators are the ways available to accomplish the goals. Methods are defined sequences of operators and goals, and Selection rules determine how to choose between more than one method [John and Kieras, 1994]. Tasks are seen as something that people do to achieve goals. The emphasis is not just on the purely physical aspects of the actions, but also on the mental process, for example, what they have to know or remember.

The technique's main strength is that it is very useful at predicting user times for routine and expertly done tasks, where there are likely to be few errors. Indeed, the technique explicitly does not deal with error, in that it assumes expert and efficient behaviour, which is both a strength and weakness in its applicability. Since there is more than one way of applying the GOMS model, it can be very versatile and useful. Variations include goal hierarchies, working memory loads, schedule tasks, lists of operators, and production systems. The variations of GOMS can be used to model different usability areas such as



auditory perception, verbal responses and eye movements [Gray et al, 1993]. The different variations result in different emphasises, and provide different performance measures, and can be used to model parallel activities and different levels of granularity. This report uses two particular versions: CMN-GOMS, which is based on the Keystroke Level Model developed by Card, Moran and Newell, and CPM-GOMS, which attempts to trace the cognitive operators being used at any time. Both these versions are useful for different aspects of the interface, thus which variation to use depends on what information is required by the study.

However, despite GOMS pedigree and history, it is still not a complete answer to cognitive modelling, and has various drawbacks. The method depends on the interaction being very structured, having easily identifiable goals and methods, with clear selection rules, and there is no guarantee that all user goals will be identified. It is best at dealing with rational, expert and error-free behaviour, so is less useful at dealing with novice situations. In design it tends to be reactive, in that it assesses the quality of a design, rather than being proactive in guiding the design.

### 3.2 CW

Cognitive walkthrough is a cognitive-based method aimed at uncovering usability issues by following the sequence of actions a user would take to perform a specific task, and by analysing at each stage how successful the user would be in performing the action correctly against psychological criteria (see [Wharton *et al*, 1994]) and is based upon the CE+ theory of exploratory learning developed by Polson and Lewis [Polson *et al*, 1992]. The method has gone through various iterations of development, being originally very theory based, and then simplified in order to make it more accessible. There has since been a return to more explicit dependence on the theory in order that decisions made can be justified by cognitive knowledge.

Since cognitive walkthrough involves tracking a user's actions whilst performing a set task or series of tasks, the method takes a task-oriented perspective, in that it concerns itself deeply with the goal structure and flow of the goals in completing the task. At every stage the interface is evaluated to determine whether or not it provides the necessary information for the user to successfully continue with the task, and what feedback the interface provides to assure the user that they are on the right path. CW tries to ensure that the interface supports and complements the problem-solving processes of the user. The analysis of the user actions is done in terms of success and failure stories. Success stories are where all the guidelines are satisfied. Failure stories are where one or all of the guidelines are not fulfilled.

There are four questions that must be answered before the analysis begins [Wharton et al, 1994]:

- Who will be the users of the system?
- What task (or tasks) will be analysed?



- What is the correct action sequence for each task and how is it described?
- How is the interface defined?

The analysts then use four questions (used as guidelines) as the basis for evaluating the interface, and ask the questions at each point in the interaction process [Wharton et al, 1994]:

- Will the users try to achieve the right effect?
- Will the user notice that the correct action is available?
- Will the user associate the correct action with the effect trying to be achieved?
- If the correct action is performed, will the user see that progress is being made toward solution of the task?

Cognitive walkthroughs primarily concentrate on ease of learning, on the basis that if something is easy to learn then it will also be easy to use and thus have less usability problems associated with it. This perspective is justified by the fact that users tend to learn features of an interface as they come across a need for that feature, rather than all at once [Wharton et al, 1994]. Therefore, ease of learning of an interface, and clear guidance in choosing features is seen as essential to interface usability.

However, one of the main criticisms of CW has been that by simplifying the underlying cognitive basis into a checklist of questions, there is no longer the support built in to justify problem identification or to provide rational and well-thought out solutions. Since then there has been a move back to a more theory-based approach, which has other difficulties associated with correct application of the method. An associated criticism is that the method is used too far down the development cycle, when bad design decisions are probably already in the system, and big changes are too expensive to be implemented. There is a substantial difference between uncovering a potential usability problem, and providing a reasoned solution to such a problem. CW also suffers in that it only considers fairly straightforward tasks, and does not take a wider view. It is limited in scope, often cannot generalise from individual results, and can be tedious and time-consuming in application. It only tends to identify less important problems, and can miss general overall problems.

Yet despite these disadvantages, CW can be straightforward to learn and apply, and can give a useful insight into a certain section of usability problems.

### 3.3 PUM

In contrast to the two methods discussed above, PUM is still at the development stage, and is thus not widely used. It attempts to bring together aspects of user modelling and system modelling into one representation. Based on a cognitive model, it is designed to be used when the system has been developed up to a point where what tasks the system is to cover, and a general idea of how those tasks can be carried out, are known [Blandford and Young *op cit.*]. A description of the knowledge that the user needs to operate the



interface successfully is written in the Instruction Language (IL), which is then compiled by the cognitive model in order to give an indication of the problems that might arise. Development of PUMS has so far concentrated on the IL, since the use of this to represent the problem space has proved to be the most difficult aspect of the method.

The IL description (in brief, for a fuller account see [Blandford and Young, 1996]) is made up from:

- the conceptual objects that the user manipulates
- relations between defined object types in terms of functional relationships and predicates
- a device description including commands, the initial state, and information displayed to the user
- user knowledge in terms of conceptual operators, initial knowledge, and user task

Therefore it is a fairly formalised description of the interface, although if the IL is not to be used to generate a running model then the description does not need to be so accurately defined and can be written in a more general and flexible form.

Two of the reasons why PUM can be so valuable are the running cognitive model which gives greater insight into the usability issues, and the actual act of describing the interface in such a detailed and explicit manner which can often draw the developer's attention to important issues. A third issue [Blandford and Young, *op cit*] is that the method's use and training encourages systems to be designed and developed from the user's perspective right from the start. PUMS main limitations are that it looks more at novice behaviour and cannot represent errors.

### 3.4 STN

State Transition Networks are a popular and well-established way of diagrammatically representing an interface [Dix et al, 1993]. There are various forms, but the simplest type has each state of the system represented by a circle, linked by lines, or transitions, which correspond to the action necessary to move from that state to another. Thus a STN is able to represent a flow or sequence of interaction in a relatively simple format.

There are some drawbacks to using STNs, in that they are not particularly able to represent concurrent actions, and help systems and escape options are difficult to represent in a clear and uncluttered manner [Dix et al, 1993]. However, for simple interaction sequences they can clearly illustrate the flow of interaction and allow redundant cycles to be identified. Not only are they easy to use, requiring little training and being quick to learn, they are also quick to write, uncomplicated, and allow an overall visual representation of the interface to be communicated to the analyst.



### 3.5 Z

Z is a formal specification notation based on set theory and first order predicate logic. It uses schemas, or collections of named objects with relationships specified by axioms. These schemas can be built up to define large specifications. Z was developed by the Programming Research Group at Oxford University Computing Laboratory in the late 1970s, and is one of the most widely used formal notations used in industry across the world for helping to develop software. It is currently undergoing international standardisation.

The mathematical base of Z means that it can be considered to be unambiguous, which makes it a powerful notation for communicating ideas and concepts. Formal methods are potentially a very useful tool in examining system usability, since not only do they allow aspects of a system to be described in detail, and the implications to be examined whilst ignoring other issues through abstraction, they also allow a design to be changed and the modifications assessed. By using a formal notation to construct a model of a system, they allow the designer to gain an insight into the structures and relationships that are of importance, and to manipulate those relationships and examine the implications of change without needing an actual implementation of a system. The idea of using a Z specification to reason about usability issues is not new [one such work being Hyde and Duke, 1996], since it's structure allows the problem space to be clearly defined and examined.

## 4. THE ROBOTIC ARM

The Advanced Manufacturing and Mechatronics Centre at Middlesex University is currently developing a robotic manipulator for use by wheelchair-bound people. The arm is intended to be used in a domestic context for everyday tasks such as feeding and grooming, and has been developed primarily to prove that a sophisticated manipulator can be produced at a reasonable cost, with usability issues being considered informally if at all. The arm consists of eight joints, powered by motors, which can move either individual joints, or the whole arm at once, via the input devices.

The input devices interface with a Windows-based application which in turn sends motor control commands in a special command language to a dedicated microprocessor, which actually controls the movement of the arm. For the purpose of the analysis, only one task is being considered, which uses only a small part of the interface. However, the task is one that will be very common to all users, and therefore will give valuable information on the usability of the interface. The task is to move the robotic arm to a certain position, without making use of any pre-taught positions, as though it were to be used to turn on a light switch. It is this kind of task that the developers of the arm consider to be a basic task, and that should be part of the core functionality of the interface. From the main menu of the application, this covers the options **Move** and **MoveArm**. **Move** allows the user to specify a particular arm joint and in what direction it can be moved, as well as controlling its speed. **MoveArm** allows the user to move the arm as a whole in a



particular direction. At present there is no feedback to the user other than that provided by the visual feedback of the arm's position.

The interface has not yet been fully implemented, but it is going to be implemented as a Windows application, using a menu format. Menu options will be selected in order to operate the arm. There are two methods of input, which can be used concurrently or as alternatives.

The gesture input system is based on a baseball cap with two sensors: one allowing movement forwards and backwards to be detected, the other allowing movement left and right to be detected. This allows a variety of unique gestures to form the gesture vocabulary. The gesture system is presently implemented so that a cursor moves along underneath the menu options continuously in turn, and if the correct gesture is made when the cursor is underneath a particular option, then that option is selected. Another gesture acts as a toggle between high and low speed of the cursor. A final gesture is an escape option, which automatically stops the arm if the arm is moving, and returns the user to the main menu.

The voice recognition system allows direct menu option selection simply by saying the menu option out loud. It is designed to be trained to individual voices, and needs resetting over time due to the way that voices change.

The use of a choice of input allows for discreet use in social situations. Also, if one method is not recognised by the computer, for instance because the gesture or word was imperfectly formed, then the alternative mode of input can be used. If both means of input are used at the same time but for conflicting inputs, then it would depend on which input was processed first as to which input was selected first. The output is not yet implemented, but will take the form of a small LCD display attached to the base of the arm showing the menus.

The users of the system are people with severe motor disabilities, unable to do elementary physical tasks unaided. These people may also have restricted head movements and voice capabilities. Their intelligence will range right across the scale, from below average to well above average. They may well be inexperienced in interacting with computer interfaces, given the nature of their disabilities, and since there is a lack of such robotic arms, they may not be sure about how to go about operating one. Therefore both in terms of the task and the style of interface the users will probably be inexperienced. However, given that the robotic arm is designed to help them live a more independent life, they will probably be very motivated to learn how to use the system.

### **Arm Task Description**

The task for all five analyses was the same, and was thought of in terms of device task, as in what the arm would accomplish, and user task, as in what the user's goal was. The information on which the analyses were based were obtained from a plan of the proposed

interface provided by B. Parsons, chief developer of the robotic arm, and notes taken of how those parts of the interface already implemented worked.

Device task:

Move gripper to position (x,y,z) from rest, with no pre-taught position.

User task:

Turn on light switch.

For a full description of how the arm works, please refer to [Parsons *et al*, 1995] and Appendix A at the end of this report.

## 5. THE FIVE TECHNIQUES APPLIED

### 5.1 The GOMS Analysis

#### Introduction

The interface to the robotic arm was analysed in two stages. The first stage analysed the interface with regard to a certain task, that of moving the arm to switch on a light, a task that would be very common in the proposed use of the arm. Since at this point the input devices were not fully implemented, the analysis could not be taken down to the level of clearly defined motor, perceptual and cognitive operators. Instead, an overall device description was produced, and the task examined in this context. This stage was written using CMN-GOMS. CMN stands for Card, Moran and Newell, and is the version of GOMS which was presented in [Card *et al*, 1983]. This version of GOMS was chosen for its simplicity of learning, and has a strict goal hierarchy, with methods represented in a program form as sequences of steps which must be performed in order. Several usability issues were noted, and suggestions for improving the interface, along with a revised device description in CMN-GOMS, were put forward. For the original device description of this interface using CMN-GOMS, and the improved device description, please refer to Appendices C and D.

The next analysis was done once the input devices had been determined, in order to examine what leverage a GOMS analysis could give on the multi-modal usability aspects of the interface. Since the interface design had also changed slightly, a new CMN-GOMS analysis was done, this time of a simple task rather than the whole interface, then the analysis was taken down to a CPM-GOMS level in order to examine more fully the cognitive, motor and perceptual aspects of the interaction. CPM stands for both Cognitive-Perceptual-Motor, and also Critical Path Method. It is based on the assumption that tasks needing different processes within the Model Human Processor as put forward by [Card *et al*, 1983] can be performed in parallel. To use CPM-GOMS, a CMN-GOMS analysis is first done of the goal hierarchy and methods in order to obtain the basic perceptual, cognitive and motor operators, which are then expressed using schedule charts.



## First Analysis

One of the most obvious features of the interface brought out by the first CMN-GOMS analysis was the amount of interaction needed for a user to move the arm. In order to move the arm to a specific position, the user had to first decide whether to move the arm as a whole, or to move an individual joint. If it was decided to move just a joint, the user then had to specify which joint, the direction of movement of the joint, to adjust the speed of the joint if necessary, start the joint moving, stop the joint moving, and then decide whether or not to continue moving that same joint or to move either the whole arm or another joint. In either of the latter choices, the user would have to return to the main menu. There was no other way of going back to the selection of joint stage. Given that the potential users are disabled and therefore have restricted interaction capabilities, this seemed a large and complicated interface for accomplishing the goal of moving the arm. Indeed, it is worth showing in full the entire list of goals to move one joint, to demonstrate both the number of stages of interaction, and their relative complexity:

### GOAL: MOVE ARM TO POSITION XYZ

```
* DETERMINE CURRENT POSITION OF ARM
* GOAL: MOVE WHOLE ARM (repeat until arm near position XYZ)
* * REMEMBER MEANINGS OF MOVE AND MOVEARM OPTIONS
* * [SELECT: GOAL: USE MOVE...if arm is in right area but not
extended
* * appropriately
* * GOAL: USE MOVEARM...if arm is not in right
area]
* * GOAL: USE MOVE
* * * PRESS MOVE OPTION
* * * GOAL: SELECT JOINT
* * * REMEMBER MEANINGS OF JOINT OPTIONS AND
END
* * * * OPTION
* * * * [SELECT: GOAL: USE BASE...if base needs moving
* * * * GOAL: USE ARM...if arm needs
moving
* * * * GOAL: USE SHOULDER...if
shoulder needs moving
* * * * GOAL: USE ELBOW...if elbow
needs moving
* * * * GOAL: USE HAND...if hand needs
moving
* * * * GOAL: USE WRIST...if wrist needs
moving
* * * * GOAL: USE PALM...if palm needs
moving
```

* * *	* GOAL: USE GRIPPER...if gripper
needs moving	
* * *	* GOAL: END OPERATION...if none
need moving]	
* * *	* GOAL: USE ELBOW
* * *	* GOAL: SELECT JOINT MOVEMENT
* * *	* REMEMBER MEANINGS OF JOINT MOVEMENT
OPTIONS	
* * *	* AND END OPTION
* * *	* [SELECT: GOAL: USE IN...if joint needs moving in
* * *	* GOAL: USE OUT...if joint needs
moving out	
* * *	* GOAL: END OPERATION...if none
need moving]	
* * *	* GOAL: USE IN
* * *	* * PRESS IN OPTION
* * *	* GOAL: MOVE JOINT
* * *	* REMEMBER MEANINGS OF OPTIONS
* * *	* GOAL: SELECT JOINT MOVE
* * *	* * REMEMBER MEANING OF OPTIONS
* * *	* * [SELECT: GOAL: END OPERATION...if
operation to end	
* * *	* * GOAL: SELECT
MOVEMENT...if operation to	
* * *	* * continue]
* * *	* * GOAL: SELECT MOVEMENT
* * *	* * REMEMBER MEANINGS OF OPTIONS
* * *	* * [SELECT: GOAL: USE GO...if speed at
correct level	
* * *	* * GOAL: SET
SPEED...if speed not at correct	
* * *	* * level]
* * *	* * GOAL: SET SPEED
* * *	* * PRESS OPTION SPEED LEVEL
* * *	* * [SELECT: GOAL: USE MIN...if
slowest speed	
* * *	* * * required
* * *	* * * GOAL: USE
SLOW...if slow speed	
* * *	* * * required
* * *	* * * GOAL: USE
MEDIUM...if medium	
* * *	* * * speed required
* * *	* * * GOAL: USE
FAST...if fast speed	



```

*      *      *      *      *      *      *      required
*      *      *      *      *      *      *      GOAL: USE
MAX...if fastest speed
*      *      *      *      *      *      *      required]
*      *      *      *      *      *      *      PRESS OPTION GO
*      *      *      *      *      *      *      GOAL: USE SLOW
*      *      *      *      *      *      *      *      PRESS OPTION SLOW
*      *      *      *      *      *      *      GOAL: FINISH MOVING JOINT(repeat until joint in
correct
*      *      *      *      *      *      *      position)
*      *      *      *      *      *      *      GOAL: SELECT STOP MOVEMENT
*      *      *      *      *      *      *      *      [SELECT:  GOAL: USE STOP...if arm
will otherwise
*      *      *      *      *      *      *      move beyond correct
position
*      *      *      *      *      *      *      GOAL: NOT USE
STOP...if arm will not
*      *      *      *      *      *      *      move beyond correct
position]
*      *      *      *      *      *      *      GOAL: USE STOP
*      *      *      *      *      *      *      *      PRESS STOP OPTION WHEN
ARM IN CORRECT
*      *      *      *      *      *      *      POSITION

```

That is fifteen goals to move one joint! This is a very large number of goals to do something which should be relatively straightforward, given that the purpose of the arm is to assist movement. To move the arm as a whole, the number of goals was slightly less at thirteen, but still large.

