Avatars, Agency and Performance: The Fusion of Science and Technology within the Arts

A doctoral thesis submitted for the Degree of Doctor of Philosophy.
The Marcs Institute, University of Western Sydney, Australia.

Richard Andrew Salmon 2014
Acknowledgements

I would like to thank the Power House Museum, Ultimo, Sydney, Australia and all the staff there, who provided the space and support for most of the data in this research project to be collected in parallel with the exhibition of the Articulated Head. Thank you to all those at The Marcs Institute at the University of Western Sydney who supported the project, including the Articulated Head Engineering Team, Professor Chris Davis, Dr Damith Herath and Dr Christian Kroos. A special thank you to Zhengzhi Zhang and Dr Richard Liebbrandt for the programming of the Event Manager/Max interface and the Thought Clouds respectively. Thank you to Colin Schoknecht and Steven Fazio for their unwavering and timely technical support. Thank you to Professor Chris Davis for his strong supervisory support in the final stages of completion of this thesis. Thank you to Professor Michael Atherton for his sort lived but influential supervisory support and last but not least to Assoc. Prof Garth Paine and Stelarc, who have been the mainstay of support in the conception and completion of this work. Together, Stelarc, Assoc. Prof Garth Paine and myself formed the Articulated Head Performance Team.
Statement of Authenticity

The work presented in this thesis is to the best of my knowledge and belief, original other than where acknowledged in the text.

I hereby declare that I have not submitted this material, either in full or part, for a higher degree at this or any other institution.

Richard Andrew Salmon, June 15th, 2014
# Table of Contents

1. **Abstract** .............................................................................................................................................. 1

   1.1 The Project ....................................................................................................................................... 1

       The purpose of this research .................................................................................................................. 1

   Introduction ........................................................................................................................................... 2

   Methodology .......................................................................................................................................... 2

   Methods .................................................................................................................................................. 2

   Findings .................................................................................................................................................. 2

2. **Introduction** ..................................................................................................................................... 4

   2.1 Stelarc’s background ....................................................................................................................... 5

       2.1.1 From Talking Head to Thinking Heads: Aims & Objectives ..................................................... 6

   2.2 History ............................................................................................................................................. 7

       2.2.1 Mirrored development: the Prosthetic and Articulated Head .................................................... 7

       2.2.2 The Projected Evolutionary Development of the Head .............................................................. 11

   2.3 Stelarc’s work and the Articulated Head set within the context of other robotic/AI chatbot artworks ......................................................................................................................... 11

   2.4 General history of ideas in relation to Human Machine Interaction ............................................. 13

       2.4.1 Anthropomorphism ................................................................................................................... 16

       2.4.2 Perceptions of agency ................................................................................................................ 17

       2.4.3 Human emotional response to virtual agents .......................................................................... 17

   2.5 The Original Plan and how this investigation differs .................................................................... 19

       2.5.1 A limitation of this investigation ............................................................................................... 20

       2.5.2 Projected incorporation of interactive scenarios ..................................................................... 20

   2.6 Technical Components of the Articulated Head ............................................................................. 21

   2.7 Operational infrastructure, maintenance and control of the Articulated Head Systems (Overview) ..................................................................................................................................... 24

   2.8 Software component based architecture ......................................................................................... 25

       2.8.1 The Thinking Head Attention Model and Behavioural System ................................................. 26

       2.8.2 The Event Manager (Overview) ................................................................................................. 27

       2.8.3 The EM/Max Interface ............................................................................................................... 28

   2.9 Where and when has this investigation taken place? .................................................................... 28

   2.10 The Big Question ............................................................................................................................ 29

   2.11 The Research Questions ................................................................................................................. 29

   2.12 This study’s contribution ................................................................................................................. 30

   2.13 The Approach: Prioritisation of Human Experience ..................................................................... 30

   2.14 What follows? ................................................................................................................................. 31

       2.14.1 Defense of the Articulated Head’s performance .................................................................... 32

       2.14.2 The Problem .......................................................................................................................... 32

3. **Methodology and methods** .................................................................................................................. 34

   3.1 Grounded Theory ............................................................................................................................... 34
3.1.1 Computer assisted qualitative data analysis software (CAQDAS) .................. 38
  3.1.1.1 Axial Coding ........................................................................................................ 39
  3.1.1.2 Frames analysis ................................................................................................... 39
3.2 Video Cued Recall ................................................................................................... 41
   Reporting Experience ..................................................................................................... 41
3.2.1 Expanded discussion of applications of Video Cued Recall ................................. 43
3.2.2 The best time to implement the interview .............................................................. 45
3.3 Phenomenology ......................................................................................................... 46
3.4 A Comparative Analysis ............................................................................................ 51
   Human ↔ human compared to human ↔ robot communication ............................... 51
3.4.1 Communication Theory .......................................................................................... 51
3.4.2 The Universal Law of Communication .................................................................... 52
3.4.3 A quick review: What was the Articulated Head again? ...................................... 53

4. The Reduction ............................................................................................................. 55
4.1 Scope of this Reduction ............................................................................................. 55
  4.1.1 Consciousness ......................................................................................................... 57
  4.1.2 The senses ............................................................................................................... 58
  4.1.3 A spiritual dimension ............................................................................................... 58
  4.1.4 Communication or information exchange, initiations and sense fields ............ 58
  4.1.5 The senses and range .............................................................................................. 59
      Imaginative variations .................................................................................................. 60
      Sight ............................................................................................................................ 60
      Hearing ....................................................................................................................... 60
      Smell .......................................................................................................................... 60
      Taste ........................................................................................................................... 61
      Touch ......................................................................................................................... 61
  4.1.6 Interdependencies and embedded enculturation .................................................. 61
  4.1.7 Sense field pick up patterns .................................................................................... 62
  4.1.8 The whole body sense ............................................................................................ 63
  4.1.9 Genesis .................................................................................................................. 64
  4.1.10 The language of communication ......................................................................... 65
  4.1.11 Non verbal communication .................................................................................. 65
  4.1.12 Sense by sense comparison .................................................................................. 66
      Sight ............................................................................................................................ 66
      Hearing ....................................................................................................................... 67
      Smell .......................................................................................................................... 68
      Taste ........................................................................................................................... 68
      Touch ......................................................................................................................... 69
  4.1.13 The revelation ........................................................................................................ 70
  4.1.14 Conclusion to the reduction ................................................................................ 72
  4.1.15 Tentative findings .................................................................................................. 73
4.2 Practical application of the Video Cued Recall Method .......................................... 75
  4.2.1 Procedure ............................................................................................................... 75
5. Setting the Scene: .............................................................................. 80

5.1 From the engineering - to performance arts perspective .................. 80
  5.1.1 The positive engineering perspective ........................................ 80
  5.1.2 Questioning the approach ......................................................... 81
  5.1.3 The negative perspective ......................................................... 81
  5.1.4 Of positive thought, critical evaluation and leaps of faith .......... 82
  5.1.5 A melioristic yet tempered view towards development ............... 83
    5.1.5.1 The happy particularity with regards to this investigation ... 84
  5.1.6 The influence of the spatio-temporal environment .................... 84

5.2 The analogous performing arts scenario: The Stage .......................... 84
  5.2.1 Illusion and Conscious Perception ........................................... 86
  5.2.2 The importance of what is said and what is not ....................... 87

5.3 Investigative framework: Evidence, not hypothesis led ................. 87

5.4 Philosophical and practical justification of approach ....................... 88
  5.4.1 The current state of play: Popular perceptions ........................ 88
  5.4.2 The role of the Arts in Robotics and vice versa ....................... 90
  5.4.3 The Tryst: The Science/Technology/Arts Nexus ....................... 91
    5.4.3.1 The contrasting methods debate ....................................... 92
    5.4.3.2 Human ↔ machine interaction: Micro versus Macro .......... 92
    5.4.3.3 Emergent Phenomena: Context and Causality: Micro versus Macro ... 94

6. The Creative New Media Propositions .............................................. 98

6.1 The Additions .............................................................................. 98

6.2 Developmental Stages and Logistical Impingements ....................... 100

6.3 Cost Implications and Technical Constraints ................................ 101
  6.3.1 Outline of the cost implications .............................................. 102

6.4 Project links to agency, affect and engagement ............................. 102

6.5 Projected data-flow diagram ....................................................... 105

6.6 Overview of the Hardware Equipment Set-up ................................ 106

6.7 Introduction to the Software Technologies Used ......................... 108
  6.7.1 Max/MSP .............................................................................. 108
  6.7.2 Jamoma ............................................................................... 109
  6.7.3 Ircam SPAT objects for Max ................................................ 109
  6.7.4 Vizzie ............................................................................... 110
# Presentation of Research Data and Findings