It was also noted that there were no short cuts. For a user to move the arm, the user had to follow the interface all the way through, which, as described above, could be quite a complicated series of steps. However, it was decided that another part of the system not under analysis did allow for pre-taught positions to be stored and used when required, which could possibly fill a need for short-cuts by an expert user.

It was noted that there was a distinction made in the interface between moving the arm as a whole, and moving an individual joint of the arm, although the sequence of interaction necessary to move the arm as a whole was almost identical to that needed to move the arm joints individually. For instance, the goal for setting the speed of movement was the same for both, as detailed below:

#### GOAL: SET SPEED

```

*      PRESS OPTION SPEED LEVEL
*      [SELECT:  GOAL: USE MIN...if slowest speed required
*      GOAL: USE SLOW...if slow speed required

```

```

*           GOAL: USE MEDIUM...if medium speed required
*           GOAL: USE FAST...if fast speed required
*           GOAL: USE MAX...if fastest speed required]
*   PRESS OPTION GO
*   GOAL: USE MIN
*   *   PRESS OPTION MIN
*   GOAL: USE SLOW
*   *   PRESS OPTION SLOW
*   GOAL: USE MEDIUM
*   *   PRESS OPTION MEDIUM
*   GOAL: USE FAST
*   *   PRESS OPTION FAST
*   GOAL: USE MAX
*   *   PRESS OPTION MAX

```

The goals of **SELECT STOP MOVEMENT** were also the same, as were others. Therefore, given that the sequences were almost identical, and in some cases individual low level goals were identical, it seemed as though the distinction between the two commands was artificial, and created an unnecessary burden on the user in having to navigate two separate paths to accomplish almost the same thing.

It was noticed that there were some levels of interaction that were apparently not necessary. When moving the joint or arm, once movement had stopped for some reason, the user is given the option to **Continue**. However, in terms of what **Continue** actually does, it is exactly the same as choosing **Go**, in that both start the movement and lead on the option which allows the user to then stop the movement or wait for movement to stop automatically. This suggests that **Continue** is in fact redundant, and could be accomplished by cycling the interface back to the **Go** option once movement has stopped, thus reducing the level of complexity of the interface.

One of the joints was named **Arm**, which could lead to confusion by a user between the arm as a whole and the specific joint. This was noticed when the goals were being written out, and is something that could possibly have been found without conducting any analysis. However, the very act of examining the interface in depth meant that this was noticed.

#### **Improved First CMN-GOMS Analysis:**

From the first analysis, a second interface was postulated, and defined using CMN-GOMS in order to show how a user might go about using the arm when the interface has been changed to reduce complexity and allow for easier manipulation. For a full description, please refer to **Appendix D**.

For instance, when analysing the interface it was felt that the number of goals needed to move one joint was, at fifteen, an excessive number. In the revised analysis, the number



of goals is reduced to twelve, which, although still large, is slightly better than previously. The entire list is shown below:

**GOAL: MOVE ARM TO POSITION XYZ**

- \* DETERMINE CURRENT POSITION OF ARM
- \* **GOAL: USE MOVE**
- \* \* PRESS MOVE OPTION
- \* \* **GOAL: SELECT JOINT/ARM**
- \* \* \* REMEMBER MEANINGS OF OPTIONS AND END OPTION
- \* \* \* [SELECT: GOAL: USE BASE...if base joint needs moving
- \* \* \* GOAL: USE ARM...if arm joint needs
- moving
- \* \* \* GOAL: USE SHOULDER...if shoulder joint
- needs moving
- \* \* \* GOAL: USE ELBOW...if elbow joint needs
- moving
- \* \* \* GOAL: USE HAND...if hand joint needs
- moving
- \* \* \* GOAL: USE WRIST...if wrist joint needs
- moving
- \* \* \* GOAL: USE PALM...if palm joint needs
- moving
- \* \* \* GOAL: USE GRIPPER...if gripper joint
- needs moving
- \* \* \* GOAL: USE WHOLE ARM...if whole of
- arm needs
- \* \* \* moving
- \* \* \* GOAL: END OPERATION...if nothing
- needs moving]
- \* \* \* **GOAL: USE ELBOW**
- \* \* \* \* PRESS ELBOW OPTION
- \* \* **GOAL: SELECT MOVEMENT**
- \* \* \* REMEMBER MEANINGS OF JOINT AND WHOLE ARM
- \* \* \* MOVEMENT OPTIONS AND END OPTION
- \* \* \* [SELECT: GOAL: USE IN...if individual joint needs moving
- in
- \* \* \* GOAL: USE OUT...if individual joint needs
- moving out
- \* \* \* GOAL: USE UP...if whole arm needs
- moving up
- \* \* \* GOAL: USE DOWN...if whole arm needs
- moving down
- \* \* \* GOAL: USE FORWARD...if whole arm
- needs moving
- \* \* \* forwards

* * *	GOAL: USE BACK...if whole arm needs
moving	
* * *	backwards
* * *	GOAL: USE LEFT...if whole arm needs
moving to the left	
* * *	GOAL: USE RIGHT...if whole arm needs
moving to the	
* * *	right
* * *	GOAL: END OPERATION...if no part of
the arm needs	
* * *	moving]
* * *	<b>GOAL: USE IN</b>
* * *	* <b>PRESS IN OPTION</b>
* * *	<b>GOAL: MOVE(repeat until arm or joint is in correct position)</b>
* * *	<b>REMEMBER MEANINGS OF OPTIONS</b>
* * *	<b>GOAL: SELECT TO MOVE</b>
* * *	* <b>REMEMBER MEANING OF OPTIONS</b>
* * *	* <b>[SELECT: GOAL: END OPERATION...if operation to</b>
end	
* * *	* <b>GOAL: USE GO...if speed at correct</b>
setting	
* * *	* <b>GOAL: SELECT SPEED...if speed</b>
at incorrect	
* * *	* <b>setting]</b>
* * *	* <b>GOAL: SELECT SPEED</b>
* * *	* <b>PRESS OPTION SPEED LEVEL</b>
* * *	* <b>[SELECT: GOAL: USE MIN...if slowest speed</b>
required	
* * *	* * <b>GOAL: USE SLOW...if slow</b>
speed required	
* * *	* * <b>GOAL: USE MEDIUM...if</b>
medium speed	
* * *	* * <b>required</b>
* * *	* * <b>GOAL: USE FAST...if fast</b>
speed required	
* * *	* * <b>GOAL: USE MAX...if fasted</b>
speed required]	
* * *	* * <b>PRESS OPTION GO</b>
* * *	* * <b>GOAL: USE SLOW</b>
* * *	* * * <b>PRESS SLOW OPTION</b>
* * *	<b>GOAL: SELECT FINISH MOVING</b>
* * *	* <b>GOAL: USE STOP</b>
* * *	* * <b>PRESS STOP OPTION WHEN ARM IN</b>
CORRECT	
* * *	* * <b>POSITION</b>



Although this is not a great reduction on the number of goals in the original analysis (from fifteen to twelve, and from thirteen to twelve), the goals have been simplified by merging them where possible with the goals for moving a joint. This means that instead of having to follow two separate paths for moving joints and the whole arm, only one path needs following, and thus continuity of interaction is assured. Also, by making the goals for moving a joint and moving the whole arm as similar as possible, it reduces the goal-forming load on the user, requiring them to make less complicated decisions.

It was noted earlier in the original analysis that the option **Continue** is a redundant feature, in that **Go** would work just as well at that stage of the interaction and possibly cause less confusion. Therefore, in the improved GOMS description, the use of **Continue** has been dropped, and the interface instead cycles back to the **Go** option once movement has stopped.

### **CPM-GOMS Analysis**

The interface was then analysed using CMN-GOMS once the input devices had been implemented, and the analysis then taken down to the CPM-GOMS level. The overall structure was very similar as that in the analyses described above, and the only major difference was the inclusion of operators for using voice and gesture. For a full description of the CMN-GOMS portion of the analysis, please refer to **Appendix E**.

The methods for using voice and gesture varied in their basic operators, with the gesture method having the most operators:

For Gesture:

- ~ retrieve meaning of option
- ~ find option on screen
- ~ wait for cursor to move to that option
- ~ retrieve correct gesture
- ~ perform correct gesture
- ~ confirm option

For Voice:

- ~ retrieve meaning of option
- ~ say option out loud
- ~ confirm option

The sequence of gesture operators is twice as many as that for voice. This is because the gesture system is dependent upon the movement of the cursor to the right option, whereas the voice system has no such system constraints.

The methods stayed the same for each goal, with the exception of the methods needed for the goal **USE STOP**, where the operators varied for both voice and gesture:

For Gesture:

- ~ look at arm
- ~ determine when arm to stop
- ~ retrieve meaning of option
- ~ retrieve correct gesture
- ~ wait for arm to reach stop position
- ~ perform correct gesture
- ~ confirm option

For Voice:

- ~ look at arm
- ~ determine when arm to stop
- ~ retrieve meaning of option
- ~ wait for arm to reach stop position
- ~ say option out loud
- ~ confirm option

For this goal, both means of interaction are constrained by the movement of the arm, in that the user has to wait for the arm to reach the desired position before the command can be given. Again, the voice system had less operators, but the difference here was that of one operator, meaning that for this goal there was little apparent difference in the effort needed to accomplish it by either input device.

This initial CMN-GOMS analysis of the input devices was then taken down to a CPM-GOMS level, in order to examine the critical path of interaction, and see to what extent GOMS could give useful information on multi-modal usability aspects of the interface. Two schedule charts were drawn for each input device, one for the general method of interaction, and one which corresponded only to the **USE STOP** goal. For a full representation of the charts, please refer to **Appendix F**.

The schedule charts for the ordinary voice and gesture operators were straightforward. In both, there was very little parallel processing, or using of perceptual, cognitive and motor abilities at the same time, because of the nature of the critical path, which did not generally need this. Again, from an initial examination, it appeared that the voice input device needed fewer operators than the gesture system, and thus would be the quicker and less intense form of interaction.

The schedule charts for the **USE STOP** goal were interesting, in that the critical path on the gesture chart was identical to that on the ordinary gesture chart in terms of the pattern of perceptual, cognitive, motor and system operators used, although the actual use of those operators was different. Instead of looking at the screen, the second gesture chart looked at the arm, and instead of waiting for the cursor to move to the correct position, in the second chart the user waits for the arm to move to the correct position, since the cursor is already in place. Thus for the two possible gesture sequences of operators, the



only substantial difference was the system time in waiting for the cursor or arm to move to the correct position. The voice critical path for the **USE STOP** goal was different to the original chart only because of the addition of the waiting for the arm to move to the correct position.

No times are given on any of the schedule charts for the duration of the operators, since there are too many uncertainties. The system response time will obviously vary, as will the time taken to vocalise a word, depending on the word, or to perform a gesture. There are also problems in establishing exactly how long a cognitive or perceptual operator may take. Therefore no timing comparisons between the charts can be given, which may lessen their usefulness in determining length of interaction.

Another problem is that although the critical paths given on the schedule charts appear reasonable and may well be correct, there are other ways of representing portions of the interaction. For example, the ordinary voice schedule chart shows the path of finding the correct option on the screen as a sequence of operators going:

Retrieve meaning of option, find option on screen, perceive info, verify info, retrieve word

whereas there are alternative ways of representing this. A naive user might first formulate goal, then read the first option, compare it to the one looked for, and if wrong scan right or down and repeat the process. Another way, utilising expert spatial memory, might have the user formulating the goal, recalling the position of the relevant option, moving eyes to that position, reading the information there, and verifying it. Therefore there is no way to explicitly determine exactly how to represent the critical paths for the methods of interaction within CPM-GOMS. Only a best interpretation can be given.

## **GOMS Conclusions**

The interface was particularly suited to a GOMS analysis, due to the goal-based nature of the task, the straightforward goal structure, and the relative simplicity of that portion of the interface under investigation. It is difficult to know if the goal structure used was the most accurate, since there may be other ways to represent goals, but this is a recurrent problem in analyses of this type. The low level operators would be the same, because the same sequence of operations will be needed. The lack of selection rules in this task analysis between significant methods shows that there was only one probable way for the user to accomplish the given goal, and that the user was unlikely to do it any other way. However, it must be noted that this goal structure was determined by the device. The user might wish to structure goals differently to how the interface allows.

The contribution of the initial CMN-GOMS analysis has been twofold. By using CMN-GOMS to analyse the interface, we have been able to reason about aspects of the interaction and discuss potential problems before the interface has been fully implemented and major commitments made. Despite not taking the model down to the keystroke level, important usability issues in terms of interface structure have been



uncovered. Problems relating to the long sequences of operators needed to accomplish a goal, the inability of the user to backtrack, the difficulty of distinguishing between MoveArm and Move, the lack of shortcuts, and other important points were clearly identified. We have also been able, using an analysis of user goals, to put forward an alternative structure to lead to a better interface. This version of GOMS has therefore been shown to have utility both in the analysis of existing interfaces and in the generation of alternative interfaces.

The second analysis makes apparent the different flavours of the two versions of GOMS used. Whereas the CMN-GOMS approach is very hierarchical, with methods represented in sequence, and operators performed in sequence, CPM-GOMS can represent parallel activity, and is thus potentially more suitable for representing multi-modal activity. However, given that only one form of input could be used at any given time, it could be argued that this was not strictly speaking a multi-modal situation, and therefore comments on CPM-GOMS's suitability for representing such forms of interaction may not be relevant.

CPM-GOMS provides a fine grain of detail for reasoning about precise operations in a particular modality. The use of schedule charts allows the different types of processing by both user and system to be apparent, and to therefore be reasoned about. The analysis clearly brings out the differences between the methods for the voice and gesture input devices, and it is thus useful for comparing them against each other. However, it is difficult to represent both modalities when used simultaneously, because we are still unable to give a definitive picture of what actually happens at the human processing level. GOMS is dependent on the Model Human Processor structure of information processing, and it is not yet known for certain as to how accurate this model is.

Another issue is that CPM-GOMS gave no basis for reasoning about important aspects of multi-modality such as modality selection. Selection may be based on many different factors such as social setting, user familiarity, ease of use, perceived times, and user capabilities. CPM-GOMS cannot model such factors, although they will be very important to the user of the robotic arm in deciding which input device to use to accomplish a given goal. Therefore CPM-GOMS, although it initially looked promising in examining the multi-modal aspects of the interface, proved to be able to contribute little useful information over and above that contributed by the CMN analysis.

In conclusion, this was a straightforward use of straightforward techniques for a straightforward interface. With reference to the use of this technique for dealing with multi-modal issues, the situation is more complex, and is hampered by the lack of clear research into multi-modal usability. The CMN-GOMS part of the analysis could not deal with multi-modal issues at all, due to its hierarchical and sequential structure. CPM-GOMS was able to look at the basic operators in considerable depth with regard to the load on human processing ability along a critical path, but suffered from the lack of clear knowledge about how humans process information, and the difficulty of representing



simultaneous interaction. There were certain important issues that could not be represented at all.

## 5.2 The CW Analysis

### Introduction

The interface was first analysed without reference to specific input devices. Then the interface was analysed with voice and gesture input devices. The two analyses are discussed at the same time for the purposes of this report. It was noted earlier [see 4. **Robotic Arm**] that the users are potentially disabled, with severe restrictions on their movements and a wide range of reasoning abilities. The action sequence for the task was that the arm needed to be moved to the right, then the arm part moved up and the gripper moved out. For a full description of the required actions for accomplishing this task, please refer to **Appendix G**.

### The Walkthrough: Step-by-Step Analysis Phase

The interaction sequence was worked through step by step, and the interaction analysed according to the following four questions:

- Will the users try to achieve the right effect?
- Will the user notice that the correct action is available?
- Will the user associate the correct action with the effect trying to be achieved?
- If the correct action is performed, will the user see that progress is being made toward solution of the task?

Rather than showing the whole of the report, four of the more interesting and illuminating stages of the walkthrough are discussed below. These are: the standard choice of options from the main menu in terms of the user wishing to move the whole arm; choosing an ordinary option from one of the menus; how the user might stop the arm when in motion; and finally, choosing to move only part of the arm. These four stages give an overall flavour of the CW as a whole, and show a portion of the range of issues uncovered.

1. Choose option **MoveArm** on the main menu.

The interface gives a list of options: **Left, Right, Up, Down, Forward, Back, Wrong, End**

This action is used when the user wishes to move the arm as a whole. The usability issues uncovered in response to the questions are shown below.

*Will the users try to achieve the right effect?*

Success story: the user knows how to interact with the system using the voice and gesture input devices, because they have been trained beforehand. They can see the arm and visually assess what movement is needed.

Possible failure story: the user may not know the difference between the options **MoveArm** and **Move**, and may be confused when trying to decide between them. The user may not know (although this is less likely) that the arm can be moved as a whole instead of individually.

*Will the user notice that the correct action is available?*

Success story: the user will know that they can select an option by nodding their head in a particular manner or by vocalising the word, because they will have been trained beforehand. The cursor will prompt the use of the gesture, the menu names will prompt the use of the voice input device.

*Will the user associate the correct action with the effect trying to be achieved?*

Success story: the user will know how to select an option to move on to the next menu because of previous training.

Possible failure story: there may be a problem if the user moves their head to look from the interface to the arm and back, in that according to how the gesture system is implemented it may be interpreted as a command. There may be a similar problem if the user is engaging in a conversation while the voice input system is operational, although this is unlikely due to the small number of possible menu names. There may be a possible implementation problem if the user pauses in middle of saying "Move arm".

*If the correct action is performed, will the user see that progress is being made toward solution of the task?*

Success story: the user will know that progress has been made because the next menu, detailing the next stage of interaction will appear, and the user will know from training that this new menu relates to the next stage of interaction.

Possible failure story: the menu that appears after the **MoveArm** option is chosen gives no indication that the whole arm is going to be moved. Feedback to the user will need to be considered at this point so that the user will know what options have been chosen.

2. Choose option **Right** on the menu.

The interface gives a list of options: **Go, Speed, Wrong, End**

This refers to the choice of direction that the user wishes the arm to be moved in. Again, the usability issues relating to this are noted.



*Will the users try to achieve the right effect?*

Success story: the user knows how to interact with the system using the voice and gesture input devices, because they have been trained beforehand.

Possible failure story: the user may not realise that the directions correspond to the user (i.e. right is to the user's right, etc.). There will need to be feedback so that user can check that right direction was chosen.

*Will the user notice that the correct action is available?*

Success story: the user will know that they can select an option by nodding their head in a particular manner or by vocalising the word, because they will have been trained beforehand. The cursor will prompt the use of the gesture, the menu names will prompt the use of the voice input device.

*Will the user associate the correct action with the effect trying to be achieved?*

Success story: the user will know how to select an option to achieve the desired effect because of previous training.

Possible failure story: there may be a problem if the user moves their head to look from the interface to the arm and back, in that according to how the gesture system is implemented it may be interpreted as a command. There may be a similar problem if the user is engaging in a conversation while the voice input system is operational, although this is unlikely due to the small number of possible menu names. There may be another possible failure story if the arm is contorted, resulting in it being difficult to establish which way is left or right according to individual joints or the arm as a whole.

*If the correct action is performed, will the user see that progress is being made toward a solution of the task?*

Success story: the user will know that progress has been made because the next menu, detailing the next stage of interaction will appear, and the user will know from previous experience that this new menu relates to the next stage of interaction.

Possible failure story: the user cannot check that the direction choice is that intended until the arm actually starts to move.

6. When the arm appears to be in the correct position, choose **Stop**.

The interface gives the options: **Go, Speed, Wrong, End**

Stopping the arm in the correct position is an important part of the interaction, and the associated problems in how to use the **Stop** option are seen clearly at this stage of the analysis.

*Will the users try to achieve the right effect?*

Success story: the user knows how to interact with the system using the voice and gesture input devices, because they have been trained beforehand.

*Will the user notice that the correct action is available?*

Success story: the user will know that they can select an option by nodding their head in a particular manner or by vocalising the word, because they will have been trained beforehand. The cursor will prompt the use of the gesture, the menu names will prompt the use of the voice input device.

*Will the user associate the correct action with the effect trying to be achieved?*

Success story: the user will know how to select an option to move on to the next menu because of previous training.

Possible failure story: there may be a problem if the user moves their head to look from the interface to the arm and back, in that according to how the gesture system is implemented it may be interpreted as a command. There may be a similar problem if the user is engaging in a conversation while the voice input system is operational, although this is unlikely due to the small number of possible menu names. Another possible failure story is that the user may not realise that they can wait until the arm has reached the limit of movement before selecting stop. It may also take too long for the user to interact with the system, thereby leaving the arm in a wrong position.

*If the correct action is performed, will the user see that progress is being made toward solution of the task?*

Success story: the user will know that progress has been made because the next menu, detailing the next stage of interaction will appear, and the user will know from previous experience that this new menu relates to the next stage of interaction. Also, the arm will be in the correct position to begin the next phase of the interaction.

9. Choose option **Arm** from the menu.

The interface gives a list of options: **In, Out, Wrong, End**

Here the user wishes to move only part of the arm, which can cause certain problems as described below.

*Will the users try to achieve the right effect?*

Success story: the user knows how to interact with the system using the voice and gesture input devices, because trained beforehand.



Possible failure story: the user might not know that in this context arm is the name of a joint, rather than referring to the whole arm, so might chose it in error thinking that it is for moving the whole arm, or not choose it thinking that it will move the whole arm.

*Will the user notice that the correct action is available?*

Success story: the user will know that they can select an option by nodding their head in a particular manner or by vocalising the word, because they will have been trained beforehand. The cursor will prompt the use of the gesture, the menu names will prompt the use of the voice input device.

*Will the user associate the correct action with the effect trying to be achieved?*

Success story: the user will know how to select an option because of previous training.

Possible failure story: there may be a problem if the user moves their head to look from the interface to the arm and back, in that according to how the gesture system is implemented it may be interpreted as a command. There may be a similar problem if the user is engaging in a conversation while the voice input system is operational, although this is unlikely due to the small number of possible menu names.

*If the correct action is performed, will the user see that progress is being made toward solution of the task?*

Success story: the user will know that progress has been made because the next menu, detailing the next stage of interaction will appear, and the user will know from previous experience that this new menu relates to the next stage of interaction.

## **Analysis**

The usability problems uncovered seem to relate mainly to the labelling of the options. There is a problem in determining the difference between the options MoveArm and Move, as the user may not know at this stage which one concerns moving the arm as a whole, and which will move individual joints. There is an associated problem in that the user may also be unaware that the arm can be moved in these two different ways. One of the joints is called Arm, which might lead the user to assume that it refers to the whole arm rather than just one particular joint.

The directions of movement of the arm relate to the user, but the user will possibly be unaware of this at that time. Another problem brought out by the analysis regards the stop option. The user may be unaware that the arm will automatically stop when it reaches its furthest extent, and also, depending on the method of interaction, the user may have difficulties in stopping the arm at the correct position.

There were very few multi-modal issues brought out by this cognitive walkthrough, and those that were seem to be very similar and of superficial importance.

## **CW Conclusions**

There are various problems with cognitive walkthrough as an analysis method, and these were discussed briefly earlier in the report. However, this analysis has shown that the method is quick and easy to apply, and can uncover several different kinds of usability problems.

The main problem uncovered using this method was that of ineffective or misleading labelling of options. By assuming the perspective of a new user, the walkthrough was able to determine which labels could be considered effective and which needed changing. The other main issue uncovered was the need for effective feedback at all stages. The feedback to the user has not yet been implemented in this system: however, as a result of this cognitive walkthrough the areas where feedback is essential have been clearly determined, and thus any implementation will benefit from this analysis.

Cognitive walkthroughs do mention that failure can occur, but do not follow it up to give more information about the consequences of such failure. This is perhaps better than GOMS, which does not explicitly cover failure at all, but still leaves some usability issues unresolved.

This CW analysis does not necessarily cover important aspects of multi-modality, because of CW's emphasis on goal-structure and previous knowledge, rather than actual means and method of interaction. It was therefore unlikely to uncover multi-modal aspects unrelated to ease of learning and goal structure, such as issues of time and complexity. CW gave no basis for reasoning about important aspects of multi-modality such as modality selection, which may be based on many different factors such as social setting, user familiarity, ease of use, perceived times, and user capabilities. CW was unable to model such factors, although they will be very important to the user of the robotic arm in deciding which input device to use to accomplish a given goal. Thus CW proved to be able to contribute little useful information.

Therefore it seems that although this method has advantages in that it is quick and relatively easy to apply, it has problems in the range of usability issues uncovered, and the difficulty of recognising certain problems. Many problems seem to be uncovered more by the skill of the analyst than as a result of the question promptings.

## **5.3 The PUM Analysis**

### **Introduction**



The first PUM analysis superficially seemed as though it would be straightforward, due to the simplicity of the interface and the lack of input and output devices to consider. However, the interface proved to be less tractable to this kind of analysis than any of the other representations, due to the particular way in which Programmable User Modelling works, in that it considers the domain tasks and the device tasks and tries to map between them. The second analysis, which incorporated the input devices, added little to the usability understanding of the multi-modal aspects of the interface.

### First Analysis

The analysis of the robotic arm initially applied the standard PUM principle of focusing on domain tasks and how they are achieved using this device. This proved to be extremely problematic, because the mapping between domain tasks and the device is non-obvious. This is mainly because of the difficulty of expressing 3-D spatial movements in cognitively meaningful ways. The PUM representation is not suited to representing 3-D spatial planning. Essentially, the user's domain task is to plan (probably using partial planning, means-ends reasoning, and reactive behaviour) a route to get the gripper from the current position and orientation to the desired position. The difficulty of this task depends on the number of degrees of freedom the user is manipulating. Our intuition is that the task will initially involve large-scale movements, followed by detailed refinement for final positioning. People do not generally think of 3-D spatial tasks in terms of Cartesian co-ordinates (x,y,z), and the fact that Cartesian co-ordinates are used as an intermediate representation scheme will make this device very difficult to use. Thus there is a distinct mis-match between how the user would move the arm, and how the device actually works. For example, to move just one joint to a new position requires a different view to the domain tasks, as borne out by the analysis below:

To move just one joint to a new position:

```
task:          joint-at (J, XYZ)
```

The objects in the domain include the joints, directions of movement, and locations:

#### OBJECTS

```
joint:         gripper, arm, base, whole-arm,...etc.  
location:     abc, def, ... xyz, ...etc.  
direction:    in, out, up, down, left, right, ...etc.
```

The same word (e.g. left/right) may have different meanings depending on which joint is being moved, which can cause confusion. There may also be semantic confusion over joint names since the user has to know which name refers to which joint.

To define the possible degrees of freedom of the joints there need to be a large collection of statements:

```
can-move(joint, direction)
```

e.g.:

```
can-move(shoulder, in)
can-move(shoulder, out) ...etc.
```

This is, at least initially, a statement of what is true about the device. It is not specified immediately whether or not the user must know these things.

The task goal is achieved by setting the arm moving in the correct direction, then waiting until it gets to the right place (or near it) and pressing "stop". The level of description is at a level such that "wait then press STOP" is considered as one conceptual operation:

```
operation stop-at (joint: J, location: L)
user-purpose: joint-at(J,L)
precondition: direction-specified()=D
               is-moving(J)
               joint-specified()=J
filter:        is-within-range(J, L)
               can-move(J, D)
               L is in direction D from current position
action:       wait then press STOP
```

This presupposes that L is not at the end of the trajectory; if this were so, then the user would not have to press stop. However, many users would be likely to anyhow, as the knowledge that L is at the end of the trajectory is additional information that the user would have to use to select between the alternative possible actions (wait-then-press-stop vs. just-wait-until-it-stops-of-its-own-accord).

Direction, speed and joint have been specified as preconditions, which means that the user must know that these things have to be specified. In fact, the device enforces these conditions in that the user cannot set the joint moving without specifying these things. These are best considered as knowledge communication goals (the user has to communicate these parameters to the device at the appropriate time).

At each menu stage, the user has to identify the most relevant menu item. They may do this by already knowing which item they require, or by recognising which is most relevant to their current goal.

For each menu, the relevance of each menu item can be specified:

```
is known or recognisable:
direction menu:
is-relevant("end", joint-specified()=Jnew)
is-relevant("left", direction-
specified()=anticlockwise)
is-relevant("up", direction-specified()=up)
...etc.
```



```

initial menu:
is-relevant("move", joint-specified()=J such that J ≠
whole-arm)
is-relevant("movearm", joint-specified()=whole-arm)
joint menu:
is-relevant("end", joint-specified()=whole-arm)
is-relevant("gripper", joint-specified()=gripper)
is-relevant("arm", joint-specified()=arm)
...etc.
speed menu:
is-relevant("min", speed-specified()=slowest)
is-relevant("slow", speed-specified()=slow)
...etc.
go menu:
is-relevant("go", is-moving(J,L))
is-relevant("speed level", speed-specified()=S such
that          S≠current-speed)
is-relevant("end", direction-specified()=Dnew)
is-relevant("end", joint-specified()=Jnew)
continue menu:
is-relevant("continue", is-moving(J,L))
is-relevant("speed level", speed-specified()=S)
is-relevant("end", direction-specified()=Dnew)
is-relevant("end", joint-specified()=Jnew)

```

According to the available documentation, "end" is available as an option for all menus except for the initial and speed menus. However, the relevance of choosing "end" is different in different contexts. In particular, it is not clear that the user would have reason to press "end" to tidy up at the end of the interaction.

For all menu selections the operation is:

```

operation select-item(I)
user-purpose G
filter          is-relevant(I,G)
action          select(I)

```

The user may have several goals (e.g. to specify joint and direction and maybe speed) and not know which is to be addressed first. This matching involves identifying the most relevant menu item for addressing any of the active goals.

The analysis so far highlights the importance of the user recognising which option corresponds to their goal(s): that the semantic meaning of each option name must be clearly distinguishable. This is not clearly the case for, for example, "move" and "move arm".