## 7.1 Text String Input/Output Analysis

### 7.1.1 User Input String Lists

### 7.1.2 Chatbot Output String Lists

### 7.1.3 Observations made from review of the phrase query lists

#### 7.1.3.1 Observation 1 – The anthropomorphic stance

#### 7.1.3.2 Observation 2 – Programming additions

#### 7.1.3.3 Observation 3

#### 7.1.3.4 Observation 4 – Conversational diversity across the data set

#### 7.1.3.5 Conclusions from the text string analysis

## 7.2 The NVIVO Coding Process and Emergent Themes

### 7.2.1 What was done?

**Coding saturation**

**Theme sorting**

### 7.2.2 Revisiting the Knot

### 7.2.3 The Narrative Emanating from the Coding

## 7.3 The Emergent Themes

### 7.3.1 Theme 1: Anthropomorphism

### 7.3.2 Theme 2: Expectations and enculturation

### 7.3.3 Theme 3: Physical Presentation and Operational Aspects

### 7.3.4 Theme 4: Mode of Interaction

### 7.3.5 Theme 5: Movement, Tracking and Control

### 7.3.6 Theme 6: Dialogue

### 7.3.7 Theme 7: Engagement

### 7.3.8 Theme 8: Emotions

### 7.3.9 Theme 9: Senses – Related

---

### 6.8 The Creative New Media Projects Development

#### 6.8.1 The Projections

#### 6.8.2 Rationale

**Execution of the prototype Text Projections**

#### 6.8.3 The Creative Additions at the Powerhouse Museum

**The Spatial Auditory System**

**The Thought Cloud Textual Projections**

---

### 6.7.5 Osculator

### 6.7.6 Open Sound Control

### 6.7.7 The CNMAT Library

### 6.7.8 The [c74] Object

### 6.7.9 Processing

### 6.7.10 Director

### 6.7.11 TeamViewer

### 6.7.12 Alchemy Player
11. Bibliography .................................................................................................................. 256
  11.1 References .................................................................................................................. 256
12. Appendices (Hard Copies) ............................................................................................. 267
  12.1 Document appendices ............................................................................................. 267
    12.1.1 Appendix 1 – Plan View Diagram of the Exhibit .............................................. 267
    12.1.2 Appendix 2 – The Auditory Visual References Table ................................... 267
    12.1.3 Appendix 3 The Expenses Table ................................................................... 267
  12.2 Electronic Appendices ............................................................................................. 267
    12.2.1 Appendix 1 – An Introduction to Head 0+ ..................................................... 267
    12.2.2 Appendix 2 – From Talking Heads to Thinking Heads ................................. 267
    12.2.3 Appendix 3 – Articulated Head – System Quick Start Guide ...................... 267
    12.2.4 Appendix 4 – THAMBS documentation ....................................................... 267
    12.2.5 Appendix 5 – Participant Information Pack .................................................. 267
    12.2.6 Appendix 6 – Participant Consent Form ....................................................... 267
    12.2.7 Appendix 7 – Participant Release Form ......................................................... 267
    12.2.8 Appendix 8 – Participant Questionnaire ....................................................... 267
    12.2.9 Appendix 9 – Ethics Amendment Approval ................................................... 267
    12.2.10 Appendix 10 – Road Traffic Accident Investigative Scenario .................... 267
    12.2.11 Appendix 11 – Creative New Media Additions Documentation ................ 267
    12.2.12 Appendix 12 – Slideshow of Powerhouse Exhibit Construction .................. 267
    12.2.13 Appendix 13 – Top 100 User Input String Words ....................................... 267
    12.2.14 Appendix 14 – Top 100 Chatbot Output String Words ............................ 267
    12.2.15 Appendix 15 – Favourite - April2012-Results Preview ............................... 267
    12.2.16 Appendix 16 – User Input String Phrase Table .......................................... 267
    12.2.17 Appendix 17 – Chatbot Output String Phrase Table ................................... 267
    12.2.18 Appendix 18 – Example 1 ................................................................. 267
    12.2.19 Appendix 19 – Example 2 ................................................................. 267
    12.2.20 Appendix 20 – Example 3 ................................................................. 267
    12.2.21 Appendix 21 – Example 4 ................................................................. 267
    12.2.22 Appendix 22 – Chatbot’s top 1000 words .................................................. 267
    12.2.23 Appendix 23 – Example 5 ................................................................. 267
    12.2.24 Appendix 24 – Top 1000 User Input String Words .................................... 267
    12.2.25 Appendix 25 – Top 10,000 User Input String Words ............................... 267
    12.2.26 Appendix 26 – Top 10,000 Chatbot Output String Words ...................... 267
    12.2.27 Appendix 27 – Example 6 ................................................................. 267
    12.2.28 Appendix 28 – See .............................................................................. 268
    12.2.29 Appendix 29 – Example 7 ................................................................. 268
    12.2.30 Appendix 30 – Example 8 ................................................................. 268
    12.2.31 Appendix 31 – Keyboard Evidence .......................................................... 268
    12.2.32 Appendix 32 – Example 9 ................................................................. 268
### List Of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-1</td>
<td>The Freakish Peak</td>
</tr>
<tr>
<td>2-2</td>
<td>The Head</td>
</tr>
<tr>
<td>2-3</td>
<td>The Articulated Head and Kiosk</td>
</tr>
<tr>
<td>2-4</td>
<td>(Herath, 2012)</td>
</tr>
<tr>
<td>4-1</td>
<td>Interaction Capture Set-up</td>
</tr>
<tr>
<td>4-2</td>
<td>Video Cued Recall Interview Set-up</td>
</tr>
<tr>
<td>5-1</td>
<td>Professor Hiroshi Ishiguro (right) sits next to the Geminoid HI-2 robot</td>
</tr>
<tr>
<td>5-2</td>
<td>Bina48 an early demonstration of &quot;mindfile&quot; transfer at Lifeanaut.com</td>
</tr>
<tr>
<td>5-3</td>
<td>Leonardo Da Vinci’s depiction of a vertical flying machine</td>
</tr>
<tr>
<td>5-4</td>
<td>A Star Trek Communicator device</td>
</tr>
<tr>
<td>6-1</td>
<td>Data flow diagram for Creative Media Augmentations</td>
</tr>
<tr>
<td>6-2</td>
<td>Simplified Augmentations Equipment Set-up</td>
</tr>
<tr>
<td>6-3</td>
<td>Installation of wiring: Left hand lab cupboard prior to hardware installation</td>
</tr>
<tr>
<td>6-4</td>
<td>Left-hand lab cupboard after hardware installation</td>
</tr>
<tr>
<td>6-5</td>
<td>The Articulated Head Set-up and Projections at SEAM</td>
</tr>
<tr>
<td>6-6</td>
<td>Prototype Projection Data Flow Diagram</td>
</tr>
<tr>
<td>6-7</td>
<td>Projections Extended Screen Display layout</td>
</tr>
<tr>
<td>6-8</td>
<td>Director projection control</td>
</tr>
<tr>
<td>6-9</td>
<td>Director projection embedded patcher</td>
</tr>
<tr>
<td>6-10</td>
<td>Director projection embedded patcher window</td>
</tr>
<tr>
<td>7-1</td>
<td>Video Cued Recall Annotations, dialogue Transcription and Commentary</td>
</tr>
<tr>
<td>9-1</td>
<td>Maslow’s Hierarchy of Needs</td>
</tr>
<tr>
<td>9-2</td>
<td>The Freakish Peak</td>
</tr>
<tr>
<td>9-3</td>
<td>Interactive environment design refinement &amp; layout</td>
</tr>
<tr>
<td>9-4</td>
<td>Enclosure Cross Section View</td>
</tr>
<tr>
<td>9-5</td>
<td>Plan View of Design Refinement</td>
</tr>
</tbody>
</table>
List Of Tables

Table 6-1 - Project links to Research Questions ................................................................. 103
Table 6-2 - The WiiMote Interface .......................................................................................... 120
Table 7-1 - User input strings - top 10 words ......................................................................... 130
Table 7-2 - Chatbot output strings - top 10 words ................................................................. 130
Table 7-3 User Input (left) Chatbot Output (right) Selected Search Words .................................. 132
Table 7-4 - Word ‘name’ phrase query list .............................................................................. 135
Table 7-5 - Word ‘like’ phrase query list .................................................................................. 135
Table 7-6 - Word ‘old’ phrase query list .................................................................................. 136
Table 7-7 - Word ‘say’ phrase query list .................................................................................. 136
Table 7-8 - Word ‘sing’ phrase query list .................................................................................. 137
Table 7-9 - Word ‘favourite’ phrase query list ........................................................................ 138
Table 7-10 - Word ‘colour’ phrase query list .......................................................................... 138
Table 7-11 - Word ‘love’ phrase query list .............................................................................. 139
Table 7-12 - Word ‘dance’ phrase query list .......................................................................... 140
Table 7-13 - Word ‘life’ phrase query list .............................................................................. 140
Table 7-14 - Word ‘cool’ phrase query list ............................................................................. 141
Table 7-15 - Word ‘robot’ phrase query list .......................................................................... 141
Table 7-16 - Word ‘Harry’ phrase query list .......................................................................... 142
Table 7-17 - Word ‘live’ phrase query list .............................................................................. 142
Table 7-18 - Word ‘eat’ phrase query list .............................................................................. 142
Table 7-19 - Word ‘happy’ phrase query list .......................................................................... 143
Table 7-20 - Word ‘time’ phrase query list .......................................................................... 143
Table 7-21 - Word ‘speak’ phrase query list .......................................................................... 144
Table 7-22 - Word ‘think’ phrase query list .......................................................................... 144
Table 7-23 - Word ‘name’ phrase query list .......................................................................... 145
Table 7-24 - Word ‘Stelarc’ phrase query list ........................................................................ 145
Table 7-25 - Word ‘think’ phrase query list .......................................................................... 146
Table 7-26 - Word ‘favourite’ phrase query list .................................................................... 146
Table 7-27 - Word ‘think’ phrase query list .......................................................................... 147
Table 7-28 - Word ‘people’ phrase query list ........................................................................ 147
Table 7-29 - Word ‘time’ phrase query list .......................................................................... 148
Table 7-30 - Word ‘hear’ phrase query list ........................................................................... 148
Table 7-31 - Word ‘talk’ phrase query list ............................................................................ 149
Table 7-32 - Word ‘understand’ phrase query list ................................................................... 150
Table 7-33 - Word ‘western’ phrase query list ...................................................................... 150
Table 7-34 - Word ‘human’ phrase query list ....................................................................... 151
Table 7-35 - Word ‘old’ phrase query list ............................................................................. 152
Table 7-36 - Word ‘color’ phrase query list .......................................................................... 152
Table 7-37 Video Cued Recall Interaction Captures ............................................................... 160
Table 12-1 - Cost Implications .............................................................................................. 271
How to navigate and read this document

**IMPORTANT**: There is a DVD supplied with this thesis that installs a ‘READ ME’ file and TWO folders, which include electronic documentation, to the top directory of the system hard drive. The file and folders are titled: ‘AAP_READ_ME.pdf’, ‘AAP_Contents’ & ‘AAP_Cases’ respectively.

To install the electronic documentation, simply double click (run) either the Mac or Windows installer - dependent on your system. It is essential that these files be placed at the top directory of the Macintosh HD/ (Mac) or C:/drive (Windows) prior to opening the electronic documentation.

**IMPORTANT**: when using the Windows installer, please be sure to choose the C:/drive as the install location when prompted, otherwise the files will automatically be installed into your applications folder by default, and will not work properly. This is not required with the Mac installer, because it makes the correct location selection for you automatically. The installed files can be trashed from the hard drive after use with no implications, no extra data is written to disc.

The main content of this document is divided into sections by major and sub-section headings as detailed in the table of contents above. It is of course recommended to read the whole document from beginning to end. However, the section and sub-section headings provide an outline of the content within.

To supplement the main printed thesis a wider body of evidence is supplied to readers on the DVD

Appendices to this thesis are divided into two main groups:

1. Those, which are included as hard copies at the end of the printed thesis - which may also be accessed via links when viewing electronically (note: only essential short documents are included as hard copies).
2. Those, which are only available when viewing the electronic documentation supplied via DVD installation (long and/or non essential shorter documents and auditory visual material). These documents can be accessed via hyperlinks provided in the Key Evidence Links Table at: [system drive]/AAP/AAP Content/Cases once DVD installation has taken place.

The presentation of evidence in the main text, documents major findings and presents key windows or snapshots into a much larger body of substantiating evidence. The larger body of substantiating evidence is far too expansive to present in the main text, or the appendices to this thesis as it runs into thousands of pages and includes several hours of audiovisual data. Hence provision of the Key Evidence Links Table that allows the reader to access key pieces of substantiating evidence quickly and easily, without the need to sift through a larger body of documentary data or appendices to verify key points made.

To be able to access all the electronic evidence presented on disc, Microsoft Word and Excel, Adobe Acrobat Reader and a browser or other media player that supports viewing of Windows Media Video files are required. Windows video codecs are used so viewing of audiovisual evidence on a Windows machine is advised if possible, but it is not essential so long as a compatible media player is installed on the viewing system. This thesis was prepared on a Mac but because NVivo¹ ("NVivo 10 research software for analysis and insight," 2012) was a Windows only software application until very recently, most of the evidential data has been collated and exported from NVivo (Windows), therefore the exports are predominantly word documents and Windows Media files.

Audio/visual clips are supplied as exported NVivo webpage links and/or direct links to related video files. This approach provides options for access of the data on a range of computer platforms, with a very wide variety of

¹ NVivo is a Computer Assisted Qualitative Data Analysis Software program supplied by QSR International
differing software situations catered for. The electronic documentation provided has been tested on a variety of Windows (XP service pack 2 and above) and Mac OS (10.5.8 and above) operating systems successfully using Firefox, Google Chrome and Safari with .wmv file browser support.
1. Abstract.

1.1 The Project
The aim of this thesis was to develop a deeper understanding of the nature and context within which interactions between a robotic exhibit (dynamic avatar and conversational agent) called the Articulated Head (AH) and a human interacting with it took place. The main focus of the research was directed at understanding the nature of the experience of avatar, conversational agent and audience interaction within the constraints and context of the Articulated Head’s exhibited status: that of an enclosed situated robotic exhibit in a public exhibition space. The Articulated Head is predominantly referred to in the past tense because the project is essentially over. The hardware components of the Articulated Head were decommissioned from exhibition at the Powerhouse Museum in Ultimo, Sydney, Australia in November 2012.

The purpose of this research
The purpose of this research was to consider how interaction between humans and the Articulated Head could be improved. Therefore, a primary goal of this investigation was to find ways in which human engagement was enhanced during these interactions.

The following research questions underpinned the investigation:

a. In which ways ‘has and can’ the avatar be given agency?

b. In which ways ‘is and can’ the avatar be multi-sensory and multi-present?

c. In which ways ‘can and do’ the characteristics of a `virtual’ performer (e.g., personality, nuance of response, context) condition affect and engagement?