As specified so far, the user's cognitive behaviour will be:

```

task          joint-at(J,XYZ)
->goal        joint-at(J,XYZ)
->operation   stop-at(J,XYZ)
->subgoals    direction-specified()=D
              is-moving(J)
              speed-specified()=S
              joint-specified()=J
**in this, J is defined via the task goal; D is defined
(via filters) as being the direction that takes J from
the current to the target location; S is undefined, so
there is no reason for it not to default to medium. So
assume this goal is already satisfied.
-> initial menu: select-item("move")
-> joint menu: select-item("J")
-> goal satisfied joint-specified()=J
-> direction menu: select-item("D")
-> goal satisfied direction-specified()=D
-> go menu: select-item("go")
-> goal satisfied is-moving(J)
-> [as joint moves through XYZ] complete operation
      stop-at(J,XYZ)
-> goal satisfied joint-at(J,XYZ)

```

This simple means-ends analysis story says nothing at all about setting and changing speeds. In terms of simply achieving user goals, there is no reason to ever change speeds. As noted in the original analysis, the user is likely to specify a higher speed to get things done faster, then a slower speed to improve accuracy. This means that the user must know what the current speed setting is, in order to know what does constitute faster or slower.

It is worth noting that for the user to know exactly when to press the STOP button they have to know about the response time of the robotic arm (or the specified joint) at the current speed. Another point is that if the joint over-shoots the target position, there is no easy "reverse direction" option. The user has to go right back to the beginning and re-specify that joint and the new direction, which will be opposite to the previous one.

### **The Second Analysis**

The second analysis of the robotic arm using PUM incorporated the input devices of the gesture system and voice recognition software. This proved to have little obvious impact on the analysis.

For example the operation stop arm was replaced by two almost identical operations:

```

operation stop-by voice (joint: J, location: L)
user-purpose: joint-at (J, L)
precondition: direction-specified () = D
              is-moving (J)

```



```
        joint-specified () = J
filter: is-within-range (J, L)
        can-move (J, D)
        L is in direction D from current position
        use of gesture device inappropriate
action: wait then "stop"
```

```
operation stop-by gesture (joint: J, location: L)
user-purpose: joint-at (J, L)
precondition: direction-specified () = D
        is-moving (J)
        joint-specified () = J
filter: is-within-range (J, L)
        can-move (J, D)
        L is in direction D from current position
        use of voice device inappropriate
action: wait then "stop"
```

The filters describing that the use of the gesture or the voice devices was inappropriate could be expanded to show the exact reasoning behind the choice of one input device over another.

These operations clearly shows the trivial nature of the alterations to the IL, and that the only changes were in filters rather than pre-conditions. This change makes little difference to the initial analysis. It is not clear as to if this will always be a filter or if it may be a pre-condition, or in other cases not be necessary at all, since some users may only be able to use one kind of input device, thus making the choice redundant. There is no indication of how a user might make such a choice.

## **PUM Conclusions**

The first PUM analysis, and the difficulties of doing such an analysis, highlighted some very important usability issues.

The strong divergence between the user's domain task and the actual device task was clearly identified almost immediately. This raised some interesting points. The device operates in three dimensions, and there is a clear difference between how the device works and how a user might manipulate something in three dimensions. Another point identified was that with this interface the goals of choosing joint, direction and speed are to be satisfied in sequence, whereas in reality, a user might want to change the order of such goals, or satisfy one without having to go back through all the other goals to reach it. This leads into the problem of the lack of a way of easily changing direction in order to accommodate overshooting, which again raises another problem of how a disabled user is likely to be able to respond with any degree of accuracy to 3D positioning. This in turns points out the need to know which speed setting is in operation, and the need to be able to make judgements as to what faster and slower actually mean in context.

Other important points were raised, such as the difficulty of distinguishing between **Move** and **Movearm**, possible confusion over joint names, and that directions for the joints will vary according to context. Finally, the meaning of the “end” command was ambiguous, in that not only did it provide an escape option back to the main menu, but it was also the way of ending the interaction once the task had been completed, thus combining two separate functions.

Therefore the PUM analysis, although problematic to apply, uncovered a high number of usability problems, and provided reasoning as to why those problems were important.

However, the second analysis which involved the input devices proved less illuminating. The alterations made to the original analysis were of a trivial nature and gave no insight into potential multi-modal usability issues. The increased complexity of interaction when two input devices are available and the user can use one or other or both was not clearly shown, and PUM could not allow for timing issues. The lack of major change in either content or emphasis of the analysis meant that other potential multi-modal issues were not addressed.



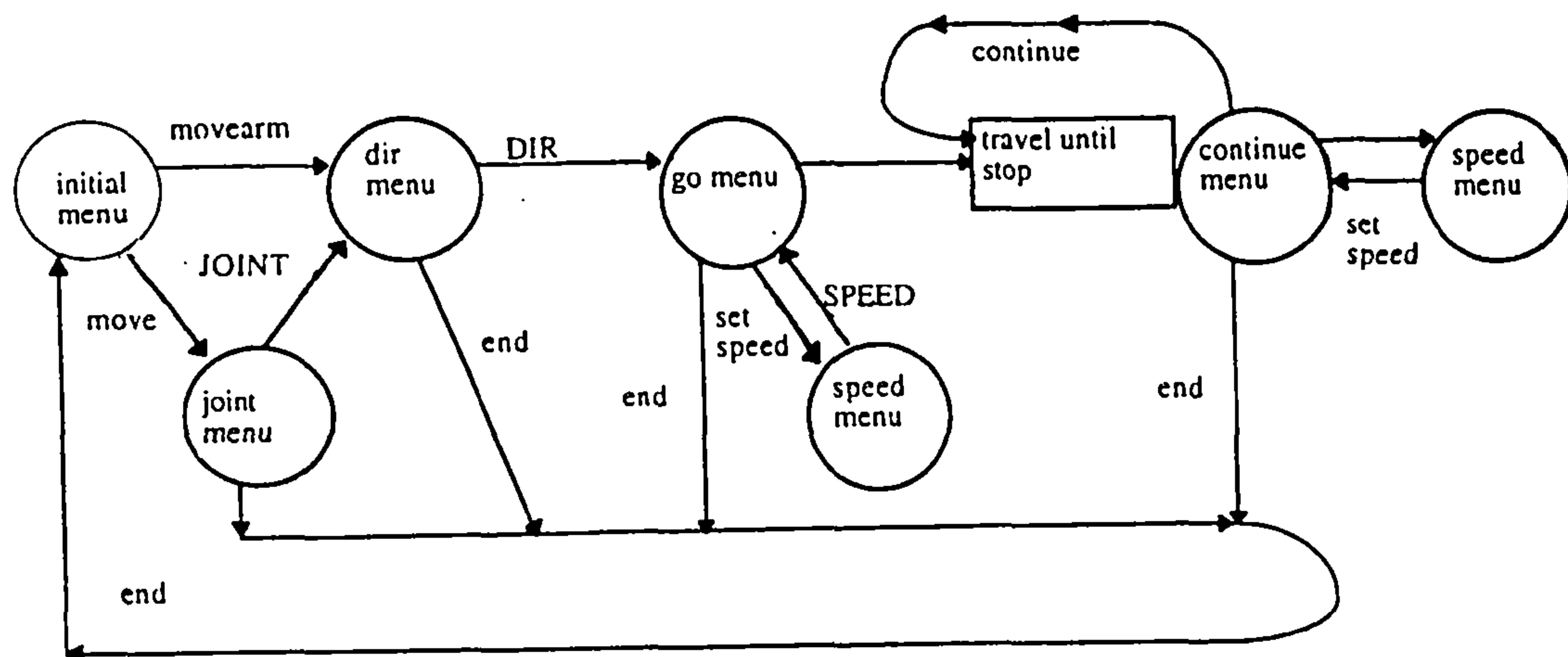
## 5.4 The STN Analysis

### Introduction

The interface of the robotic arm was initially analysed using STN when the input devices were undetermined. A second analysis was then done after some changes had been made to the structure of the interface, and the input devices determined. The interface was found to be particularly suited to STN analyses due to the small number of states and user-actions, and the relative simplicity of that portion of the interface under investigation.

### First Analysis

From the information provided by B. Parsons and observations of the working of the system, a State Transition Network was drawn as follows:



By explicitly representing the interaction choices in a diagrammatic form, it was straightforward to follow through a path of interaction corresponding to a given task, and to reason about the usability issues raised. Even though the diagram is not particularly complex, and the portion of the interface under investigation fairly small, several usability issues immediately become apparent.

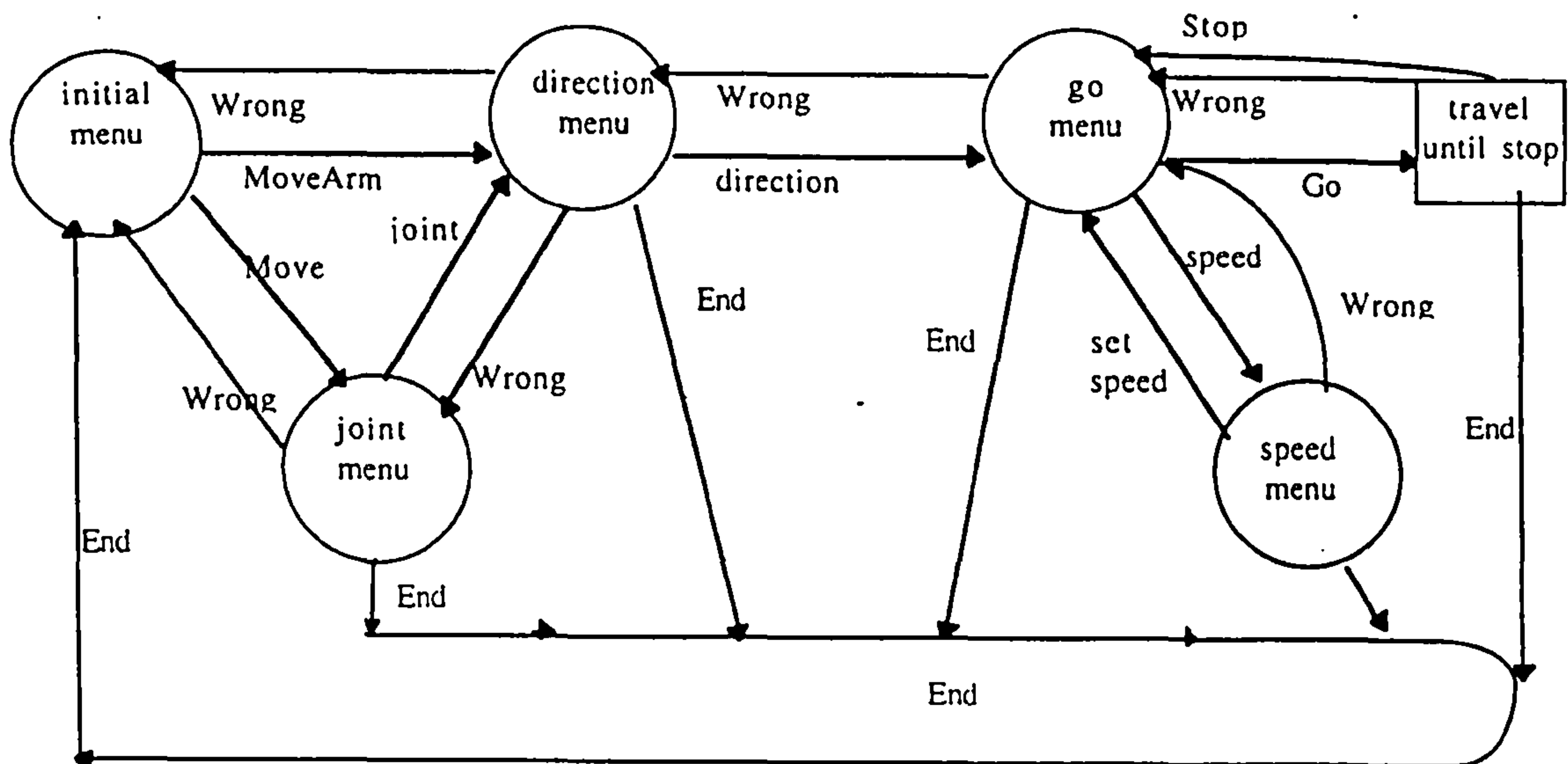
The main issue raised was that of the possible redundancy of the **Continue** option, which appears to add little to the functionality of the arm and instead creates more states for a user to navigate through, thus potentially making the system more confusing and longer to operate. Arguably, if the **Continue** menu also contained an option to change the direction of movement of the arm or joint, then this separation could be justified, but at present this is not the case, and if **Continue** was to have a change direction option, then so should **Go**, which would still make the **Continue** operation redundant.

The lack of any means of returning to the **direction** menu without going back to the original state was another area of concern, because it means that the user has to repeat a large section of interaction, whereas ideally the user would have more immediate access to make such a change.

The similarity between the steps of moving a joint and that of moving the whole arm were made very apparent by the diagram, which, allowing for the direction menu to be context specific, showed that the steps taken were in fact identical. This provides a strong argument towards combining the two different options into one, making the interface sequence cleaner and preventing an artificial distinction between two different types of movement.

### Second Analysis

A second analysis of the interface was done once the choice and means of input had been determined, and a new STN diagram was drawn (see below). It was apparent that very little on the diagram had changed, and that those changes that had been made had little reference to the multi-modal aspects of the interface. Therefore, information about multi-modal issues could not be determined from the diagram. This may be because STN diagrams represent the interaction states rather than the actual process or means of interaction, and cannot explicitly represent issues of time or simultaneous interaction.





## **STN Conclusions**

The STN diagram, therefore, despite being a very simple representation of the interface, nevertheless allowed important usability issues relating to the sequencing of operations and the possible redundancy of operations to become apparent. However, more complex usability issues were not readily made apparent through using this representation. The STN does not provide any inbuilt usability criteria, but depends upon someone analysing the representation in order to understand any possible usability issues. There is no in-built justification for the raising of such issues, other than that simplicity of flow and non-redundancy of operations is generally considered to be important in the design of successful interfaces. Therefore it is dependent more upon the skill of the analyst to recognise the importance of what is shown.

With regard to multi-modal usability issues, the diagram did not give any clear help or useful feedback to the analyst. This may be due to the nature of the diagram, which concentrates more on the states of the interaction rather than the process or means of interaction. The STN gave no basis for reasoning about important aspects of multi-modality such as modality selection, which may be based on many different factors such as social setting, user familiarity, ease of use, perceived times, and user capabilities. The STN cannot model such factors although they will be very important to the user of the robotic arm in deciding which input device to use to accomplish a given goal. Therefore the STN, although it was easy to use and apply, proved to be able to contribute little useful information regarding multi-modal usability.

## **5.5 The Z Analysis**

### **Introduction**

The analysis in Z of the robotic arm interface was done to see if a formal notation not normally associated with usability investigation could give any leverage on the interface issues. A specification of the interface was first produced without reference to the input devices. A second specification including the input devices was then produced in order to see what leverage Z could give on the multi-modal usability issues.

### **The specifications**

Both specifications abstract the operations represented in the interface in order that the operations can be examined without implementation detail, allowing the specifications to be examined to see what usability issues have become apparent. For a complete description of the specifications, please refer to **Appendix H**.

### **Usability Analysis of the First Specification**

There are several usability issues which the process of constructing the specification has made apparent. The most obvious is the similarity of the **Move** and **MoveArm**

commands, and how they could feasibly be combined into just one operation rather than kept as two separate ones. The advantages of combining the commands are that it would simplify the interface, and reduce confusion between what the two options do.

The specification showed plainly the similarity of the **Go** and **Continue** options. There seems to be little need for a separate **Continue** option, so it could be removed, thus simplifying the interface and reducing potential user confusion.

The fact that the interface is strongly ordered, in that there is no way for a user to backtrack and change the part of arm or direction selected without having to go back to the initial menu and starting the whole process again, was brought out by the specification.

The issues mentioned above were issues that other analyses of the robotic arm interface had already uncovered. However, the use of a Z specification, since it involves examining in detail each operation and what is allowable for any given state of the interface, did raise an issue which other analyses seem to have missed. This is that the interface offers options to the user that cannot in fact be carried out. For example, if the arm reaches its limit of movement, the interface displays the options **Continue**, **Speed Level**, **End**. Yet both **Continue** and **Speed Level** are irrelevant: **Continue** because the arm cannot move any further; and **Speed Level** because any speed selected now will not be kept when **End** is chosen and the interface returns to the initial menu.

### **Usability Analysis of the Second Specification**

The second specification differed from the first only in the addition of the input devices. This proved to have little effect on the usability analysis, since the new operations were merely distinguishing between the two available forms of input, and describing how to use them. There was no way of determining which might be the best choice from the Z specification, or how a user might determine which one to choose. Therefore the multi-modal aspects of the interaction could not be adequately described using Z.

### **Z Conclusions**

The specifications in Z of the portion of the interface under investigation helped to define several usability issues due its abstract nature. The **Move** and **MoveArm** commands, and **Go** and **Continue**, were shown to be sufficiently similar as to question the advisability of leaving them separate. The **Move** and **MoveArm** operations were compressed into one operation in this specification. Some states of the interface were found to be inapplicable in reality, and the problem of backtracking was also made apparent by the preconditions necessary to be satisfied before the interaction could continue. The first Z specification therefore provided an excellent means of examining the problem space and allowing usability issues to be examined, and also for postulating alternatives without having to re-implement the interface.



However, the multi-modal specification, where the two input devices were described, added little to the understanding of the usability of the interface due to the way that they were constructed. Z was unable to give insight into issues of choice and selection between the devices, and could not handle timing matters.

## **6. CONCLUSIONS**

The five analyses of the robotic arm have tested the scope of the different techniques, their limitations, and their ability to identify both general and multi-modal specific usability issues.

### **Ease of use and level of support**

All the analyses were done by the first author, with some input, when needed, from the second author. Thus it is possible to reflect on the learnability and usability of these techniques, based on this experience.

The five techniques were very different in terms of level of support for the analyst within the technique, and varied greatly in their ease of use. GOMS, PUM and CW, being overtly designed for usability assessment, had well defined procedures, whereas STN and Z, being used in software engineering for other purposes than evaluating interfaces, were less constrained. CW and STN were the most straightforward techniques to learn and apply, GOMS being slightly more difficult, and PUM and Z, as the most formal of the five, being the most difficult since they involved tight specification of the properties of the interface. GOMS, PUM and Z also involved the most training in order to learn the particular notations and means of application.

With CMN-GOMS, the notion of goals, operators, methods and selection rules meant that there was a clear structure to the analysis. The analysis could be performed in a structured and constrained manner. First, high-level goals could be determined, then the analysis taken down level by level until the final operators were determined. This was an iterative approach, but one which allowed for clarity and insight at all levels. However, there were problems at the CPM level, due to lack of concrete knowledge about the internal workings of the brain. At this level, the analyst had to use best judgement since various alternatives could be considered. Also, there was difficulty in determining the relative merits of one particular goal structure against another. This again was left to the judgement of the analyst. The main difficulty in learning this technique was in learning how and under what conditions to apply it, rather than any inherent level of difficulty.

CW seemed well-defined in that there was a clear set of questions, and that specific information about the users and task had to be provided from the beginning. The questions meant that the walkthrough was done in a structured manner and that at every stage searching questions about what the user was likely to do had to be answered. It could be done straight through in one go, avoiding lengthy iterations. However, the questions could be answered in many different ways, and quite often points could be

found which did not fit neatly into one question or another, but which were still necessary to the evaluation. The CW also relies on the skill of the analyst in answering the questions correctly with reference to the potential user, without necessarily knowing much about the potential user. Thus although it was a straightforward technique to learn, it was problematical to apply, in that the questioner needed experience in order to determine how and when to ask the questions.

PUM used the highly structured and constrained IL as a means of assessing the interface, which helped the analysis be consistent, accurate and complete. Having to specify prior knowledge, and what could and could not be done meant that the interface could be comprehensively covered. However, there was often confusion about the difference between a filter and a pre-condition, and having to define everything meant that the IL could be difficult to write. Admittedly, the IL allowed for imprecise definitions, but even with this the technique was less than straightforward. There was also difficulty in actually starting the analysis: where to start was a problem. This technique involved many iterations, with the analysis only becoming clear towards the end, however, the effort did mean that many insights could be obtained during the process. The particular notation also took time to learn.

STN, given that it was a diagrammatic approach meant for other uses within software engineering, had little overt support for the analyst. However, the structure of the diagrams with nodes of states and arcs of operations meant that it was straightforward to learn, easy to apply, and quick to assess. Only one or two iterations were required before a complete representation was obtained.

Being a specification notation, Z had no overt support for usability assessment. However, it resembled the PUM IL, in that every property had to be specified. It suffered from the same kinds of problems as PUM, in that precise definitions were not always easy to determine, the analysis took several iterations before it was complete, and there was difficulty in deciding where to start the analysis. The notation, based on set theory and first-order predicate logic, and using schemas to represent the states of the system, also required a certain amount of training.

### **General scope of techniques**

The five techniques differed in the kind of general usability issues uncovered, and how well they could deal with various forms of user behaviour, as can be seen by the following table which sets out the usability issues uncovered by each technique.

The use of CMN-GOMS allowed the interface to be examined in terms of how well structured it was, the naming of options, and identified problems relating to the length of interaction, the redundancy of operations, and the lack of short-cuts. CMN-GOMS therefore was able to examine and identify a wide range of potential usability issues, and, using CMN-GOMS, an alternative structure could be proposed. However, despite this, CMN-GOMS suffered from being unable to represent behaviour other than that of expert



users, and was unable to deal with failure, or users incorrectly using the interface to achieve a goal.

The CW acknowledged the existence of failure explicitly, but did not follow through as to the consequences of such failure, merely stating that failure could possibly occur at whatever point and giving reasons why. Another drawback was that the interaction was considered stage by stage rather than as a whole, thereby possibly missing certain issues relevant to the interface overall. The CW was best able to deal with issues of misleading names of options, and the importance of giving feedback information to the user, since the technique is aimed at emulating the behaviour of novice users, to whom such issues are important. Thus the technique was not as wide-ranging as GOMS, although it did deal with some different issues.

PUM was found to cover a large range of potential usability problems. Amongst other issues, the analysis made clear the lack of mapping between what the user wanted to accomplish, and how the device would attempt to accomplish it. However, again, user-failure was an issue insufficiently covered, and the analysis assumed consistent and expert use by people with exceptional memories, which might not be a typical user's profile.

The STN, despite being the quickest and easiest of the techniques to learn and apply, did allow various usability issues to be uncovered, the most obvious being that of the redundancy of some options, and the possibly misleading nature of some names. However, the diagram made no reference to the level of skill or knowledge of the user, being device oriented, and thus was unable to cover a wider range of issues, including failure.

The Z specification, being mainly device oriented with little explicit representation of the user, was unable to cover a wide range of issues, but was able to shed light on the problem of ordering, the inability of the interface to allow backtracking, and the existence of some actually unavailable choices. It was also able to show how operations might be more effectively and simply represented.

The table on the next page gives a brief summary of the kinds of problems that each technique was able to uncover. Both CMN- and CPM-GOMS are included under the general column GOMS.

### **Multi-modal scope of techniques**

With regard to multi-modal issues, such as the possible conflicts between voice and gesture input, the five techniques were found to be less than adequate. Admittedly, this is just as much due to the lack of information about properties of multi-modal interaction than the limitations of the individual techniques, but given the increasing complexity of interfaces this was seen as a cause for concern, since such properties would only become apparent through expensive empirical testing.

Other techniques such as syndetics [Duke *et al*, in submission] and Cognitive Task Analysis [May and Barnard, 1994] were not used, since there is little or no published information on their application to interface analysis.

**Summary table of usability problems**

PROBLEM	GOM S	PUM	CW	STN	Z
long sequence of operators to move arm	X	-	X	-	-
inability to backtrack	X	X	-	X	X
difficulty of choosing between Move Arm or Move	X	X	X	-	-
Lack of short cuts	X	-	-	-	-
Continue versus Go: Continue seen as redundant	X	-	-	X	X
Confusion over joint called Arm	X	X	X	-	-
Gesture input with twice as many operations as voice because dependent on cursor movement	X	-	-	-	-
problem if head moved to look at arm while gesture system operational may be interpreted as a command	-	-	X	-	-
if user pauses in middle of saying "Move arm"...	-	-	X	-	-
if user engaged in conversation...	-	-	X	-	-
lack of feedback after Move Arm selected, no indication that whole arm is to be moved	-	-	X	-	-
problems of determining left and right, especially when arm contorted	-	-	X	-	-
user cannot check direction choice until arm starts to move	-	-	X	-	-
time taken to interact with system to stop arm	-	-	X	-	-
similarity between moving joint and moving whole arm	X	-	-	X	X
illegal options	-	-	-	-	X
mismatch between way that arm works and way that user would move arm	-	X	-	-	-
Not clear that End returns user to main menu	-	X	X	-	-
End having two meanings	-	X	X	-	-



GOMS seemed initially to be best able to handle multi-modal interaction issues, since the CPM-GOMS level looks at the different motor, perceptual and cognitive activities of the user, and also attempts to address the issue of time to a limited extent. However, CPM-GOMS suffers from a lack of complete knowledge about how humans process information, and was unable to shed much light on the multi-modal interaction process, although it was able to give comparative information about the two methods of input. This is also true of CMN-GOMS. CW was unable to give much insight into multi-modal usability issues because of the technique's emphasis on goal-structure and user knowledge rather than actual means and method of interaction. Thus it was unlikely to uncover multi-modal issues such as timing and complexity.

PUM was unable to shed any light on multi-modal usability issues since it dealt with interaction in a procedural and knowledge-based manner. There was no way of showing complexity of interaction or of defining the decision process between two choices of input device, except at the most basic of levels.

The STN concentrated on the states and stages of the interaction more than the process or means of interaction, and therefore was unable to give any useful information regarding multi-modal usability issues.

Z was similar to PUM in its inability to handle multi-modal interaction except at a trivial level, since it is based on states of operation rather than wider issues involving choice and complexity. However, since Z is not designed as a method for analysing any usability issues, let alone those of a multi-modal nature, this is hardly surprising.

All of these techniques were goal-based to some extent, and were concerned more with the ordering of the interaction and the knowledge possessed by the user rather than the nature of the interaction. Thus their limitations with regard to multi-modal usability issues centered around the same issues, an emphasis on interaction ordering and being unable to deal with simultaneous complexity. This possibly implies that multi-modal interaction cannot be explained or described in these terms and that notations which use different interaction paradigms may be more appropriate for describing such systems.

### **Final conclusions**

Any report of this nature suffers from certain limitations. The work reported here could be approached in a number of different ways, and the conclusions refined and developed appropriately. Since all five analyses were done mainly by one person, previously inexperienced in such techniques, it may be that other, more experienced analysts may have obtained different results and drawn alternative conclusions as a result.

However, this report is of interest because by using just one analyst, the weaknesses and strengths of the five techniques can be assessed in comparison to each other, not just in terms of the usability issues uncovered, but also in terms of their learnability and

usability. For a technique to be useful and to be exploited by designers of interactive systems, it must be clearly defined and straightforward to learn and apply. No matter how powerful a technique might be, if it is difficult to learn and time-consuming in application, designers will prefer to use other methods. Of the five techniques assessed, CW and STN were the most straightforward to apply, with GOMS next, and PUM and Z hampered by the need to learn a specific notation and the degree of formalism necessary.

The five techniques all to a greater or lesser extent relied on the skill of the analyst in determining usability problems. Z and STN were obviously the most heavily dependent on the analyst having relevant insights, in that they are techniques not normally applied to usability evaluation. However, GOMS, CW and PUM also relied significantly on the ability of the analyst to identify areas of potential concern. This demonstrates that even techniques which purport to support the analyst and acknowledge user behaviour have limitations in their application and are dependent to some extent on the analyst's ability.

A technique in order to be widely acceptable has to be seen to cover a wide range of usability issues, in order to justify the time spent in applying it. The five techniques between them covered a wide spectrum of usability issues. It is always problematic to determine which usability problems are of more importance than others, but even allowing for a subjective approach, certain points are apparent. GOMS, CW and PUM were the most successful techniques in this particular set of analyses, in that they uncovered the highest number of usability issues, which is not surprising since they are techniques specifically designed for this purpose. What was interesting was that both the STN and Z could be used to successfully reason about some aspects of the interaction.

The issue of multi-modal usability added another dimension to the analysis of the techniques considered here. Such issues could not be examined in any depth by the five techniques for a variety of reasons. Firstly, multi-modal usability is an area as yet not fully defined. Secondly, the techniques emphasised the state nature of the interface, and had more emphasis on goals and knowledge to achieve interaction than the actual process of interaction itself, which may be of more importance in multi-modal systems. Thirdly, the interface examined was multi-modal only in that it had two means of interaction, rather than that both means were likely to be used at the same time, although that possibility was allowed. None of the techniques could represent important issues affecting modality selection such as social setting, user familiarity, ease of use, perceived times, and user capabilities.

This report has clearly shown how existing modelling notations are useful in describing interfaces, and that they can uncover a large number of problems. It was hardly surprising that techniques designed for usability analysis were more useful than other notations designed for other purposes, but all representations allowed usability issues to be assessed to some extent. It is interesting that both Z and STN allowed meaningful analysis of usability issues, when they are techniques not normally used in the area of usability evaluation. This shows that all modelling techniques can give some insight into usability problems, because they allow the interface to be clearly described. Determining the best



overall technique from those used in these analyses was difficult, since it depended on several factors, but it seemed that GOMS and CW gave the most insights. However, with regard to multi-modal usability, only CPM-GOMS was at all useful, and even then only at a minor level. Therefore there seems to be a need both for techniques which can deal with multi-modal usability issues, and further research into the nature of multi-modal interaction itself.

## 7. BIBLIOGRAPHY

- ANZAI, Y (1994) Human-Robot-Computer Interaction: A New Paradigm of Research in Robotics. *Advanced Robotics VIII*. 4, 357-369.
- BARNARD, P. J. & MAY, J. (1993) Cognitive Modelling for User Requirements. In P.F.Byerley, P.J.Barnard and J.May, Eds *Computers, Communication and Usability: Design Issues, Research and Methods for Integrated Services*. Elsevier Science Publishers B.V.
- BARNARD, P. J. & HARRISON M. D. (1989): Integrating Cognitive and System Models in Human Computer Interaction. In A.Sutcliffe, L.Macauley, Eds *People and Computers V* Cambridge University Press 87-103
- BLANDFORD, A. E. & YOUNG, R. M. (1996) Specifying user knowledge for the design of interactive systems. *Software Engineering Journal*. 11.6, 323-333.
- CARD, S. K., MORAN, T. P. AND NEWELL, A. (1983). *The Psychology of Human Computer Interaction*, Hillsdale : Lawrence Erlbaum.
- DILLER, A. (1994) *Z: An Introduction to Formal Methods*, John Wiley and Sons Ltd, Chichester
- DIX, A., FINLAY, J., ABOWD, G. & BEALE R. (1993) *Human-Computer Interaction*, Prentice Hall International
- DUKE, D. J. & HARRISON, M. D. (1995): Event model of human-system interaction. *Software Engineering Journal*, January 1995
- DUKE, D. J., BARNARD, P. J., DUCE, D.A. & MAY, J. (in submission) Syndetic Modelling. Submitted to *Human Computer Interaction*
- GRAY, W. D., JOHN, B. E., & ATWOOD, M. E. (1993) Project Ernestine: Validating a GOMS Analysis for Predicting and Explaining Real-World Task Performance. *Human-Computer Interaction VIII* 237-309
- HINCHEY, M. G. & AND BOWEN, J. P. (1995): Applications of Formal Methods FAQ. In M.G.Hinchey and J.P.Bowen Eds *Applications of Formal Methods* Prentice Hall International (UK)
- HYDE, J.K. & DUKE, D.J. (1996) Fusion Engines and Melting Pots. In F.Bodart and J.Vanderdonckt, Ed. *Design, Specification and Verification of Interactive Systems '96*. Springer-Verlag/Wien



- JOHN, B. & KIERAS, D. (1996) The GOMS family of user interface analysis techniques: comparison and contrast. *ACM Transactions on CHI.* , 3, 320-351.
- JOHN, B. & KIERAS, D. (1994) The GOMS Family of Analysis Techniques: Tools for Design and Evaluation. In *Human-Computer Interaction Institute Technical Report CMU-HCII-94-106*
- JOHN, B. & KIERAS, D. (1996): Using GOMS for User Interface Design and Evaluation: Which Technique? To appear in *ACM Transactions on Computer-Human Interaction*. Available from <http://ai.eecs.umich.edu/people/kieras/musette.html>
- JOHN, B. & VERA, A. (1992): A GOMS Analysis of a Graphic, Machine-Paced, Highly Interactive Task *CHI'92*, May 3-7, 251-258
- KIERAS, D. (1994): A Guide to GOMS Task Analysis. Unpublished paper available from <http://ai.eecs.umich.edu/people/kieras/musette.html>
- MACK, R. L. & NIELSEN, J. (1994): Usability Inspection Methods: Executive Summary. In R.L.Mack, J.Nielsen, eds. *Usability Inspection Methods* John Wiley and Sons
- MAY, J. & BARNARD, P. J. (1995) The Case for Supportive Evaluation During Design. In *Interacting with Computers VII*, 2, 115-143
- MAY, J. & BARNARD, P. J. (1994) A Cognitive Task Analysis of the EuroCODE Exemplar Material . AMODEUS2 Project Document UM/WP22
- PARSONS, B., PRIOR, S. D & WARNER, P. R. (1995) An Approach to Designing Manipulator Controller Software Employing Context Dependent Command Interpretation, *European Conference on Assistive and Rehabilitation Technology*, Lisbon, October 1995
- POLSON, P., LEWIS, C., RIEMAN, J., & WHARTON, C. (1992) Cognitive Walkthroughs: a method for theory-based evaluation of user interfaces. In *International Journal of Man-Machine Studies*
- PREECE, J., ROGERS, Y., SHARP, H., BENYON, D., HOLLAND, S., & CAREY, T. (1994) *Human Computer Interaction*, Addison-Wesley UK
- RANN, D., TURNER, J. & WHITWORTH, J. (1994) *Z: A Beginner's Guide*, Chapman and Hall, London
- SPIVEY, J.M. (1989). *The Z Notation: A Reference Manual*, Prentice-Hall International.