Auditory-visual creative new media additions (see section 6) were added to the Articulated Head’s exhibit space with the express intention that they would help contribute to enhancing the interactive experience for the active audience visiting the Articulated Head. These auditory-visual additions were tested to determine whether they would enhance active audience engagement and thereby improve the exhibits interaction design.
**Introduction**

An introduction the Articulated Head and Stelarc (the concept artist) are given in Section 2 followed by a historical and contextual view of the exhibit in relation to other robotic artworks. The general history of human-machine interaction and established concepts are then presented, followed by an exposition of the operational architecture of the Articulated Head, including other practical and theoretical considerations pertinent to the investigation.

**Methodology**

The methodological framework within which this investigation was situated had at its centre Grounded Theory and Phenomenology. The reasons for choosing this methodological framework are explained in Section 3.

**Methods**

The methods utilised in data collection were:

1. Electronic storage of textual input/output conversational data that was exchanged between the human and the machine in these interactions.
2. Video Cued Recall Interviews and questioning of research participants in relation to their interactions.
3. Researcher observational data.

The methods used in analysis of data were:

1. A word frequency/phrase analysis of textural input/output data.
2. An Interpretive Phenomenological analysis of the Video Cued Recall Interviews and their associated interaction data using a computer assisted qualitative data analysis software (CAQDAS) program called NVIVO ("NVivo 10 research software for analysis and insight," 2012) (see Section 7).

**Findings**

The findings from this research identified and examined various phenomena that occurred during both the recorded Video Cued Recall Interviews, and the observed human machine interactions that took place. A significant finding of this investigation was the identification of “The Freakish Peak
Phenomenon” (see Figure 2-1 section 2.4.3 & Figure 9-2 section 9.2.12). This Phenomena and other findings as examined in Sections 7, 8 & 9 formed the basis of “The Blueprint of Emergent Recommendations”, a group of suggested design refinements presented in a sequence of diagrams (Figures 9-3, 9-4 & 9-5) and their accompanying text in Section 9. The suggested design refinements take direct account of the findings from this investigation and target improvement of the interaction design of this and similar types of interactive exhibits accordingly.

Findings from this research are of relevance and consequence to other similar interactive exhibits and situated robotic designs that incorporate Chatbots and involve human machine interaction. The recommendations for realignment of the design and layout are also of consequence for the conclusions of this work in that they directly address the core question of this research project: How can interaction between humans and machines be improved? The design refinements link back to the research questions a, b, and c on page 1 of this thesis in that they offer a range of suggestions as to how the avatar can be given agency, can be multisensory and multi present, and how the characteristics of a 'virtual' performer can enhance the affect and engagement of its audience. The suggested refinements directly address several unresolved issues raised by the artist Stelarc, and the wider performance team comprising Stelarc, Associate Professor Garth Paine and myself during installation of this exhibit in the Powerhouse Museum, as well as those issues identified by analysis of the research participant Video Cued Recall Interviews and associated data.
2. Introduction

The Australian Research Council (ARC) and the National Health and Medical Research Council (NHMRC) worked together to fund the Thinking Systems initiative in 2005/6. This funding initiative was aimed at supporting cross-disciplinary research.

This thesis and the research conducted to complete it, stem from that Thinking Systems initiative grant. The Thinking Systems initiative was a large-scale multi-institutional cross-disciplinary project involving a range of research initiatives related to thinking systems, robotics, human ↔ computer and human ↔ robot interaction.

The Thinking Head Project based at The MARCS Institute (Institute, 2012) at the University of Western Sydney ("University of Western Sydney," 2009) was one of three successful applications for this research initiative;

1. From Talking Heads to Thinking Heads: A Research Platform for Human Communication Science
2. Optimizing autonomous system control with brain-like hierarchical control systems
3. Thinking Systems: Navigating Through Real and Conceptual Spaces

The Thinking Head project was a large scale, five-year art-science research project that owed its inception to Stelarc’s Prosthetic Head. The Prosthetic Head incorporated Head0+ as detailed in E-Appendix 1: An Introduction to Head Zero (Luerssen & Lewis, 2008). The project brought together speech scientists, psychologists, software developers, artificial intelligence experts and artists.

The research reported herein is based upon one particular component of the Thinking Head project, The Articulated Head. This was a part of the large scale multi-institutional and cross-disciplinary project called the “Thinking Head Project” (Herath, Kroos, Stevens, Cavedon, & Premaratne, 2010), and
was conceptually the artistic vision of well-known Australian performance artist, Stelarc (Zebington, 2012).

2.1 Stelarc’s background
“Stelarc is an Australian performance and installation artist who has internationally exhibited, presented and performed in Europe, Asia, North America, South America and Australia. He has visually probed and acoustically amplified his body. He has made three films of the inside of his body, filming three meters of the internal space of his lungs, stomach and colon. Between 1976 and 1988 he completed 25 body suspension performances with hooks into the skin. He has used medical instruments, prosthetics, robotics, Virtual Reality systems, the Internet and biotechnology to explore alternate, intimate and involuntary interfaces with the body. He has performed with a Third Hand, a Virtual Arm, a Stomach Sculpture and Exoskeleton, a 6-legged walking robot. His Prosthetic Head is an embodied conversational agent that speaks to the person who interrogates it.

As part of the Thinking Head Project he collaborated with The MARCS Institute at the University of Western Sydney to develop the Articulated Head, the Floating Head and the Swarming Heads projects. He is surgically constructing and stem cell growing an Ear on Arm that will be electronically augmented and Internet enabled. The first surgical procedures occurred in 2006. He has recently been performing with his four avatar and automaton clones on his Second Life site, exploring actual-virtual interfaces and gesture actuation of his avatar with Kinect.


In 1997 he was appointed Honorary Professor of Art and Robotics at Carnegie Mellon University, Pittsburgh USA. From 2000-2003 he was Senior Research Fellow at the Nottingham Trent University, Nottingham UK where he developed the MUSCLE MACHINE, a 5m diameter-walking machine using pneumatic rubber muscles. In 2003 he was awarded an Honorary Degree of Laws by Monash University. He received a New Projects grant from the Australia Council in 2010 to develop a micro-robot. Stelarc was Senior Research Fellow and Visiting Artist at the MARCS Auditory Labs (now known as The MARCS Institute) at the University of Western Sydney, Australia, between 2006 - 2011. In 2010 he was also awarded the Prix Arts Electronica Hybrid Arts Prize. He is currently (at the time of writing this document) Chair in Performance Art, School of Arts, Brunel University London, Uxbridge, UK. Stelarc’s artwork is represented by the SCOTT LIVESEY GALLERIES in Melbourne” [Stelarc - personal communication – biographical notes]. Stelarc is now a Professor at Curtin University("Professor Stelarc;" 2013)

2.1.1 From Talking Head to Thinking Heads: Aims & Objectives

The Australian Research Council grant application related to the Thinking Head; E-Appendix 2: From Talking Heads to Thinking Heads clearly outlines the aims and objectives projected at the time. The application states, “the overarching goal is to establish a Talking Head Research Platform” that will;

- seed development of a new generation of talking head technology through gradual elaboration of a Thinking System that allows (a) implementation and integration of hardware and software components from various technology disciplines within Human Communication Science (HCS); and (b) iterative testing and
development via integration of research and researchers across a wide range of HCS disciplines; and

- provide a challenging and maieutic environment that enables researchers to address (a) significant questions in HCS and Embodied Conversational Agent (ECA) research, and (b) key questions in individual research areas, which could not be addressed without the implementation and development of the Thinking Head.

The Thinking Head Project and more specifically the Articulated Head has provided this challenging and maieutic environment, serving as a stage for the testing of various software engineering developments and the creative new media arts projects, which form a part of this research work. The Articulated Head, which was installed as an interactive exhibit at the Powerhouse Museum, Ultimo, Sydney, Australia between February 2011 and November 2012, has been a central focus for several published papers, some of which are referenced elsewhere in this document. The Articulated Head exhibit; its immediate environment and its interacting audience are the central concern of all the research detailed forthwith.

2.2 History

2.2.1 Mirrored development: the Prosthetic and Articulated Head

The Prosthetic Head also referred to as Head0 (zero) is an implementation of an Embodied Conversational Agent (ECA) based on and used by the Artist Stelarc (Luerssen & Lewis, 2008).

Stelarc describes the Prosthetic Head’s evolutionary developments into a number of other project iterations from his own perspective as follows:

“The Prosthetic Head project was realized in San Francisco with the assistance of three programmers. Karen Marcello (system configuration & Alice-bot customization), Sam Trychin (text to speech customization of 3D animation) and Barrett Fox (3D modeling and animation). The development of the Prosthetic Head was conducted in consultation with Dr. Richard Wallace, creator of Alice-bot and Artificial Intelligence Mark Up Language (AIML).
The Prosthetic Head was the first art project of mine that involved an interactive installation using verbal exchange. The aim was to construct an automated, animated and reasonably informed artificial agent. If the head responded appropriately, with convincing answers and facial expressions, it would be an effective - or “intelligent”- agent. It is an embodied conversational agent with real-time lip-syncing. That is; if you ask it a question via a keyboard, it searches its database for an appropriate answer and speaks the response. Rather than an illustration of an artificial intelligence, it is seen more as a conversational system that coupled to a human head is capable of carrying on a conversation. The idea was that - as its database extended, the head would become more unpredictable in its response. The artist would no longer be able to predict what his head says. The head’s database had a repertoire of stories, definitions and philosophical insights preprogrammed.

The 3D model was a 3000 polygon mesh, skinned with the artist’s face texture so that it somewhat resembled him. Its eyeballs; tongue and teeth are separate moving elements. In retrospect it was a kind of digital portrait of the artist. As well as its general database, it was also programmed with personal details of the artist as well as the artist’s opinions, theories and arts practice.

The Prosthetic Head is projected as a 5m high head. In the installation at the Australian Centre for the Moving Image (ACMI) (Thood, 2012) in Melbourne, an ultrasound sensor detected when someone approached the keyboard. The Prosthetic Head then turned to confront the user and initiated the conversation by asking the user who they were and what they are doing there. The Prosthetic Head being projected 5m in height gave the head a disembodied but strong feeling of presence. As well as carrying on a conversation, it generated its own poetry-like verse, and its own song-like sounds. These were different each time you asked it. So it was creative in the sense that it generated unexpected juxtapositions of words and sounds.
The Prosthetic Head also generated the Walking Head and the Partial Head. These works preceded the Thinking Head research project led by The MARCS Institute (Institute, 2012) at the University of Western Sydney (UWS) ("University of Western Sydney," 2009). The Walking Head was a 2m diameter, 6-legged, autonomous walking robot with a human head displayed on its LCD screen. The scanning ultrasound sensor detected if anyone entered the gallery space, it then selected from its library of possible movements and performed a choreography that lasted for approximately one and a half minutes. It then sat and went to sleep until the next visitor came into the gallery. With the Partial Head, the artist’s face was scanned, as was a hominid skull. We then did a digital transplant producing a composite human-hominid face. Using that data we 3D printed a scaffold, seeded it with living cells to grow a living skin, a third face. If the Prosthetic Head was a kind of digital portrait of the artist, the Walking Head was a robotic portrait of the artist and the Partial Head was a bioengineered portrait.

As part of the Thinking Systems research funding the Prosthetic Head AIML Software was added to with the Head0+ version, this task was conducted by Martin Luerssen of Flinders University ("Flinders University in Adelaide, South Australia - Flinders University," 2012). We could script facial expressions within the AIML text. The significant realization was that an intelligent agent might have to be more than a virtual embodiment to become a more seductive and interactive agent.

The Prosthetic Head became the Articulated Head, then the Floating Head (a collaboration with NXI Gestatio in Montreal (Gestatio NXI, 2012)) and finally the Swarming Heads (a collaboration with the Robotics Lab at The Marcs Institute (Institute, 2012) at the University of Western Sydney ("University of Western Sydney," 2009)). With all these iterations of an embodied conversational agent there is an interest in alternate architectures, and what vocabularies of behaviour generate aliveness in these systems. Not only in the virtual behaviour of the heads, but in their physical components and in the interaction of the two. For the Articulated Head an LCD screen imaging the Prosthetic Head was fixed to a 6 degree-of-freedom industrial robot arm,
allowing a 3D task envelope of interaction. The challenge with the Floating Head was that, its responses, compared with the Articulated Head, were much slower. However, this did not seem to affect the interaction with people. For example the Floating Head would hover low to the floor but when someone approached it, it would rise and move forwards towards you, acknowledging your presence. If there were a group of people around it, it would display curious behaviour moving sideways and turning one way then another, and if there were a crowd of people it would become nervous and rise higher. The Swarming Heads are a cluster of seven small wheeled robots each displaying the Prosthetic Head in their mounted screens. A Kinect motion sensor was also mounted at the front of each robot, allowing the visitor to interact with them. This multiple embodiment has the potential of becoming the most seductive system. Not only can a user interact with each robot through gesture, but also the cluster of robots might be able to generate emergent behaviours in their interactions. So as well as avoidance behaviour, predator and prey and flocking behaviour is possible. Also, if the robots are able to perform head recognition when they turn and face each other, then that could trigger conversational exchanges between them. The Swarming Heads could also become a Skype platform, enabling remote users to interact with other remote users through movement as well as image and voice”. [Stelarc – personal communication]

Stelarc’s overview of the various iterations of the Prosthetic Head, as manifested in the subsequent project developments detailed above, alludes to some aspects of how the Head, as an avatar, might be given agency - including ways in which it could be multi-sensory and multi-present. Suggestions for ways in which the characteristics of the ‘virtual’ performer (e.g., personality, nuance of response, context) condition affect and engagement are also inferred in Stelarc’s overview. Agency is of prime interest to this investigation. However, it must be noted that this study, although not ignoring the various other iterations of the Prosthetic Head as implemented in projects such as the Floating Head and the Swarming Heads, is firmly focused on the Articulated Head robotic installation, its interacting audience - and the environment in which these interactions took place.
2.2.2 The Projected Evolutionary Development of the Head
The research and development plan as detailed in E-Appendix 2: From Talking Heads to Thinking Heads puts forward a range of aspects for consideration in development of the Thinking Head, under four different subsection titles; Embodiment, Performance, Interaction and Evaluation respectively. The Embodiment title incorporates aspects such as speech recognition, speech synthesis, and dialog management systems to allow the Head to converse. The Performance title incorporates new media arts and exploration of the Embodied Conversational Agent (ECA) as a virtual performer. The Interaction title incorporates a focus on auditory-visual (AV) speech, and its integration with key aspects of intelligence: prediction, interaction and learning. The Evaluation title has a focus on controlled experiments on behavioural interactions between the Head and research participants; the findings of these experiments would drive Head development.

2.3 Stelarc’s work and the Articulated Head set within the context of other robotic/AI chatbot artworks.
Stelarc’s work is situated within the context of a number of other artistic works, involving artificial intelligence and robotic style projects that have taken place in various locations around the world in recent years. Projects such as Mari Velonaki’s Fish-Bird project which has “wheelchairs, that can communicate with each other and with their audience through the modalities of movement and written text.” (Velonaki, 2006). The wheelchairs are conceptually ‘in love’ and they write letters to each other, which are left on the floor for an audience to see. The audience can disturb the wheelchairs, which can result in them modifying the content of letters dropped to be less personal or emotional. The Articulated Head project has some parallels with the Fish-Bird project in the respect that it can ‘sense’ the presence of an audience and modify its behavior as a result, it also has preprogramed messages, which are delivered in speech rather than text.

Another artwork that shares some parallels with the Articulated Head is Ken Feinegold’s robotic/animatronic chatbot art work “In all his work, Feingold
programs his creations so that the conversations are continually changing: the robots’ dialogue, he has said, is neither scripted nor random, but instead designed to mimic “personality”: “a vocabulary, associative habits, obsessions, and other quirks ... which make their conversations poetic, surprising, and often hilarious,”(Maclaughlin, 2004)

Norman White’s helpless robot “is an artificial personality that responds to the behavior of humans by using its electronic voice which speaks a total of 512 phrases. The speech that is delivered depends on its present and past experience of “emotions” ranging from boredom, frustration, arrogance, and overstimulation”.(White, 2011) The Articulated Head had a significantly wider vocabulary of phrases than the Helpless Robot and it was capable of a range of facial expressions that indicated emotions such as frowning and smiling.

Many robotic projects taking place around the world share parallels with the Articulated Head to differing degrees. Kismet and Cog are two sociable robots built at Massachusetts Institute of Technology (MIT). Cynthia Breazeal, a roboticist at Massachusetts Institute of Technology, has worked extensively with the robot called Kismet, a rather beautiful robot with big red/pink lips, eyes and ears. Kismet can talk and give human child like responses to verbal cues from Cynthia, who is something like a parent to Kismet. Breazeal comments in a YouTube video interview(Anders, 2006) that she would eventually like to give Kismet a face. Both Cog and Kismet can perform various motor skills (“Kismet,” 2011) and both have the ability to engage an audience by seemingly paying attention to them. In her book Designing Sociable Robots, Cynthia Breazeal points out “many skills can be thought of as fixed action patterns (FAPs). Each FAP consists of two components, the action component and the taxis (or orienting) component. For Kismet, FAPs often correspond to communicative gestures where the action component corresponds to the facial gesture, and the taxis component (to whom the gesture is directed) is controlled by gaze. People seem to intuitively understand that when Kismet makes eye contact with them, they are the
locus of Kismet’s attention and the robot’s behavior is organized about them. This places the person in a state of action readiness where they are poised to respond to Kismet’s gestures. (Breazeal, 2004, p. 149)

Using facial expressions to signify emotions is something that has been explored by many sociable robots; Kismet and the Articulated Head were capable of facial expression. “Whether we consciously recognize it or not, expressing emotion is often done as a means of communicating our feelings within the context of a target audience. (Harris & Sharlin, 2011, p. 442)

Cog was a moving robot that was built to look and act like a human (Anders, 2006) with a set of sensors and actuators which try to approximate the sensory and motor dynamics of a human body. The Articulated Head shared some similar functions and attributes of robots like Cog and Kismet, and placed an audience in readiness to respond in similar ways to those described above. The visual presentations of Cog, Kismet and the Articulated Head are very different, yet all three share the ability to make gestures of one form or another using goal driven motor skills. “People differentiate humans and computers along the lines of intentionality but initially equate robots and computers. However, the tendency to equate computers and robots can be at least partially overridden when attention is focused on robots engaging in intentional behavior.” (Levin, Killingsworth, Saylor, Gordon, & Kawamura, 2013, p. 161) All three robots have the ability to sense an audience and respond. “It is known that simple saliency mechanisms in the human visual attention system, trigger multiple reflexive behaviors” (Ruesch et al., 2008, p. 962). Cog, Kismet and the Articulated Head also demonstrate multiple reflexive behaviours in response to sensing an audience.

2.4 General history of ideas in relation to Human Machine Interaction.

The Articulated Head project was related to artificial intelligence (AI) in that AI did form a part of the performance of the avatar, through the semantic analysis of text input and subsequent choice of speech output. A focus on
the nature of the Articulated Head’s performance is considered in more
detail later (see section 5.2).

One of the early exponents of AI was Alan Turing. The Turing test measures a
machines ability to demonstrate human-like capabilities in performance,
such that a human is unable to distinguish between the machine and a
human in that specific aspect of the machines performance.

The Turing test was conceived using a text only system (see
https://en.wikipedia.org/wiki/Turing_test) and “seems to provide a scientific,
objective, criterion of what is being discussed — but with the rather odd
necessity of 'imitation' and deceit coming into it, for the machine is obliged
to assert a falsity, whilst the human being is not”.(Hodges, 2012)

The deceit cited in the paragraph above is not unusual in scientific
experiments. It is often necessary to conceal the truth in experiments that
involve human participants, in order to maintain the integrity of the results
(by trying to reduce the influence of demand characteristics, strategies and
expectations). Examples of this practice exist in many published experiments
such as that of the Reeves and Nass experiment cited later in Section 2.3.1.
The reason for raising this point is that this investigation identifies the existence
of a step-function in audience engagement during interaction with the
Articulated Head. The said step-function has been titled “The Freakish Peak”
phenomenon, because it has a similar but opposite step-function effect to
that of Dr Mori’s so called “Uncanny Valley theory” (see section 2.4.3). The
step-function effect of the Freakish Peak phenomenon, when shown in a
graphic representation of the effect, creates a peak rather than a valley in
the graph (see sections 2.4.2, 2.4.3 & Figure 2-1). The word “Freakish” was
chosen because of the astonishment the interacting audience experiences
at the Freakish Peak and because the terms “Uncanny” and “Freakish” are
interchangeable; a test was employed to see whether the effect of The
Freakish Peak phenomenon was repeatable. This test necessarily concealed
a truth about the interaction from the interacting audience (see 2.4.2, 2.4.3 &
Figure 2-1).
The Articulated Head’s performance was designed to give an interacting audience the perception that it was aware and thinking, that the machine, (the Articulated Head) was intelligent. Cognitive sciences study the cognition of the human mind\(^2\) and use the term “Strong Artificial Intelligence” (Strong AI) to refer to a view that the cognitions of the human mind can be modeled and recreated as a set of inter-functioning computer programs. The design and functions of the Articulated Head and its embodied conversation agent did in many ways show adoption of these assumptions of Strong AI. However, investigations into human machine interaction and the “problem of human machine communication” (Suchman, 1987, 2007) have presented “challenges to the assumptions of ‘strong’ AI and Cognitive Sciences” (Duguid, 2012, p. 4) by arguing that the planning of an end-to-end interaction within the context of a human machine communication, where the assumption that the situated actions of humans in that interaction would naturally follow the interaction plan are misguided. According to Suchman it is not possible to control all the variables that might enter into the context of an interaction flow, and she showed through her research that rigid plans of interaction designers are often challenged by the subsequent unpredictable actions of both machines and humans during interaction flows. “In other words, while projecting the future is useful, those projections rely on future activity that the projections themselves cannot fully specify” (Radics, N.D, p. 1). If Suchman is correct, then end-to-end micro dynamic interaction design of the Articulated Head’s conversational agent and interactive performance with its active audience is not possible, therefore one must look for other ways in which to show enhancement of engagement in interaction between humans and machines.

Suchman’s work can be traced back to Hubert Dreyfus and Martin Heidegger. Heidegger was a significant exponent and figure in the development of the school of phenomenology. Phenomenology forms a significant part of the underpinning methodological framework (see section 2)

---

\(^2\) The term ‘the human mind’ should be taken to mean the embodiment of the human biological brain matter and all the influences that this physical situation may bring to bear upon perceptual experience and the exercise of free will therein.
3) from within which this investigation is conducted. “Through Suchman, Heidegger has provided central tools for the critique of AI and cognitive Science and the general understanding of human-machine interaction and communication.” (Duguid, 2012, p. 7). An exposition of phenomenology comes later in this document (see section 3.3) but as a brief introduction, phenomenology involves the study of human experience of existence as perceived within the context of consciousness. A known phenomenon in the study of human machine interaction is that of personification.

### 2.4.1 Anthropomorphism

Personification describes human attribution of human-like characteristics and capabilities to a non-human entity such as a machine. The human attribution that takes place in personification is called anthropomorphism. A study was conducted by Byron Reeves and Clifford Nass titled “How People Treat Computers, Television, and New Media Like Real People and Places” in 1996. In Nass and Reeves (Reeves & Nass, 1996b), a chapter titled “The Media Equation” details an experiment which showed that humans are predisposed to being polite to computers in much the same way as they exchange pleasantries with other humans. This anthropomorphic stance is an interesting phenomenon. The experiments showed that “when a computer asks a user about itself, the user will give more positive responses than when a different computer asks the same question” and “because people are less honest when a computer asks about itself, the answers will be more homogeneous than when a different computer asks the same question” (Reeves & Nass, 1996b, p. 21). In this experiment the research participants had no idea what the experiment was testing, and the concealment of the true purpose of the testing from the participant was purely for the purpose of retaining the integrity of the result. The reason for mentioning these experiments is that the human approach in assuming the anthropomorphic stance in interactions with a machine plays a significant role in the findings of this study. The Freakish Peak phenomenon identified earlier (see section 2.4) appears to be triggered by a human’s heightened perception of agency in the Articulated Head’s conversational agent performance (see 2.4.2, 2.4.3, Figure 2-1 & 9.2). A test was employed to see
whether the Freakish Peak phenomenon was repeatable (see section 7.3.4 page 183,184,185) and this test employed a similar concealment.