WHARTON, C., RIEMAN, J., LEWIS, C. & POLSON, P. (1994). *The Cognitive Walkthrough Method: A Practitioner's Guide*. In J. Nielsen and R. Mack, Ed. *Usability Inspection Methods*, 105-140. Wiley: New York.



## 8. APPENDICES

### Appendix A: The Robotic Arm

At present there is no feedback to the user.

Device task:

Move gripper to position XYZ from rest, with no pre-taught position.

User task:

Turn on light switch.

When arm is at rest, there is an initial menu. There is a choice for the user between **MoveArm**, which governs the movement of the arm as a whole, and **Move**, which governs the movement of individual joints.

So to move the arm the user must chose between **Move** and **MoveArm**.

If **MoveArm** is chosen, since it may be easier to move whole arm, a choice of further options will appear:

**Left, Right, Up, Down, Forward, Back, End**

**End** will return the user to the initial at rest menu. The others will select which direction (in relation to the user) the arm will move in.

If a direction such as **Up** is chosen, a choice of further options appears:

**Go, Speed Level, End**

**Go** will start the arm moving at medium speed until **Stop**, the option that appears if **Go** is chosen is selected, or the arm reaches the limit of movement.

**Speed Level** will change the speed of the arm to either **Min, Slow, Medium, Fast** or **Max**.

**End** will return the user to the initial at rest menu.

If **Go** is chosen, the arm will continue moving until **Stop** is selected, or the arm reaches the limit of movement. When **Stop** is selected, or when the arm reaches the limit of movement, a further choice of options appears:

**Continue, Speed Level, End**

**Continue** will allow the arm to continue moving at the preset speed until **Stop**, the option that appears if **Continue** is chosen is selected, or the arm reaches the limit of movement. When **Stop** is selected, or when the arm reaches the limit of movement, the options return to:

**Continue, Speed Level, End**

and the user can cycle through these until the arm is in the correct position in that direction and chooses **End** to return to the main menu in order to choose a new direction.

If **Speed Level** is selected, there is a choice of speeds given as listed above, and once the user has chosen the new speed, the menu returns to the options:

**Continue, Speed Level, End.**

When the arm is in the right position in one direction, the user then selects **End** to return to the initial at rest menu, and then has the choice of selecting **MoveArm** and choosing a new direction for the arm to move in, and repeating the choices listed above, or selecting **Move**, which can move each individual joint of the arm.

If **Move** is selected, the following joint options appear:

**Base, Arm, Shoulder, Elbow, Hand, Wrist, Palm, Grip, End**

If **Shoulder** was selected, the following options would appear:

**In, Out, End**

Where **In** and **Out** refer to the direction of movement of the joint (in the case of **Shoulder** moving either **In** towards the user on a horizontal plane of movement or **Out** away from the user on the horizontal plane), and **End** returns the user to the initial at rest menu.

If **In** was selected, the following options would appear:

**Go, Speed Level, End**

These are exactly the same as in the **MoveArm** section, and the options are now on exactly the same as from this option level in the **MoveArm** section, so are therefore not listed.

The user can cycle through these two paths until the arm is in the correct position.



## Appendix B: The Multi-Modal Robotic Arm

Device task:

Move gripper to position XYZ from rest, with no pre-taught position.

User task:

Turn on light switch.

There are two methods of input, which can be used concurrently or alternately.

A gesture system is based on a baseball cap with two sensors: one allowing movement forwards and backwards to be detected, the other allowing movement left and right to be detected. This allows a variety of unique gestures to form the gesture vocabulary. The way that the gesture system is presently implemented, a cursor moves along underneath the menu options continuously in turn, and if the correct gesture is made when the cursor is underneath a particular option, then that option is selected. Another gesture acts as a toggle between high and low speed of the cursor. A final gesture is an escape option, which automatically stops the arm if the arm is moving, and returns the user to the main menu.

The voice recognition system allows direct menu option selection simply by saying the menu option out loud. It is designed to be trained to individual voices, and needs resetting over time due to the way that voices can change.

The use of two different means of input allows for discreet use in social situations. Also, if one method is not recognised by the computer, for instance because the gesture or word was imperfectly formed, then the alternative mode of input can be used.

If both means of input are used at the same time but for conflicting inputs, then it would depend on which input was processed first as to which input was selected first.

The output is not yet implemented, but will take the form of a small LCD display attached to the base of the arm showing the menus.

When arm is at rest, there is an initial menu. There is a choice for the user between **MoveArm**, which governs the movement of the arm as a whole, and **Move**, which governs the movement of individual joints.

So to move the arm the user must chose between **Move** and **MoveArm**.

If **MoveArm** is chosen, since it may be easier to move whole arm, a choice of further options will appear:

### **Left, Right, Up, Down, Forward, Back, Wrong, End**

**Wrong** will return the user to the menu at the level above, in effect saying that the wrong option was chosen and allowing the user to backtrack.

**End** will return the user to the initial at rest menu. The others will select which direction (in relation to the user) the arm will move in.

If a direction such as **Up** is chosen, a choice of further options appears:

### **Go, Speed, Wrong, End**

**Go** will start the arm moving at medium speed until **Stop**, the option that appears if **Go** is chosen is selected, or the arm reaches the limit of movement.

**Speed** will give a menu of **Min, Slow, Medium, Fast, Max, Wrong, End**.

**Wrong** will return the user to the menu above.

**End** will return the user to the initial at rest menu.

The other options allow the user to set the speed of the arm, which, once selected, will return to the **Go, Speed, Wrong, End** menu.

If **Go** is chosen, the arm will continue moving until **Stop** is selected, or the arm reaches the limit of movement. When **Stop** is selected, which is necessary even if the arm reaches the limit of movement, i.e. the user cannot progress with the interaction until **Stop** is selected, a further choice of options appears:

### **Go, Speed, Wrong, End**

**Go** will allow the arm to continue moving at the preset speed until **Stop**, the option that appears if **Go** is chosen is selected, or the arm reaches the limit of movement. When **Stop** is selected, or when the arm reaches the limit of movement, the options return to:

### **Go, Speed, Wrong, End**

and the user can cycle through these until the arm is in the correct position in that direction and chooses **End** to return to the main menu in order to choose a new direction.

If **Speed** is selected, there is a choice of speeds given as listed above, and once the user has chosen the new speed, the menu returns to the options:

### **Go, Speed, Wrong, End.**



When the arm is in the right position in one direction, the user then selects **End** to return to the initial at rest menu, and then has the choice of selecting **MoveArm** and choosing a new direction for the arm to move in, and repeating the choices listed above, or selecting **Move**, which can move each individual joint of the arm.

If **Move** is selected, the following joint options appear:

**Base, Arm, Shoulder, Elbow, Hand, Wrist, Palm, Grip, Wrong, End**

**Wrong** returns the user to the menu layer above, in this case the main menu.

**End** returns the user to the main menu.

If **Shoulder** was selected, the following options would appear:

**In, Out, Wrong, End**

Where **In** and **Out** refer to the direction of movement of the joint (in the case of **Shoulder** moving either **In** towards the user on a horizontal plane of movement or **Out** away from the user on the horizontal plane). **Wrong** returns the user to the joint menu, and **End** returns the user to the initial at rest menu.

If **In** was selected, the following options would appear:

**Go, Speed, Wrong, End**

**Go** will start the arm moving at medium speed until **Stop**, the option that appears if **Go** is chosen is selected, or the arm reaches the limit of movement.

**Speed** will give a menu of **Min, Slow, Medium, Fast, Max, Wrong, End**.

**Wrong** will return the user to the menu above.

**End** will return the user to the initial at rest menu.

The other options allow the user to set the speed of the arm, which, once selected, will return to the **Go, Speed, Wrong, End** menu.

If **Go** is chosen, the arm will continue moving until **Stop** is selected, or the arm reaches the limit of movement. When **Stop** is selected, which is necessary even if the arm reaches the limit of movement, i.e. the user cannot progress with the interaction until **Stop** is selected, a further choice of options appears:

**Go, Speed, Wrong, End**

**Go** will allow the arm to continue moving at the preset speed until **Stop**, the option that appears if **Go** is chosen is selected, or the arm reaches the limit of movement. When **Stop** is selected, or when the arm reaches the limit of movement, the options return to:

**Go, Speed, Wrong, End**

and the user can cycle through these until the arm is in the correct position in that direction and chooses **End** to return to the main menu in order to choose a new direction.

If **Speed** is selected, there is a choice of speeds given as listed above, and once the user has chosen the new speed, the menu returns to the options:

**Go, Speed, Wrong, End.**

These are exactly the same as in the **MoveArm** section. The user can cycle through these two paths until the arm is in the correct position.



## Appendix C: CMN-GOMS Description of Device

### GOAL: MOVE ARM TO POSITION XYZ

```

*   DETERMINE CURRENT POSITION OF ARM
*   GOAL: MOVE WHOLE ARM (repeat until arm near position XYZ)
*   *   REMEMBER MEANINGS OF MOVE AND MOVE ARM OPTIONS
*   *   [SELECT: GOAL: USE MOVE...if arm is in right area but not
extended
*   *   appropriately
*   *   GOAL: USE MOVE ARM...if arm is not in right
area]
*   *   GOAL: USE MOVE
*   *   *   PRESS MOVE OPTION
*   *   *   GOAL: SELECT JOINT
*   *   *   *   REMEMBER MEANINGS OF JOINT OPTIONS AND
END
*   *   *   *   OPTION
*   *   *   *   [SELECT: GOAL: USE BASE...if base needs moving
*   *   *   *   GOAL: USE ARM...if arm needs
moving
*   *   *   *   GOAL: USE SHOULDER...if
shoulder needs moving
*   *   *   *   GOAL: USE ELBOW...if elbow
needs moving
*   *   *   *   GOAL: USE HAND...if hand needs
moving
*   *   *   *   GOAL: USE WRIST...if wrist needs
moving
*   *   *   *   GOAL: USE PALM...if palm needs
moving
*   *   *   *   GOAL: USE GRIPPER...if gripper
needs moving
*   *   *   *   GOAL: END OPERATION...if none
need moving]
*   *   *   *   GOAL: USE BASE
*   *   *   *   *   PRESS BASE OPTION
*   *   *   *   GOAL: USE ARM
*   *   *   *   *   PRESS ARM OPTION
*   *   *   *   GOAL: USE SHOULDER
*   *   *   *   *   PRESS SHOULDER OPTION
*   *   *   *   GOAL: USE ELBOW
*   *   *   *   *   PRESS ELBOW OPTION
*   *   *   *   GOAL: USE HAND
*   *   *   *   *   PRESS HAND OPTION
*   *   *   *   GOAL: USE WRIST

```





* * * *	* * *	[SELECT: GOAL: USE MIN...if
slowest speed		
* * * *	* * *	required
* * * *	* * *	GOAL: USE
SLOW...if slow speed		
* * * *	* * *	required
* * * *	* * *	GOAL: USE
MEDIUM...if medium		
* * * *	* * *	speed required
* * * *	* * *	GOAL: USE
FAST...if fast speed		
* * * *	* * *	required
* * * *	* * *	GOAL: USE
MAX...if fastest speed		
* * * *	* * *	required]
* * * *	* * *	PRESS OPTION GO
* * * *	* * *	GOAL: USE MIN
* * * *	* * *	* PRESS OPTION MIN
* * * *	* * *	GOAL: USE SLOW
* * * *	* * *	* PRESS OPTION SLOW
* * * *	* * *	GOAL: USE MEDIUM
* * * *	* * *	* PRESS OPTION MEDIUM
* * * *	* * *	GOAL: USE FAST
* * * *	* * *	* PRESS OPTION FAST
* * * *	* * *	GOAL: USE MAX
* * * *	* * *	* PRESS OPTION MAX
* * * *		GOAL: FINISH MOVING JOINT(repeat until joint in
correct		
* * * *	* * *	position)
* * * *	* * *	GOAL: SELECT STOP MOVEMENT
* * * *	* * *	[SELECT: GOAL: USE STOP...if arm
will otherwise		
* * * *	* * *	move beyond correct
position		
* * * *	* * *	GOAL: NOT USE
STOP...if arm will not		
* * * *	* * *	move beyond correct
position]		
* * * *	* * *	GOAL: USE STOP
* * * *	* * *	* PRESS STOP OPTION WHEN
ARM IN CORRECT		
* * * *	* * *	POSITION
* * * *	* * *	GOAL: USE NOT STOP
* * * *	* * *	* WAIT UNTIL ARM HAS
FINISHED MOVING		





* * *	* [SELECT: GOAL: USE LEFT...if position XYZ is to
left of	
* * *	current arm position
* * *	GOAL: USE RIGHT...if position
XYZ is to right of	
* * *	current arm position
* * *	GOAL: USE UP...if position XYZ is
above current	
* * *	arm position
* * *	GOAL: USE DOWN...if position
XYZ is below	
* * *	current arm position
* * *	GOAL: USE FORWARD...if
position XYZ if in front	
* * *	of current arm position
* * *	GOAL: USE BACK...if position
XYZ is behind	
* * *	current arm position
* * *	GOAL: END OPERATION...if arm
already in good	
* * *	position]
* * *	GOAL: USE LEFT
* * *	* PRESS LEFT OPTION
* * *	GOAL: USE RIGHT
* * *	* PRESS RIGHT OPTION
* * *	GOAL: USE UP
* * *	* PRESS UP OPTION
* * *	GOAL: USE DOWN
* * *	* PRESS DOWN OPTION
* * *	GOAL: USE FORWARD
* * *	* PRESS FORWARD OPTION
* * *	GOAL: USE BACK
* * *	* PRESS BACK OPTION
* * *	GOAL: END OPERATION
* * *	* PRESS END OPERATION OPTION
* * *	GOAL: MOVING ARM
* * *	REMEMBER MEANINGS OF OPTIONS
* * *	GOAL: SELECT ARM MOVE
* * *	* REMEMBER MEANING OF OPTIONS
* * *	* [SELECT: GOAL: END OPERATION...if
operation to end	
* * *	GOAL: SELECT
MOVEMENT...if arm to be	
* * *	moved]
* * *	GOAL: END OPERATION