2.4.2 Perceptions of agency

Human perception of agency involves the imputation that a person or artifact can act to achieve a goal with autonomy. Humans can (and often do) attribute agency to machines and robots. There is a hypothesis emergent\(^3\) from this research investigation that I have called “The Freakish Peak”; this hypothesis is intimately connected with the perception of agency. The reasons for choosing this particular term are explained earlier in section 2.4. The Freakish Peak hypothesis posits that the degree of perceived human agency attributed to the machine by a human, has a threshold at which “suspension of disbelief” (Reeves & Nass, 1996a) becomes relatively automatic. A belief that the machine has real human capabilities such as sense and thought processing is triggered once this threshold has been crossed (See figure 2-1 ‘The Freakish Peak’). The design refinement of the Articulated Head presented in a sequence of figures in Section 9 is specifically aimed at enhancing the Articulated Heads performance so that people’s perception of it more readily reaches this threshold, thereby triggering suspension of disbelief in its interacting audience. The design refinement directly addresses the research questions at the centre of this study (see page 1). In what follows, I outline some ideas concerning human responses to artificial agents.

2.4.3 Human emotional response to virtual agents

In 1970 Mori put forward the so-called “Uncanny Valley” hypothesis. (Mori, 1970) The theory posits that visual representations of characters that were designed to be human-like, such as those used for films or robots can have a stage in their character development where they fall into the Uncanny Valley. This Uncanny Valley is a stage in the character development, close to

---

\(^3\) The meaning of the term "emergence" such as in the phrase “there is a hypothesis emergent from this research” is more aligned with the common use and meaning of the word rather than the technical phenomenological use of the term such as in Heidegger’s use of the word, which refers to the revealing of the various modes of “Being in beings”, the details of which are beyond the scope of this thesis as imparted and explained in more detail shortly (see Section 3).
the point at which the character is approaching the threshold of being convincingly human, where the character actually becomes much less humanlike and can repel humans, possibly because of a sense of fear or disbelief at perceived deformities. The Uncanny Valley theory is seen demonstrated in contemporary films such as Beowulf (Warner Bros, 2013a). The Articulated Head, whilst showing some human-like features, definitely did not fall into the Uncanny Valley category in terms of its visual presentation, as it was very obviously a machine. However, humans interacting with the Articulated Head did have emotional responses (see section 7). Dr Mori’s Uncanny Valley hypothesis is graphically represented in a diagram in a paper titled “Subjective ratings of robot video clips for human likeness, familiarity, and eeriness: An exploration of the uncanny valley” (MacDorman, 2006), where MacDorman plots the theoretical relationship between familiarity and human-likeness. The Freakish Peak phenomenon (see Figure 2-1 below) shows a close but inverse correlation to Dr Mori’s Uncanny Valley hypothesis ((MacDorman, 2006)), in that the graphical representation shows a peak rather than a valley in the diagram. The Freakish Peak is represented in Figure 2-1 below:

**The Freakish Peak Phenomenon**

[Diagram showing the Freakish Peak phenomenon]

*Figure 2-1 The Freakish Peak*
This investigation’s emergent hypothesis is that there is a threshold in the performance of human-like avatars and robots that incorporate Chatbots. This threshold is where a suspension of disbelief in the human-like capabilities of an avatar or robot can be triggered in a human by their own perception of the degree to which the avatar or robot demonstrates agency – just beyond this threshold the human momentarily has belief in the human-like capabilities of the robot.

2.5 The Original Plan and how this investigation differs

Referring back to section 2.2.2 and the projected evolutionary development of the Head, embodiment, performance, interaction and evaluation all feature in various sections of the scholarship detailed herein, and these sections can be mapped back to aspects of the original plan in some respects - but they do differ from the original plan in important ways, for example: Embodiment in the original plan focuses on the Embodied Conversational Agent (ECA), whereas this study, though not ignoring that aspect, is much more focused upon the implications brought to bear upon the experience of human ↔ machine interaction as the result of the embodiment of the human mind\(^4\) participating in this human ↔ machine interaction. In terms of evaluation, rather than using controlled experiments on behavioural interactions between the Articulated Head and research participants to drive Articulated Head development, this study focuses on understanding the participant experience of human ↔ machine interaction in a relatively unconstrained public exhibition space in order to gather an empirical body of evidence, which is then subjected to an Interpretive Phenomenological Analysis\(^5\). This process has been adopted in order to address the key research questions designed to help answer the big question: how can interaction between humans and machines be improved? (with a specific focus on this particular interactive environment).

\(^4\)The term ‘the human mind’ should be taken to mean the embodiment of the human biological brain matter and all the influences that this physical situation may bring to bear upon perceptual experience and the exercise of free will therein.

\(^5\)Interpretive phenomenological analysis is a qualitative research approach committed to the examination of how people make sense of their life experiences – and what happens when the flow of lived experience takes on a particular significance for people” (Smith, J.A. Flowers, P. & Larkin, M, 2009)
To define exactly what constitutes an ‘improvement’ in terms of human ↔ machine interaction is difficult and subjective in nature, as what is perceived by one human as an improvement may not be by another. Any change made to an interaction design and the surrounding interactive environment will alter the nature of the interaction-taking place. Each change will involve trade-offs where improvements made in one aspect or area of the interaction are made at the expense of other aspects and areas of that same interaction. Therefore a definition of the term ‘improvement’ is not possible in this context. However, any proposed changes to the interaction design and interactive environment in which this human ↔ machine interaction takes place should cater ‘to and for’ natural flow in human ↔ machine interaction, and should remove or circumvent barriers and obstacles to interaction flow that have been identified through research. Improvements’ should make interactions user-friendlier by making the machine receptive to user needs.

2.5.1 A limitation of this investigation

Many of the outcomes and related ideas emergent from this investigation, presented as part of the data analysis and conclusions to this work, are also relevant with regard to human ↔ machine interaction in a range of other different and varying contexts. However, it should be noted here that the Articulated Head, whilst capable of movement in its immediate three dimensional space, was essentially a situated robotic exhibit; it was hinged to the ground and not horizontally mobile from the base, i.e., it did not have feet or wheels - and this restriction or constraint of the interactive environment under investigation in this study, will render some aspects of the findings from this investigation irrelevant to human ↔ machine interaction where this restriction or constraint is not a feature of the machine or robot in question in other interactive environments.

2.5.2 Projected incorporation of interactive scenarios

The research plan E-Appendix 2: From Talking Heads to Thinking Heads proposed to evolve the Prosthetic Head (Livesey, 2012) into a Thinking Head (MARCS, 2012), by the incorporation of scenarios with projective statements such as:
“1. A customer service bill-enquiry/complaint scenario, initially with limited vocabulary and options, will allow increasing sophistication over successive iterations to enable user-driven provision of information. This test-bed will stretch understanding of communication beyond simple information transfer: repeated interactions will require maintenance of user models, including recognizing and acting upon contexts in which particular interlocutors may express frustration.

2. A second language (L2) acquisition instruction scenario will explore the increasing sophistication of dialog capabilities in a task that will allow evaluation of children and adults in a practical learning situation, where language is both the research focus and of the educational domain. An immersive language learning situation will explore whether students learn more effectively with or without the Embodied Conversational Agent (ECA); and with or without effective, shared experience, and multimodal communication”.

Not all of the originally projected aspirations in the research plan came to fruition during the development of the Articulated Head. However, many aspirations did bear fruit, and although the project and associated research evolved in slightly different ways to those originally anticipated, the original projections and final outcomes from the project do display considerable alignment.

2.6 Technical Components of the Articulated Head
An understanding of the Articulated Head’s technical components and operational capabilities is important to what follows in this investigation because the extent to which these features defined the latitude for agency afforded the Articulated Head is relevant.

The Articulated Head hardware included a six-degrees of rotation industrial robotic arm with a standard flat screen computer monitor mounted on the end effector. The computer screen displayed a three-dimensional representation of the head and face of Stelarc;
The Head was capable of spatial movement via the robotic arm, speech and a small variety of facial expressions. The Articulated Head was surrounded by an aluminum framed glass enclosure in each of the exhibition spaces it occupied during the course of this investigation. The enclosure was a health and safety measure requested and approved by the relevant health and safety authorities at both the University of Western Sydney and the Powerhouse Museum respectively in order to protect any audience members from injury that could be caused by the industrial robotic arm’s movement in the public exhibition space.

A commercially developed software conversational agent called the ‘ALICE ChatBot’ ("Artificial Linguistic Internet Computer Entity," 2012) handled speech, in avatar communication with its audience. The Head was also furnished with some sensory capabilities both auditory and visual in nature, which allowed it to track various elements pertaining to its possible audience.

An attention model, invented and programmed by Dr Christian Kroos utilising a software programming environment called MatLab (MATLAB, 2012) used sensory information, probabilities, weightings and thresholds to define and control motor goals relating to various behavioural aspects of the Articulated Head’s movement by directing the industrial robotic arm within the interactive installation enclosure.

A carefully designed text to speech engine drove animation of the facial, jaw and mouth movements of this three-dimensional representation, based
on the phonetic structure of the text to speech content input. By triggering a prerecorded library of facial point animation maps simultaneously with a correlating phonetic audio sample library, the system provided an interacting audience with what many visitors to the Articulated Head perceived to be fairly convincing synchronization of facial, jaws and lip movement with the auditory speech output. It should be noted here that Head zero did not have a co-articulation model. This meant that every articulation of a phoneme was the same no matter the context it was in.

An independent stereo camera tracking system established audience locational information, which was utilised by the Thinking Head Attention Model & Behavioural System (THAMBS). THAMBS - programmed in MatLab (MATLAB, 2012) by Dr Christian Kroos from The MARCS Institute (Institute, 2012) at the University of Western Sydney (“University of Western Sydney,” 2009) could, for example, manipulate the robotic arm’s position to make the head face the current position of a person in the audience. The intention of including such functionality was to make it appear that the Articulated Head had turned to focus its attention upon a person in the audience. The computers system as a whole had a component based system architecture and involved an Event Manager (EM), which played a central role in coordinating the Articulated Head’s functions. The conversational agent ALICE (“Artificial Linguistic Internet Computer Entity,” 2012) Chatbot, was installed as part of the Articulated Head installation. The ALICE Chatbot allowed an interacting audience to converse with the Articulated Head using text, entered via a keyboard mounted on a kiosk just outside a glass safety barrier that separated the robotic arm from its interacting audience (see Figure 2-4 below). THAMBS operations were visualised on the kiosk screen with the purpose of giving a person typing at the keyboard a visual representation of what the THAMBS system was doing.
The Articulated Head was presented in the Powerhouse Museum as an Embodied Conversational Agent and Artificial Intelligence (AI) - A Thinking Head (TH). The exhibit title, labeling and advertising material that promoted the Articulated Head’s presence in the museum all contributed to the suggestion that the Articulated Head was capable of thought and learning, that it was capable of exhibiting attributes of an existential being. The text presented to visitors of the museum exhibit informed patrons that they could ask the Articulated Head any kind of question and that he would reply, that he could converse with them and they could hold a conversation with him.

2.7 Operational infrastructure, maintenance and control of the Articulated Head Systems (Overview)

Dr Damith Herath, a research engineer working on the project, was responsible for several sections of software coding and the hardware control systems of the Articulated Head. Dr Herath provided an overview of the Operational infrastructure, maintenance and control of the Articulated Head Systems in the document provided as electronic documentation in E-

---

6 The Articulated Head is referred to as “he” rather than “it” here to reflect and further illustrate the bias of the exhibit labeling that did indeed refer to the Articulated Head as “he”.

---
Appendix 3: AH_QuickStartGuide. The guide provides diagrams of the hardware setup as implemented in the Powerhouse Museum.

2.8 Software component based architecture

Figure 2-5 above shows a flow diagram of the operational framework of the Articulated Head as implemented in the Powerhouse Museum. The diagram does not show every aspect of the operating system environment - the creative additions, (see section 6) are not included in this flow diagram. Some of the modular software framework shown as rectangular boxes in the flow diagram in Figure 2-5 above, in particular the Data Logger, Sonar Proximity Client, Audio localizer, and the Face Tracker boxes, were in fact only partially operational or nonoperational for the majority of the duration of this study. The reasons for partial or non-operation of these modules and their influences, if any, during audience interactions with the Articulated Head, are discussed later under the themed data analysis headings of section 7. Some minor variations to the software component based
architecture of the Articulated Head exhibit were tested at various points throughout the Articulated Head’s presence in the Powerhouse Museum. Although the impact of the said variations upon the main findings of this investigation are thought to have been negligible, based on the observation that dramatic noticeable changes to the interactions taking place did not occur, and research participant reporting of the interactive experience was also not noticeably modulated, where that impact of the variations is thought to have been a contributory factor modulating important aspects of the analysis of data in this study, the contributory factor is mentioned and any possible modulatory influences upon the findings related to the particular section of the data being analysed are discussed.

2.8.1 The Thinking Head Attention Model and Behavioural System.

The Articulated Head required a behaviour system to control its physical actions when interacting with its audience. The plan was to provide a behavior system that would act as an attention model, defining how the Head would appear to pay attention to its interacting audience, this system was named THAMBS. THAMBS is an acronym for the Thinking Head Attention Model and Behavioural System. Dr Christian Kroos who was a research fellow working on the Articulated Head project described THAMBS in the document E-Appendix 4: THAMBS documentation as “biologically inspired control software implemented in MatLab (MATLAB, 2012) for robotic platforms with sensing capabilities.”. The document also provides a list of publications, which impart theoretical and technical considerations and implementation details related to THAMBS development.

In essence THAMBS consisted of a small number of sub-system routines that collected information from system sources such as the co-ordinates generated by the industrial robotic arm and those values generated by the auditory and visual sensors. The auditory and visual sensors consisted of a stereo tracking camera and a stereo microphone set-up, which helped indicate the position of a person in the audience within the Articulated Head’s sensor pick-up fields. The THAMBS system used the gathered endogenous system and sensory information to process a perceptual event. The perceptual event subsequently defined an attended event, which then
had a motor goal attached. The attended event was then processed to generate a response from the Articulated Head. The arm appeared to direct its focus towards the attended event (i.e. the person in the audience who was within the Articulated Head’s sensor pick-up fields) because the appropriate motor commands were sent from THAMBS to the industrial robotic arm to move it into the rotational, vertical and horizontal position which pointed towards the motor goal (the coordinates of the person in the audience who was within the Articulated Head’s sensor pick-up fields). The range of parameter values collected by THAMBS were subjected to a number of filtering algorithms, which used attributes such as probability, weightings and threshold values to determine the degree of confidence with which an appropriate motor goal could be identified, in order to make decisions about which events the THAMBS system selected to pay attention to. THAMBS thresholds were intentionally used for the reduction of the rapidity of the Articulated Head’s responses to potentially attended events, especially when the number of potentially attended events increased, such as when there were many people present within the sensor pick-up fields at the same time. This reduction in rapidity of response was necessary because the rapid switching between one potentially attended event and the next, resulted in an erratic and somewhat disjointed behavior exhibited by the Articulated Head unto its audience. This disjointed behaviour was detrimental to engagement because it made it appear that the Head was easily distracted and notionally not interested in the person who was interacting with it. Since engagement was a performance goal, the said disjointed behaviour, as an outcome was undesirable. This observation is discussed in more detail in section 7.3.7 under Theme 7 Engagement.

2.8.2 The Event Manager (Overview)

The Articulated Head’s Event Manager, programmed in JAVA by Dr Damith Herath and Zhenghi Zhang as indicated in Figure 2-5 (D. Herath, 2012) had a central role to play in terms of the collection, scheduling, execution and distribution of information throughout the rest of the Articulated Head’s software and hardware systems. The Event Manager was a scheduler, it kept a list of all events and event requests throughput from the systems hardware and software apparatus and executed them in order according to event
priority flags. In this respect the Event Manager was very similar to the event processing systems present in many commercial interactive multimedia software programs such as Max (Cycling 74, 2012) or Adobe Director ("Adobe (Software)," 2012) for example.

### 2.8.3 The EM/Max Interface

The Event Manager/Max interface was designed to pass information out from the Event Manager to the Max-programming environment. Details such as the text strings for both user input, and the Articulated Head ALICE Chatbot response strings were passed to Max over a TCP/IP local area network connection. Open Sound Control (OSC) protocol messages passed data to Max. The creative additions were set up to run on a separate computer and operating system from the Event Manager in order to minimize Articulated Head down time, whilst testing and development of the creative additions took place. The passed data was then used for the purposes of displaying the text or related textual strings in the projections detailed in section 6.8.1. Other data such as the X Y Z transitional position coordinates of the robotic arm were also transmitted from the event manager through the Event Manager/Max interface to Max as and when necessary. These parameter values allowed for the positioning of the robots voice within the spatial auditory system as detailed in section 6 (The Creative New Media Propositions) where a more detailed description of the operational networking protocol, programming and implementation of the Event Manager/Max Interface are given.

### 2.9 Where and when has this investigation taken place?

2.10 The Big Question.
The big question identified in the grant application E-Appendix 2: From Talking Heads to Thinking Heads was:

- How can interaction between humans and machines be improved?

This question will be referred to simply as the big question from now on. The critical importance of the human present in the human ↔ machine interaction referenced in the big question has become the key object of focus within this investigation and the ramifications of this critical focus are presented throughout the following sections of this thesis.

2.11 The Research Questions
The key research questions presented on page 1 and reiterated below were specifically designed to help answer the big question. The key research questions have one main line of enquiry, which draws from the three open-ended subsequent questioning strands (a, b & c). The research seeks to examine avatar, conversational agent and audience interaction; it does so through the following questions:

a. In which ways has and can the avatar be given agency?

b. In which ways is and can the avatar be multi-sensory and multi-present?

c. In which ways can and do the characteristics of a `virtual' performer (e.g., personality, nuance of response, context) condition? affect and engagement?

These questions are reiterated and explored in detail throughout this document and many aspects of embodiment, performance, interaction and evaluation become features and/or concerns of the themed qualitative data analysis (see Section 7). Section 8 presents key concepts that help to explain the findings from the themed qualitative data analysis, and inform

---

7 The word ‘condition’ is used as a verb here; examination of the characteristics of the ‘virtual performer’ and how those characteristics affect the performance and engagement of an active audience is referred to herein as an aspect of conditioning of the interaction.
design refinements put forward in a blueprint of recommendations emergent from the investigation (see Section 9).

2.12 This study’s contribution.
Many roboticists have developed humanoid robots in order to explore human ↔ machine interaction, one of the best known of these developers is probably Professor Hiroshi Ishiguru of Osaka University (“Osaka University,” 2012). Social robots such as Cog and Kismet mentioned earlier, the Home Exploring Robotic Butler (HERB) (Srinivasa, 2012) or Actroid-DER also known as the uncanny valley girl named Yume (Ishiguro, 2006) have been developed to explore human ↔ machine interaction. The Articulated Head was just one such robotic project developed in part, to explore human ↔ machine interaction.

The key contribution that this particular research project makes to knowledge about human ↔ machine interaction, i.e. where this study helps to fill a gap, is that, many investigations of human interaction with robots are focused on the engineering aspect of the robot. The current investigation was motivated from a performance arts perspective where prioritisation of human experience and human perception of the robotic installations performance are used to inform potential improvements to the design of the interactive environment in question. This investigation has explored an audience’s interactive experience with a Robot and Embodied Conversational Agent (ECA) installation in a relatively unencumbered public exhibition space employing the methods of Grounded Theory, Video Cued Recall and Interpretative Phenomenological Analysis (Smith. J.A et al., 2009).

Many of the problems of robotic development are approached from a predominantly scientific and/or engineering perspective, where technical design and development of electronic software, hardware and mechanical aspects of the project become of paramount importance and take precedence over the subjective perception and lived experience of the robots performance as perceived by an interacting audience.
The Articulated Head’s performance and this research in particular could be seen as somewhat nebulous from the scientific engineering perspective but this study does ask pertinent questions that are very often ignored from an engineering perspective, because this study is audience focused. The study prioritises human experience and asks the end user (humans) what they actually think and feel during interactions, what matters to them and indeed what does not.

2.14 What follows?
What follows is a careful build up of all the elements necessary to understand the philosophical approach adopted by this investigation and its contribution to knowledge about how human ↔ machine interaction can be improved in this, and similar interactive environments. Section 3 imparts the methodological framework and method employed in this investigation, followed by a phenomenological reduction and comparison of human ↔ human and human ↔ machine communication (see Section 4). Section 5 purposefully presents an analogous performance arts scenario elucidating the interactive and performance focused framework within which this investigation has been situated. An electronic Appendix also considers the creative new media arts investigative terrain within which this enquiry has been situated, through the presentation of an analogous road traffic accident scenario.

The following text documents a phenomenological investigation of the human ↔ machine interaction that has taken place between an audience and the interactive robotic exhibit that was called the Articulated Head. A critical focus on the meaning of the word interaction is absolutely crucial in all that follows. The aim of this investigation was to develop and impart a clear understanding of the human ↔ machine interaction under investigation and how it might be improved, and to test and evaluate two artistic propositions - creative new media projects, which were designed to enhance engagement of the audience in this human ↔ machine interaction. The Articulated Head installation proved a maeutic environment enabling the testing of the following artistic “propositions”: 1. The inclusion of some textual projections: 2. A spatial auditory system. Section 6 provides the
rationale for these artistic propositions and the technical details of their execution.

2.14.1 Defense of the Articulated Head’s performance

In defense of the Articulated Head, it was successful as an exhibit and it should be made clear here that there is no suggestion that the Articulated Head did not engage its audiences, indeed it did – but the question here is to what extent? - and how can audience engagement in this type of human ↔ machine interaction be improved upon? To answer these questions one must inevitably approach the enquiry with critical analysis based on the assumption that there is scope for improvement in this human ↔ machine interaction. Therefore the audience engagement - or lack thereof, in the human ↔ machine interactions that took place between the Articulated Head and its public audience should be considered the main focus of this thesis - as implicitly identified as the subject under investigation in the big question. The design refinement presented in section 9 which is informed in part by the Freakish Peak hypothesis (see section 2.4.3) should be considered as the primary outcome of this investigation.

2.14.2 The Problem

Human ↔ machine interaction and how it might be improved is a huge entangled web of interrelated issues or strands tied up in a complex knot. This knot is well represented in the interactions that have taken place between the Articulated Head and its public audience. Therefore what follows is effectively the gradual untying of the knotted issues and strands into what might upon first impressions appear to be loose ends - but then, having separated out these strands to help inform an interpretive narrative and refine a clear understanding of what was actually happening in this interactive environment, a strategic vision of what might help improve interactions emerges from the interpretation, presentation and analysis of the empirical human centered evidence collected. The findings from this data analysis then lead to something of a re-platting of all these strands, into what might be considered a blueprint, a group of outcomes and related ideas emergent from the investigation that show how interaction between
humans and machines can be improved within the context of this, and similar interactive environments.

A concerted effort has been made to cross reference research data to emergent themes in the data analysis. Phenomena identified through data analysis are examined in conjunction with the presentation of key concepts and theoretical perspectives that support their exposition. These expositions then lead to emergent ideas as to how this type of human ↔ machine interaction can be improved. The relevance of the information presented for consideration in this investigation is also rigorously indicated throughout the work that follows.
3. Methodology and methods

The methodological framework chosen in order to gain a greater understanding of the Articulated Head’s audience and their interactions with the robotic exhibit was that of Grounded Theory and Phenomenology. The methods employed include a Phenomenological Reduction and comparison of human ↔ human compared to human ↔ robot communication, Video Cued Recall Interviews and Computer Assisted Qualitative Data Analysis. These methods were selected for both practical and philosophical reasons.