* * * * *	* PRESS END OPERATION OPTION
* * * * *	<b>GOAL: SELECT MOVEMENT</b>
* * * * *	* REMEMBER MEANINGS OF OPTIONS
* * * * *	* [SELECT: GOAL: USE GO...if speed at
correct level	
* * * * *	* GOAL: SET
SPEED...if speed not at correct	
* * * * *	* level]
* * * * *	* <b>GOAL: USE GO</b>
* * * * *	* PRESS GO OPTION
* * * * *	* <b>GOAL: SET SPEED</b>
* * * * *	* PRESS OPTION SPEED LEVEL
* * * * *	* [SELECT: GOAL: USE MIN...if
slowest speed	
* * * * *	* * required
* * * * *	* * GOAL: USE
SLOW...if slow speed	
* * * * *	* * required
* * * * *	* * GOAL: USE
MEDIUM...if medium	
* * * * *	* * speed required
* * * * *	* * GOAL: USE
FAST...if fast speed	
* * * * *	* * required
* * * * *	* * GOAL: USE
MAX...if fastest speed	
* * * * *	* * required]
* * * * *	* * PRESS OPTION GO
* * * * *	* * <b>GOAL: USE MIN</b>
* * * * *	* * PRESS OPTION MIN
* * * * *	* * <b>GOAL: USE SLOW</b>
* * * * *	* * PRESS OPTION SLOW
* * * * *	* * <b>GOAL: USE MEDIUM</b>
* * * * *	* * PRESS OPTION MEDIUM
* * * * *	* * <b>GOAL: USE FAST</b>
* * * * *	* * PRESS OPTION FAST
* * * * *	* * <b>GOAL: USE MAX</b>
* * * * *	* * PRESS OPTION MAX
* * * * *	* <b>GOAL: FINISH MOVING ARM</b> (repeat until arm in
correct	
* * * * *	* position)
* * * * *	* <b>GOAL: SELECT STOP MOVEMENT</b>
* * * * *	* [SELECT: GOAL: USE STOP...if arm
will otherwise	



* * * *	* *	move beyond correct
position		
* * * *	* *	GOAL: NOT USE
STOP...if arm will not		
* * * *	* *	move beyond correct
position]		
* * * *	* *	GOAL: USE STOP
* * * *	* *	* PRESS STOP OPTION WHEN
ARM IN CORRECT		
* * * *	* *	* POSITION
* * * *	* *	GOAL: USE NOT STOP
* * * *	* *	* WAIT UNTIL ARM HAS
FINISHED MOVING		
* * * *	* *	GOAL: SELECT ARM CONTINUE
* * * *	* *	* [SELECT: GOAL: USE CONTINUE...if
speed correct		
* * * *	* *	GOAL: SET
SPEED...if speed incorrect		
* * * *	* *	GOAL: END
OPERATION...if operation		
* * * *	* *	complete]
* * * *	* *	GOAL: USE CONTINUE
* * * *	* *	* PRESS CONTINUE OPTION
* * * *	* *	GOAL: END OPERATION
* * * *	* *	* PRESS END OPERATION
OPTION		
* * * *	* *	GOAL: SET SPEED
* * * *	* *	* PRESS OPTION SPEED LEVEL
* * * *	* *	* [SELECT: GOAL: USE MIN...if
slowest speed		
* * * *	* *	* required
* * * *	* *	* GOAL: USE
SLOW...if slow speed		
* * * *	* *	* required
* * * *	* *	* GOAL: USE
MEDIUM...if medium		
* * * *	* *	* speed required
* * * *	* *	* GOAL: USE
FAST...if fast speed		
* * * *	* *	* required
* * * *	* *	* GOAL: USE
MAX...if fastest speed		
* * * *	* *	* required]
* * * *	* *	* PRESS OPTION CONTINUE
* * * *	* *	* GOAL: USE MIN

*	*	*	*	*	*	*	* PRESS OPTION MIN
*	*	*	*	*	*	*	<b>GOAL: USE SLOW</b>
*	*	*	*	*	*	*	* PRESS OPTION SLOW
*	*	*	*	*	*	*	<b>GOAL: USE MEDIUM</b>
*	*	*	*	*	*	*	* PRESS OPTION MEDIUM
*	*	*	*	*	*	*	<b>GOAL: USE FAST</b>
*	*	*	*	*	*	*	* PRESS OPTION FAST
*	*	*	*	*	*	*	<b>GOAL: USE MAX</b>
*	*	*	*	*	*	*	* PRESS OPTION MAX



## Appendix D: Improved GOMS Description of Device

### GOAL: MOVE ARM TO POSITION XYZ

- \* DETERMINE CURRENT POSITION OF ARM
- \* GOAL: USE MOVE
- \* \* PRESS MOVE OPTION
- \* \* GOAL: SELECT JOINT/ARM
- \* \* \* REMEMBER MEANINGS OF OPTIONS AND END OPTION
- \* \* \* [SELECT: GOAL: USE BASE...if base joint needs moving
- \* \* \* GOAL: USE ARM...if arm joint needs
- moving
- \* \* \* GOAL: USE SHOULDER...if shoulder joint
- needs moving
- \* \* \* GOAL: USE ELBOW...if elbow joint needs
- moving
- \* \* \* GOAL: USE HAND...if hand joint needs
- moving
- \* \* \* GOAL: USE WRIST...if wrist joint needs
- moving
- \* \* \* GOAL: USE PALM...if palm joint needs
- moving
- \* \* \* GOAL: USE GRIPPER...if gripper joint
- needs moving
- \* \* \* GOAL: USE WHOLE ARM...if whole of
- arm needs
- \* \* \* moving
- \* \* \* GOAL: END OPERATION...if nothing
- needs moving]
- \* \* \* GOAL: USE BASE
- \* \* \* \* PRESS BASE OPTION
- \* \* \* GOAL: USE ARM
- \* \* \* \* PRESS ARM OPTION
- \* \* \* GOAL: USE SHOULDER
- \* \* \* \* PRESS SHOULDER OPTION
- \* \* \* GOAL: USE ELBOW
- \* \* \* \* PRESS ELBOW OPTION
- \* \* \* GOAL: USE HAND
- \* \* \* \* PRESS HAND OPTION
- \* \* \* GOAL: USE WRIST
- \* \* \* \* PRESS WRIST OPTION
- \* \* \* GOAL: USE PALM
- \* \* \* \* PRESS PALM OPTION
- \* \* \* GOAL: USE GRIPPER
- \* \* \* \* PRESS GRIPPER OPTION
- \* \* \* GOAL: USE WHOLE ARM

*	*	*	* PRESS WHOLE ARM OPTION
*	*	*	GOAL: END OPERATION
*	*	*	* PRESS END OPERATION OPTION
*	*	*	GOAL: SELECT MOVEMENT
*	*	*	REMEMBER MEANINGS OF JOINT AND WHOLE ARM
*	*	*	MOVEMENT OPTIONS AND END OPTION
*	*	*	[SELECT: GOAL: USE IN...if individual joint needs moving
in			
*	*	*	GOAL: USE OUT...if individual joint needs
moving out			
*	*	*	GOAL: USE UP...if whole arm needs
moving up			
*	*	*	GOAL: USE DOWN...if whole arm needs
moving down			
*	*	*	GOAL: USE FORWARD...if whole arm
needs moving			
*	*	*	forwards
*	*	*	GOAL: USE BACK...if whole arm needs
moving			
*	*	*	backwards
*	*	*	GOAL: USE LEFT...if whole arm needs
moving to the left			
*	*	*	GOAL: USE RIGHT...if whole arm needs
moving to the			
*	*	*	right
*	*	*	GOAL: END OPERATION...if no part of
the arm needs			
*	*	*	moving]
*	*	*	GOAL: USE IN
*	*	*	* PRESS IN OPTION
*	*	*	GOAL: USE OUT
*	*	*	* PRESS OUT OPTION
*	*	*	GOAL: USE UP
*	*	*	* PRESS UP OPTION
*	*	*	GOAL: USE DOWN
*	*	*	* PRESS DOWN OPTION
*	*	*	GOAL: USE FORWARD
*	*	*	* PRESS FORWARD OPTION
*	*	*	GOAL: USE BACK
*	*	*	* PRESS BACK OPTION
*	*	*	GOAL: USE LEFT
*	*	*	* PRESS LEFT OPTION
*	*	*	GOAL: USE RIGHT
*	*	*	* PRESS RIGHT OPTION
*	*	*	GOAL: END OPERATION





\* \* \* \* **GOAL: USE NOT STOP**  
\* \* \* \* \* **WAIT UNTIL ARM HAS FINISHED MOVING**



## Appendix E: CPM-GOMS Analysis

### GOAL: MOVE SHOULDER OUT BY 20 CMS

\* DETERMINE CURRENT POSITION OF ARM

#### \* GOAL: USE MOVE

\* \* CHOOSE MOVE OPTION

\* \* (For Gesture:

\* \* ~ retrieve meaning of option

\* \* ~ find option on screen

\* \* ~ wait for cursor to move to that option

\* \* ~ retrieve correct gesture

\* \* ~ perform correct gesture

\* \* ~ confirm option

\* \* For Voice:

\* \* ~ retrieve meaning of option

\* \* ~ say option out loud

\* \* ~ confirm option)

#### \* GOAL: USE SHOULDER

\* \* PRESS SHOULDER OPTION

\* \* (For Gesture:

\* \* ~ retrieve meaning of option

\* \* ~ find option on screen

\* \* ~ wait for cursor to move to that option

\* \* ~ retrieve correct gesture

\* \* ~ perform correct gesture

\* \* ~ confirm option

\* \* For Voice:

\* \* ~ retrieve meaning of option

\* \* ~ say option out loud

\* \* ~ confirm option)

#### \* GOAL: USE OUT

\* \* PRESS OUT OPTION

\* \* (For Gesture:

\* \* ~ retrieve meaning of option

\* \* ~ find option on screen

\* \* ~ wait for cursor to move to that option

\* \* ~ retrieve correct gesture

\* \* ~ perform correct gesture

\* \* ~ confirm option

\* \* For Voice:

\* \* ~ retrieve meaning of option

\* \* ~ say option out loud

\* \* ~ confirm option)

#### \* GOAL: MOVE JOINT (repeat until joint in correct position)

\* \* GOAL: SET SPEED

```

*      *      *      PRESS OPTION SPEED LEVEL
*      *      *      (For Gesture:
*      *      *      ~ retrieve meaning of option
*      *      *      ~ find option on screen
*      *      *      ~ wait for cursor to move to that option
*      *      *      ~ retrieve correct gesture
*      *      *      ~ perform correct gesture
*      *      *      ~ confirm option
*      *      *      For Voice:
*      *      *      ~ retrieve meaning of option
*      *      *      ~ say option out loud
*      *      *      ~ confirm option)
*      *      *      GOAL: USE SLOW
*      *      *      *      PRESS OPTION SLOW
*      *      *      *      (For Gesture:
*      *      *      *      ~ retrieve meaning of option
*      *      *      *      ~ find option on screen
*      *      *      *      ~ wait for cursor to move to that option
*      *      *      *      ~ retrieve correct gesture
*      *      *      *      ~ perform correct gesture
*      *      *      *      ~ confirm option
*      *      *      *      For Voice:
*      *      *      *      ~ retrieve meaning of option
*      *      *      *      ~ say option out loud
*      *      *      *      ~ confirm option)
*      *      *      GOAL: USE GO
*      *      *      *      PRESS GO OPTION
*      *      *      *      (For Gesture:
*      *      *      *      ~ retrieve meaning of option
*      *      *      *      ~ find option on screen
*      *      *      *      ~ wait for cursor to move to that option
*      *      *      *      ~ retrieve correct gesture
*      *      *      *      ~ perform correct gesture
*      *      *      *      ~ confirm option
*      *      *      *      ~ For Voice:
*      *      *      *      ~ retrieve meaning of option
*      *      *      *      ~ say option out loud
*      *      *      *      ~ confirm option)
*      *      *      GOAL: USE STOP
*      *      *      *      PRESS STOP OPTION WHEN ARM IN CORRECT POSITION
*      *      *      *      *      (For Gesture:
*      *      *      *      ~ look at arm
*      *      *      *      ~ determine when arm to stop
*      *      *      *      ~ retrieve meaning of option
*      *      *      *      ~ retrieve correct gesture

```



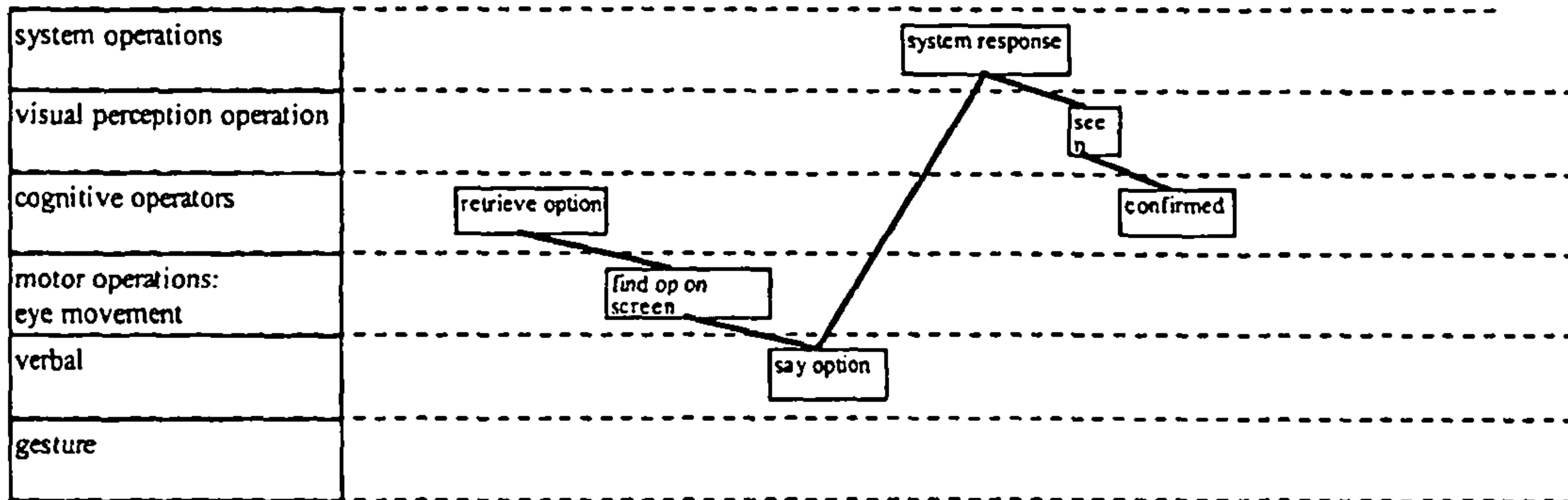
```

*      *      *      ~ wait for arm to reach stop position
*      *      *      ~ perform correct gesture
*      *      *      ~ confirm option
*      *      *      For Voice:
*      *      *      ~ look at arm
*      *      *      ~ determine when arm to stop
*      *      *      ~ retrieve meaning of option
*      *      *      ~ wait for arm to reach stop position
*      *      *      ~ say option out loud
*      *      *      ~ confirm option)
*      GOAL: END OPERATION
*      *      PRESS END OPTION
*      *      (For Gesture:
*      *      ~ retrieve meaning of option
*      *      ~ find option on screen
*      *      ~ wait for cursor to move to that option
*      *      ~ retrieve correct gesture
*      *      ~ perform correct gesture
*      *      ~ confirm option
*      *      For Voice:
*      *      ~ retrieve meaning of option
*      *      ~ say option out loud
*      *      ~ confirm option)

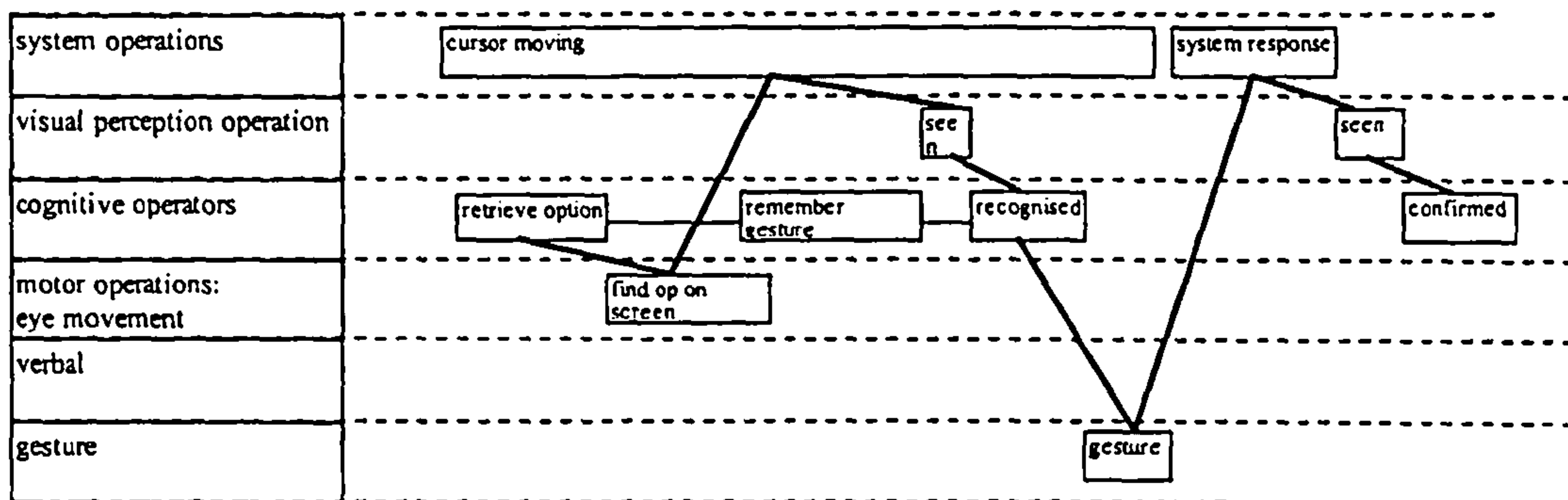
```

## Appendix F: CPM-GOMS Schedule Charts

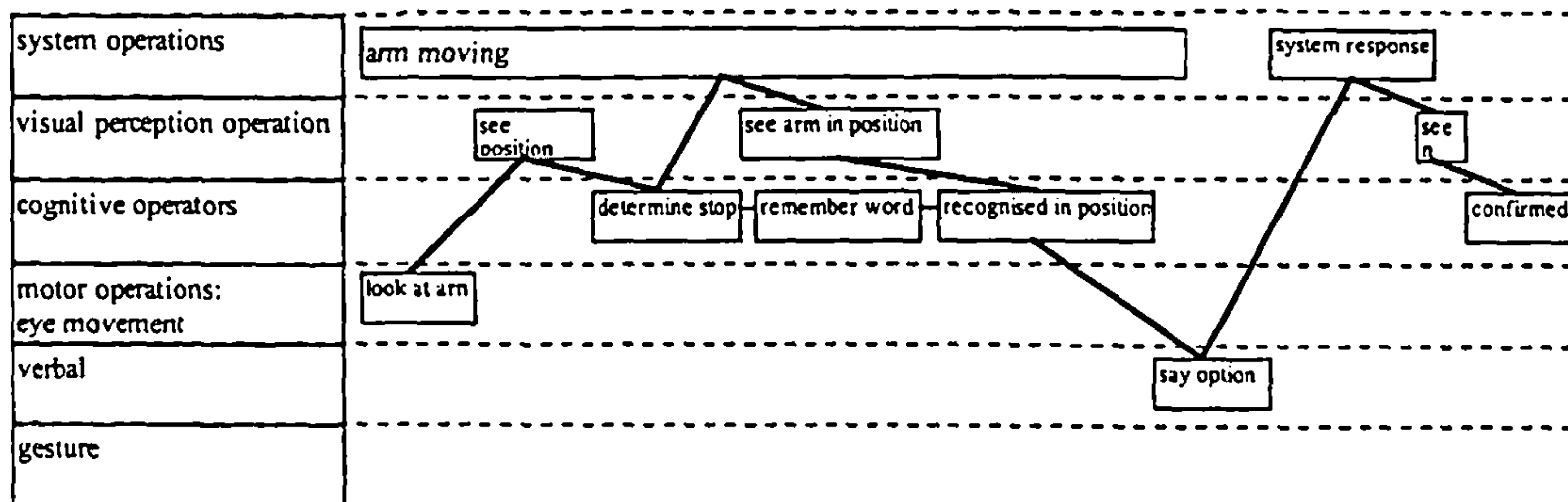
### Verbal Input Schedule Chart:



### Gesture Input Schedule Chart:

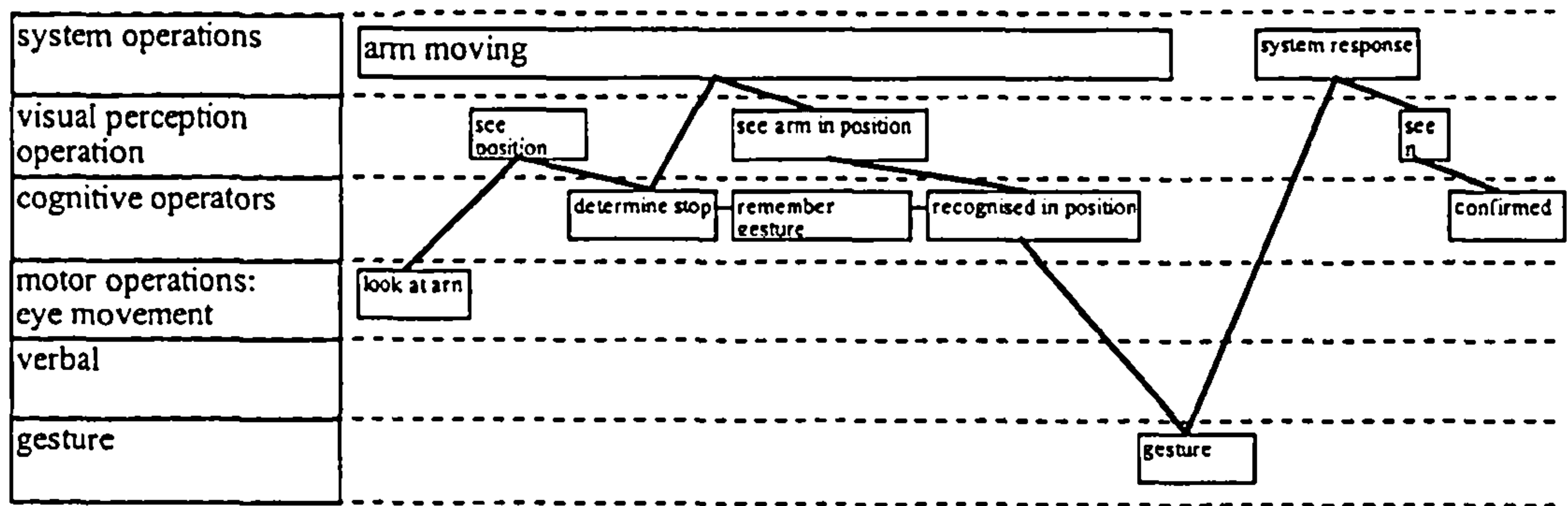


### Verbal Stop Schedule Chart:





### Gesture Stop Schedule Chart:



## Appendix G: Cognitive Walkthrough Analysis

**What is the correct action sequence for each task and how is it described:**

Assume that the arm needs to be moved to the right, then the arm part moved up and the gripper moved out.

The required actions for accomplishing this task for the second analysis are as follows:

(Action sequence starts at main options menu, which includes the options **Move** and **MoveArm**.)

1. Choose option **MoveArm** on the main menu.

The interface gives a list of options: **Left, Right, Up, Down, Forward, Back, Wrong, End**

2. Choose option **Right** on the menu.

The interface gives a list of options: **Go, Speed, Wrong, End**

3. Choose option **Speed** to set speed.

The interface gives a list of options: **Min, Slow, Medium, Fast, Max, Wrong, End**

4. Choose option **Min** for arm to move at slowest speed.

The interface gives a list of options: **Go, Speed, Wrong, End**

5. Choose option **Go** to start arm moving.

The interface gives the option **Stop**.

6. When the arm appears to be in the correct position, choose **Stop**.

The interface gives the options: **Go, Speed, Wrong, End**

7. Choose option **End** to return to the main options menu.

The interface gives the options (amongst others) **Move** and **MoveArm**.

8. Choose option **Move** from the menu.

The interface gives the options **Base, Arm, Shoulder, Elbow, Hand, Wrist, Palm, Grip, Wrong, End**

.

9. Choose option **Arm** from the menu.

The interface gives a list of options: **In, Out, Wrong, End**

10. Choose option **Out** from the menu.

The interface gives a list of options: **Go, Speed, Wrong, End**

11. Choose option **Speed** to set speed.



The interface gives a list of options: **Min, Slow, Medium, Fast, Max, Wrong, End**

12. Choose option **Min** for arm to move at slowest speed.

The interface gives a list of options: **Go, Speed, Wrong, End**

13. Choose option **Go** to start arm moving.

The interface gives the option **Stop**.

14. When the arm appears to be in the correct position, choose **Stop**.

The interface gives the options: **Go, Speed, Wrong, End**

15. Choose option **End** to return to the main options menu.

The interface gives the options (amongst others) **Move** and **MoveArm**.

16. Choose option **Move** from the menu.

The interface gives the options **Base, Arm, Shoulder, Elbow, Hand, Wrist, Palm, Grip, Wrong, End**.

17. Choose option **Grip** from the menu.

The interface gives a list of options: **In, Out, Wrong, End**

18. Choose option **Out** from the menu.

The interface gives a list of options: **Go, Speed, Wrong, End**

19. Choose option **Speed** to set speed.

The interface gives a list of options: **Min, Slow, Medium, Fast, Max, Wrong, End**

20. Choose option **Min** for arm to move at slowest speed.

The interface gives a list of options: **Go, Speed, Wrong, End**

21. Choose option **Go** to start arm moving.

The interface gives the option **Stop**.

22. When the arm appears to be in the correct position, choose **Stop**.

The interface gives the options: **Go, Speed, Wrong, End**

23. Choose option **End** to return to the main options menu.

The interface displays the main options menu.

## Appendix H: Z Specification

### Types

The specification begins with the declaration of the types that are to be used.

ARMPART ::= wholearm | base | arm | shoulder | elbow | hand | wrist | palm | grip

This declares the various parts of the arm, or the whole of the arm, which can be moved. The wholearm is declared as an armpart in order that the schemas might be as simple and straightforward as possible. This is explained in greater detail later on in the specification.

DIRECTION ::= in | out | left | right | up | down | forward | back

All the directions available are declared to be of the type DIRECTION.

SPEED ::= min | slow | medium | fast | max

All the speeds available are declared to be of the type SPEED.

MOTIONSTATUS ::= static | moving

The arm is either stationary or in motion, as defined by the type MOTIONSTATUS.

[POSITION] is currently left only as a type rather than fully defined.

There is only one function to be defined:

extremes : DIRECTION → POSITION

This function states that the extreme position of the robotic arm is a function of the direction, or, in other words, the extreme limit of the arm depends on what direction the arm is moved in.

### The State Schema

The robotic arm is defined by the following schema (words within \* are comments):

RoboticArm	
speed: SPEED	* speed of arm movement *
pos: POSITION	* current position of arm *
motion: MOTIONSTATUS * whether the arm is in motion or not *	

The arm is always in some position, and is always either static or moving. A speed is always set, regardless of whether the arm is in motion or not.

### The Initial and Final States

Initially, the robotic arm only has two things defined: the speed is set to medium by default, and the arm is not yet in motion.



InitRoboticArm
RoboticArm'
motion' = static speed' = medium

Since the End option returns the user to the main menu, effectively it re-initialises the arm whenever it is used, hence the following schema:

End
Δ RoboticArm InitRoboticArm

### Preparing for Motion

In the following two schemas, *armPartToBeMoved* and *dir* are, effectively, additional state variables. However, they have not been included in the state schema, because it is possible for the robotic arm to exist in a state in which neither of these variables are defined.

SelectArmPart
Δ RoboticArm armPartToBeMoved' : ARMPART armPartToBeMoved? : ARMPART
armPartToBeMoved' = armPartToBeMoved?

SelectArmDirection
Δ RoboticArm dir' : DIRECTION dir? : DIRECTION
dir' = dir?

It is appropriate to combine the above two schemas into a single operation which includes selection of both the arm part to be moved and the direction. As one cannot choose a direction until the arm part has been selected, this *SelectArmPartAndDirection* operation effectively becomes the prerequisite for commencing motion.

### SelectArmPartAndDirection

$\Delta$  RoboticArm  
SelectArmPart  
SelectArmDirection

armPartToBeMoved = wholeArm  $\Rightarrow$  dir  $\in$  {in, out}  
armPartToBeMoved  $\neq$  wholeArm  $\Rightarrow$  dir  $\in$  {in, out}

As the speed is always set to something, it has been maintained as a state variable. The following operation describes the user changing the speed, but this is not a prerequisite for motion.

### SelectSpeed

$\Delta$  RoboticArm  
speed? : SPEED

speed' = speed?

## Starting and Finishing Motion

When the arm starts to move, either by using the Go or the Continue command, all that changes in this specification is that the motion variable is now changed to moving. Since both Go and Continue are identical, only one of the schemas needs to be used, however they have both been listed to demonstrate their equivalence.

### Go

$\Delta$  RoboticArm  
SelectArmPartAndDirection

motion' = moving  
extreme = extremes dir

### Continue

$\Delta$  RoboticArm  
SelectArmPartAndDirection

motion' = moving  
extreme = extremes dir

There are two ways of stopping the motion of the arm. Either the limit of movement is reached, or else the user decides to stop the arm. Therefore these two options can be combined into the following operation, made up of two schemas:



$\text{Stop} \cong \text{ReachedLimit} \vee \text{ChoseStop}$

ReachedLimit
$\Delta \text{RoboticArm}$ Go
pos' = extreme motion' = static

ChoseStop
$\Delta \text{RoboticArm}$
motion' = static

### The Multi-Modal Changes

When the specification was revised to take into account the input devices, the specification only changed slightly.

A new type was declared, showing the choice of input devices:

$\text{MODE} ::= \text{voice} \mid \text{gesture}$

The main change was the inclusion of a choice of two new operations in the selection schemas:

$\text{Input} \cong \text{UseVoice} \vee \text{UseGesture}$

The operation *Input* is made up of two operations, *UseVoice* and *UseGesture*, only one of which will apply.

UseVoice
mode: MODE
mode = voice

In *UseVoice*, the voice mode has been used to select the command.

UseGesture
mode: MODE
mode = gesture

In *UseGesture*, the gesture mode has been used to select the command. Only once the cursor is in the right position will the gesture be made.

This *Input* operation is then included into the selection operations, and the schema is complete:

End

$\Delta$ RoboticArm
Input
InitRoboticArm

SelectArmPart

$\Delta$ RoboticArm
Input
armPartToBeMoved' : ARMPART
armPartToBeMoved? : ARMPART

---

armPartToBeMoved' = armPartToBeMoved?
---------------------------------------

SelectArmDirection

$\Delta$ RoboticArm
Input
dir' : DIRECTION
dir? : DIRECTION

---

dir' = dir?
-------------

SelectSpeed

$\Delta$ RoboticArm
Input
speed? : SPEED

---

speed' = speed?
-----------------

Go

$\Delta$ RoboticArm
SelectArmPartAndDirection
Input

---

motion' = moving
extreme = extremes dir



Continue

$\Delta$  RoboticArm  
SelectArmPartAndDirection  
Input

motion' = moving  
extreme = extremes dir

ReachedLimit

$\Delta$  RoboticArm  
Go  
Input

pos' = extreme  
motion' = static

ChoseStop

$\Delta$  RoboticArm  
Input

motion' = static