3.1 Grounded Theory

The research data collected in this study related to Articulated Head audience interactions comes in a variety of forms such as observational records, notes and diary entries. However, most of the data collected and analysed was in the form of either recorded text string exchanges between the Articulated Head and its interacting audience or video recordings, which captured the audiovisual environment and as much of the action from both the Articulated Head and the interacting audience as possible, including the research participants cued reporting in a Video Cued Recall Interview (explained in detail later). This data was collected with one specific application in mind: It was analysed and coded in a computer assisted qualitative data analysis software application called NVIVO, for the purposes of building an understanding of the experience of avatar, conversational agent and audience interaction. The data forms an empirical body of evidence laden with clues, which inform, generate and support theories and recommendations as to how one might enhance an audience’s engagement with this type of interactive exhibit, and ultimately find answers to the big question: How can interaction between humans and machines be improved?

Grounded Theory is a method of investigation developed by Barney Glaser and Anselm Strauss (Glaser, Strauss, & Strutzel, 1968) The method was developed when they worked together on a sociologists project, researching dying hospital patients. Glaser and Strauss later came to a
disagreement over the practical details of how to conduct a Grounded Theory study. Their disagreement led both to publish books effectively arguing their case and detailing the method, as it should be conducted from their own perspectives respectively. "The controversy between Glaser and Strauss boils down to the question of whether the researcher uses a well defined 'coding paradigm' and always looks systematically for 'causal conditions,' 'phenomena/context, intervening conditions, action strategies' and 'consequences' in the data, or whether theoretical codes are employed as they emerge in the same way as substantive codes emerge, but drawing on a huge fund of 'coding families'" (Keele, 2005). When considering competing strategies in qualitative research Amedeo Giorgi weighs up the benefits of description versus interpretation arguing “that the interpretive approach is appealing under certain conditions that have not always been fully specified and that certain other conditions lend themselves to other qualitative research practices.”(Giorgi, 1992) Many studies have been conducted since employing one, other or both of these strategies and regardless of which approach is adopted both strategies present both benefits and pitfalls.

Grounded Theory, regardless of the methods actually employed when coding data, has a few defining features. Its main defining feature is one that it shares with phenomenology and the phenomenological reduction. That is, one does not enter into the research with a hypothesis; rather the practice of a conducting Grounded Theory study and/or Phenomenological Reduction will lead to or generate a set of theories. Once data is collected, it is coded to highlight the key points coming out of the data. The codes are put into groupings related to similar concepts and then, from the groupings, themes are formed. These themes are fairly wide groupings of concepts, which in turn lead to a set of explanations about the research subject.

Glaser’s view in application of Grounded Theory is that any and all data related to the subject under investigation that can be gathered by the researcher is valid data. That is, qualitative and quantitative data, treated with respect to all conditions under which the subject under investigation
might be experienced, regardless of whether the data constitutes concrete or anecdotal evidence, it may be included in the investigation into the subject at hand. Data that helps to answer the researcher’s main question, “what is going on here?” and contributes to the set of explanations that emerge is useful (Interviews, observation, notes, questionnaires, video, audio, transcripts, memos).

The process of comparing different types of data over a cross section of research participants and sources becomes the starting point for coding and note taking (memo’ing). Coding themes can emerge quickly and constant comparison of research data helps to solidify and support the fit of theories to data. Theories generated are never right or wrong, they are just a closer or looser fit to the data being analysed. The closer the fit of a theory to the data set, the more relevant it is for cross checking by the rigors of purposeful new sampling. As theories emerge, further sampling of appropriate data types and coding take place to make sure that the emergent theories remain a close fit to the data, if not, then new theories and coding of categories and their associated attributes will be generated. When the fit of theories are supported by new sampling and coding, (i.e. the emergent theories are supported by new data collection and new themes) and the categories and their associated attributes do not expand anymore, then further sampling ceases as the point of theme saturation is reached. This is the point at which sorting takes place. The process of sorting is grouping nodes, notes, and memos into similar types and arranging them in the most appropriate order to render your theories transparent.

Reference to Glaser and the distinction between “emergence and forcing” (Dick, 2005) of theories from data is made recursively throughout literature on the subject. Criticisms of Grounded Theory point to the questions surrounding its validity as a theory and suggest that “it is impossible to free oneself of preconceptions in the collection and analysis of data” (Thomas & James, 2006)
The above criticisms are similar to those leveled at Phenomenology and the Phenomenological Reduction, which also calls for *epoché*, the ‘bracketing off’ or ‘putting aside’ of our presumptions and preconceptions about the experience that is under investigation. In order to unveil essences of the experience of our own ‘*lifeworld*’ that may otherwise be hidden from us, is to temporarily suspend the “sedimented layers of consciousness built up through our temporal experiences” (Cogan, 2012), this is clearly a very difficult, if not impossible thing to do.

The conflict and contradictions between ‘hypothesis testing’ as oppose to ‘emergent’ research methodology and philosophy have been hotly debated elsewhere (Langdridge, 2007). Bob Dick points to the fact that most of us “will have been exposed to hypothesis testing research” (Dick, 2005) and that this indoctrination has to some extent, to be “unlearned” in order not to misjudge Grounded Theory based on “criteria which make sense only for the methodology for which they were developed”. He further points out that Glaser himself “suggests two main criteria for judging the adequacy of the emerging theory from a Grounded Theory study: that it fits the situation; and that it works – that it helps the people in the situation to make sense of their experience and manage the situation better.” (Dick, 2005)

The above criteria fit the methods employed in this investigation, and the methods have worked to help the people involved to make sense of the situation that they were in during this investigation. Furthermore, Glaser’s criteria are a very close fit to the central aims of this research project being; to find ways to enhance audience engagement with interactive exhibits like the Articulated Head, and to find answers to the big question: how can interaction between humans and machines be improved? - in that, the central aim’s main objective is to help manage the situation under investigation (*human ↔ machine interaction*) better.

The key concepts (see section 8), help to explain the findings from the themed data analysis presented in Section 7, and inform a blueprint of recommendations for how interaction between humans and machines can
be improved in Section 9. All the emergent recommendations presented in the blueprint in Section 9 are essentially based on an interpretive phenomenological analysis of empirical human centered research data collected during this investigation.

3.1.1 Computer assisted qualitative data analysis software (CAQDAS)

There are a range of criticisms leveled at the use of computer assisted qualitative data analysis software and the way in which it might influence data analysis (Richards, 1999, p. 418). Some arguments (Bringer, Johnston, & Brackenridge, 2006, p. 248) point to the problem of conducting simple quantitative word searches on qualitative data and then misrepresenting this as qualitative analysis, whilst others comment that researchers may choose their theories and methods to suit a particular computer assisted qualitative data analysis software program to make their job easier. Although it is probably true that the use of software in coding qualitative data makes the retrieval of information related to specific codes and the general organization and sorting of data easier, the expansion of the amount of data that can be captured and analysed has an equal and opposite effect where good rigorous qualitative data analysis is concerned. The way that software tutorials can indoctrinate and steer young researchers towards specific modes of operation and choices of method without proper regard for other equally or possibly more appropriate methods of investigation related to their specific area of investigation has also been questioned (Richards, 1999, p. 412).

However, the key advantages of computer assisted qualitative data analysis software lie in the way in which the software has been developed to support “a weaving of rich primary sources with commentary and discussion and analysis” (Richards, 1999, p. 414). Lyn Richards goes on to explain, “If this is done well, then the distinction between data and commentary disappears as the threads of a weaving lose distinction. They remain separate threads but contribute collectively to colour and patterns and the construction of new understanding, the primary qualitative goal” (Richards, 1999). Computer assisted qualitative data analysis software can be used to help manage code and sort large amounts of data retaining the flexibility to view, sort and
ask complicated questions of the data and coding and obtain answers quickly. Paper based research becomes far too cumbersome and can be very slow to facilitate similar results, indeed time and financial constraints may make some questions that are easily answered through digital analysis virtually impossible in a paper based research project.

3.1.1.1 Axial Coding
Axial coding is the separating out of the component parts of core themes during qualitative data analysis. Axial coding is a mixture of inductive and deductive reasoning that helps inform the process of relating the various concepts and categories that have been identified in the coding process in a Grounded Theory study, to each other.

“Strauss and Corbin (1990, 1998) propose the use of a ‘coding paradigm, to include categories related to (1) the phenomenon under study, (2) the conditions related to that phenomenon (context conditions, intervening - structural- conditions or causal conditions), (3) the actions and interactional strategies directed at managing or handling the phenomenon and (4) the consequences of the actions/interactions related to the phenomenon”.
(Wikipedia, 2006)

3.1.1.2 Frames analysis
Frame analysis is not a single concept but a “wide range of approaches that have been subsumed under the heading of frame analysis” (Koenig, 2004). Others have presented an overview of these approaches: (Benford & Snow, 2000): (D'angelo, 2006): Scheufele 1999)

Frame analysis terminology includes “Narrative fidelity” and “Empirical credibility” which are cited as “mechanisms that render frames particularly viable” (The Cathy Marsh Centre for Census and Survey Research, 2013). Particularly frequently occurring frames are labeled “Master frames” and Master frames often suggest culturally poignant themes. Narrative fidelity is defined, as “the congruence of a frame with the life experience of its addressees” (Gamson & Modigliani 1989) (Koenig, 2004).
Empirical credibility is a measure of how well a frame appears to fit with the detail of actual events from a subjective rather than objective point of view. Empirical credibility essentially “denotes the ease with which audiences reconcile a frame with what they consider their possibly mediated experiences” (The Cathy Marsh Centre for Census and Survey Research, 2013). One could say that empirical credibility has subjectively been attributed to the philosophy and methods adopted in the pursuit of this investigative research by the researcher, in that; he comments within the body of this text that the philosophy and methodology adopted fits well with his previous thought on the process of investigation and the pre-established belief that understanding should be an inductive and deductive process of discovery.

The terms “conflict”, “human interest” and “economic consequences” are frames that are frequently occurring in media discourse and in particular, the news. Frames have been described as; “basic cognitive structures, which guide the perception and representation of reality. On the whole frames are not consciously manufactured but are unconsciously adopted in the course of communicative processes. On a very banal level, frames structure which parts of reality are noticed.” (Koenig, 2004) and; “principles of selection, emphasis and presentation composed of little tacit theories about what exists, what happens and what matters (Gittin 1980 p. 6)” (Koenig, 2004). There has been a shift from subconscious to consciously adopted and identified frames. Michael Polyani contributed much to social science and philosophical discourse surrounding tacit knowledge, “central to Michael Polanyi’s thinking was the belief that creative acts (especially acts of discovery) are shot-through or charged with strong personal feelings and commitments (hence the title of his most famous work Personal Knowledge)” (Smith, 2003). The construction of tacit knowledge during interaction and subsequent the actions of some of the humans interacting with the Articulated Head, when considered within the frame of its performance and the perceived agency therein - as experienced by those humans, led to the Freakish Peak hypothesis emanating from this study.
3.2 Video Cued Recall

Video Cued Recall was chosen as an appropriate method for capturing the lived interactive experience of research participants interacting with the Articulated Head for a number of reasons, as detailed below.

Reporting Experience

One of the clear and valid concerns about how research in the arts elucidates direct experience relates to the methods associated with gathering reports of individuals’ personal experience. At issue here is the accuracy of such reporting, in terms of how it correlates to the actual experience during interaction with the exhibition or performance work, as distinct from a re-imagining, or reconstruction of that experience from memory. Such reconstructions from memory could be substantially faulty, because the individual may be so engaged/consumed by the art experience (abstract and often subliminal) at the time of engagement that noting details or even qualities of that experience in parallel with having the experience itself cannot occur. This is particularly true in situations where no information or priming has occurred prior to the interaction with the artwork. Simultaneous reflection and engagement clearly also describe a split experience, which would not correlate to a non-research participants art experience.

One approach to managing the above issue(s) is the technique, Video Cued Recall (VCR). Video Cued Recall involves making an audiovisual recording of the participant during a direct experience of the artwork. The recording should in no way interfere with the participant’s ability to engage with the work. The researcher undertaking the Video Cued Recall session should not provide the participant with any information that is not available to the general public, or prime them for their forthcoming experience of the artwork in any way. Once the participant has completed their engagement with the work (which must not be constrained by the researcher), the researcher takes the participant to another space. During this session the video of the participants engagement with the artwork is played back to them very shortly after their interaction. Participants are then asked for their
thoughts about the experience of engaging with the artwork, including reflection on their reactions during the interaction.

This reporting may be augmented by the researcher asking the participant to describe: how they were feeling at the moments reflected in the video recording? What they were doing? The nature of their engagement with the artwork? And other associated questions as appropriate to the research being undertaken. The response from the participant was therefore cued, or contextualised by what they were viewing in the video playback. The video playback can be paused, rewound and replayed at any time during the recall session so that the participant can review or revisit moments that occurred during interaction as often as they wish. Such pausing and review is often necessary due to the fact that the participant has a considerable amount to say about particular moments of interaction, where the interaction may, in and of itself, only take a short period of time.

Video Cued Recall is used in ethnography and other related humanities research areas where illuminating direct experience is a principle objective. The method of Video Cued Recall addresses the phenomenological questions at the heart of this investigation - that is, evaluation of interaction. It allows discussion about the response and to some degree avoids the fact that self-reporting is difficult to interrogate. The video acts as a context for review and reflection.

Raingruber states that Video Cued Recall is “a research approach for accessing relational, practice-based, and lived understandings” and furthermore that “the goal of the phenomenologist is to develop direct contact with lived experience and to bring to light the meanings woven into the fabric of the Lebenswelt” (Raingruber, 2003, p. 1155)

It is possible, that during the recall session, the subject may react to what is viewed on the videotape rather than recalling the actual experience as it occurred at the time of the recorded episode (Lyle, 2003). As such, participant recollections may not represent the conscious or unconscious
cognitions taking place at the time of the video taped episode (Lyle, 2003), however the ability of the researcher to interrogate the responses during the session, assists in deliberations as to the degree to which the response is immediate or constructed.

Phenomenological philosophy and the practical method of Video Cued Recall are intrinsically linked in focus upon capturing the conscious lived experience of a person within a specific scenario. Reflecting on the fact that this investigation is a process of discovery and does not pursue the scientific method of establishing hypotheses prior to the investigation because the research method used herein seeks to “bring to light the meanings woven into the Lebenswelt” (Raingruber, 2003, p. 1155) rather than to test whether specific meanings exist, render Grounded Theory and Phenomenology the ideal methodological framework, and Video Cued Recall an excellent method for ‘getting at’ the essences of the audience’s experience of interaction with the Articulated Head. “A phenomenological methodology will involve the collection of naturalistic first person accounts of experience, recognizing the need to account for the influence of the researcher on the data collection and analytical process.” (Langdridge, 2007, p. 87) Video Cued Recall (VCR) sessions were an ideal way of capturing the lived experience of interactions with the Articulated Head and subsequent reports made by the person who participated in the interaction, with very little influence from the researcher being brought to bear upon the reports given.

3.2.1 Expanded discussion of applications of Video Cued Recall

Video Cued Recall is used in many situations such as sport, health, crime investigation and educational teaching research.

Described as “a research approach for accessing relational, practice-based, and lived understandings” (Lyle, 2003, p. 875). Video Cued Recall has disadvantages, for example; videotaping can influence participant behaviour. The “possibility that the subject is reacting to what is viewed on the videotape rather than recalling the taped episode” (Lyle, 2003, p. 863) can cause problems. As quoted previously “Participant recollections may
not represent the conscious or unconscious cognitions taking place at the time of the video taped episode”. (Lyle, 2003) Video Cued Recall has been used in crime investigation situations and an abstract from an article in Psychological Science titled ‘Recalling a Witnessed Event Increases Eyewitness Suggestibility: The Reversed Testing Effect’ (Chan, Thomas, & Bulevich, 2009) shows “that immediate cued recall actually exacerbates the later misinformation effect for both younger and older adults”, thus Video Cued Recall can have adverse effects upon the progress of certain types of investigation.

However, Video Cued Recall as referenced in a qualitative health research paper “benefits from minimal intervention in the activity” (Raingruber, 2003) and provides “access to nonverbal influences afforded by video taping”, which “draws participant comments on; facial expressions, voice tones, posture shifts, spatial relationships, levels of intensity, silences and overall pace that would otherwise not have been noticed or remembered”. (Raingruber, 2003, p. 1165)

Video Cued Recall allows the researcher to code using qualitative data analysis techniques by accessing the material and coding for different categories on each pass, this is especially useful in the analysis of events that are never twice the same such as the nuances of a winning finishing stroke in a tennis match or a pivotal moment in a marriage counselling conversation.

Depth of reflection in the Video Cued Recall interviews “can yield much richer commentary than the use of probing questions” (Raingruber, 2003, p. 1160) or the use of questionnaires and other feedback instruments. “Experience is best described not with words but by embodied responses, practices, sensory perceptions and tones or climates of feeling” (Raingruber, 2003, p. 1168) “Video cued narrative reflection puts participants back into their experience while allowing them sufficient distance to recollect their thoughts and feelings” (Raingruber, 2003, p. 1168)

Just as Video Cued Recall is used in sport, education and teaching research to gain a greater understanding of - and to reflect upon the nuances of
events taking place in these situations, developing a stronger understanding of the lived experience of the person interacting with the Articulated Head, has given the best chance of understanding what facets represent barriers to fluidity in communication, engagement and interaction with the Articulated Head. Having understood more about the barriers that existed, one can then begin to hypothesise and make prescriptions designed as antidotes to the barriers identified, hence the emergent Freakish Peak hypothesis (see section 2.2.2) and The Blueprint of Emergent Recommendations presented as part of the conclusions to this investigation in Section 9. “Phenomenology acts as a framework for the method of Video Cued Recall” (Paine 2010) and “the goal of the phenomenologist is to develop direct contact with lived experience and to bring to light the meanings woven into the fabric of Lebenswelt.” (Raingruber, 2003)

3.2.2 The best time to implement the interview

Video Cued Recall interviews are normally conducted via presentation of video recorded material to the participant accompanied by open-ended question/s. Questions should be presented “to the subject as soon as possible after or during the viewing of the videotape” (Lyle, 2003, p. 863)

The use of Video Cued Recall for reporting teacher interactive thinking in an educational environment shows; “retrospective accounts of thinking may involve immediate interference and generative processes. The stimulated recall video tape produces a ‘new view’, which is subject to the meta-analysis and reflection that was not available to the individual at the time of the original episode” Yinger (1986) as cited in (Lyle, 2003, p. 864). Lyle then references Gass (2001) who acknowledges, “There is evidence of increasing recall decay with consecutive, delayed and non recent protocols”.

For the reasons above, interviews in this study were conducted directly after the original videoed events wherever possible to avoid the influences of “recall decay” (Lyle, 2003, p. 864) and to minimize the time allowed for meta-analysis and reflection prior to the participants interview. The “evidence of increasing recall decay with consecutive, delayed and non recent protocols” (Lyle, 2003, p. 864) provides some ratification of the
secondary process of using repeated recall participants at various stages in the creative project development and installation whenever possible. It was anticipated that using repeat participants for the purposes of the evaluation of the impact of the creative outcomes proposed might help - as a time gap between each iteration of their Video Cued Recall sessions allowed for significant recall decay, yet repeat participants would have some sedimentary layers of experience to draw upon, which was expected to help them identify changes made to the exhibits performance between recall sessions.

### 3.3 Phenomenology

Phenomenology is a historical movement driven by the thinking of philosophers such as Edmund Husserl, Martin Heidegger, Merleau-Ponty & Jean-Paul Sartre - amongst others that have contributed to the development of this philosophical doctrine. An in-depth examination of the philosophical enquiry conducted by these philosophers is outside the scope of this thesis, primarily because the phenomenological research approach adopted in this research investigation is located outside the realm of phenomenological philosophical enquiry (see p50 paragraph 4). However, the methodological research investigation detailed herein is informed by some of the ideas put forward by phenomenological philosophers, that is why some of the key concepts and terminology of Phenomenology with various references to these philosophers are present in what follows.

Phenomenology includes the following terms:

1. **“Phenomenology”** – (a compound of the Greek words phainomenon and logos) is the study of human experience and the way in which things are perceived as they appear to human consciousness” (Langdridge, 2007)

2. **Lifeworld** – the concrete lived experience of a person – the ‘real world’ as experienced by ones consciousness.

3. **Epoché** – The ‘bracketing off’ or ‘putting aside’ of a human’s presumptions and preconceptions about the experience that is under
investigation in a phenomenological reduction. To temporarily suspend the “sedimented layers of consciousness built up through our temporal experiences” (Cogan, 2012) in order to unveil constancies and essences of the experience, revealing that which may be hidden from us by our unquestioned acceptance of our ‘lifeworld’ without any critical reflection. To allow us to “describe the things themselves” (Langdridge, 2007, p. 17) Epoché establishes the “phenomenological attitude” or perspective from which experience is to be taken. (Ihde, 2007, p. 29)

4. **Noema** (*noematic correlate*) – “that which is experienced” (Ihde, 2007, p. 29)

5. **Noesis** (*noetic correlate*) – the way in which it is experienced. (Ihde, 2007, p. 29)

Don Ihde, an American philosopher of technology and science, explains the relationship between noema, noesis and intentionality, identifying the key landscape and summing up phenomenology nicely in the following quote: “Within experience overall, there is that which is experienced, that called the object-correlate or noematic correlate. And in strict correlation with the noema, there is the act of experience or the experiencing that was the “subject correlate” or the noetic act. Here, as a correlative rule, it is maintained by intentionality simply that for every object of experience there is an act or “consciousness” that apprehends that object, and for every act there is an “intended” correlate, although some may not be fulfilled (empty). (Ihde, 2007, p. 29)

This correlation as the phenomenological model gives phenomenology its characteristic shape. Anything outside the correlation lies suspended under the previous terms of epoché. Thus any object-in-itself and equally any subject-in-itself remains “outside” phenomenology.

6. **Intentionality** – this does not mean intention of doing something (an act) – ‘as is’ in common use of the word, rather it speaks of consciousness and the fact that we are always conscious of something - (the object of consciousness)” (“the act of experience or the experiencing” (Ihde, 2007, p. 29)). The correlative relationship held
between ‘noema’ and ‘noesis’ as referenced above is labeled intentionality.

Husserl, later in his work, considered that the person who is experiencing could take a transcendental turn by reflexively stepping outside of the correlative relationship between that which they are experiencing and the way in which they are experiencing it, to take a “Gods eye view” as Merleau-Ponty put it, this aspect of Husserl’s philosophy has been rejected by almost all who have followed him (Sawicki, 2006).

Nevertheless, reflexivity plays an important role in the performance of the phenomenological reduction. The person conducting the exercise reflexively considers the influence that their own position and experience of the lifeworld may have brought to bear upon the observations and conclusions made during the processes of the phenomenological study they are conducting - and what impact it may have had on the study findings.

7. **The phenomenological psychological reduction** – builds on the start made by epoché above. The reduction starts with a detailed description of the things (objects of consciousness) that appear to us. The description avoids any interpretation of meaning at this stage. The rule of horizontalization is employed where the researcher gives equal value to each and every ‘thing’ that is described as an object of consciousness and avoids any categorization or value giving to the things or elements of the description.

8. **Imaginative variation** – this stage is sometimes employed as another way of shedding light on the meaning of an experience being reported by a research participant, where one imagines various alterations to features of the experience under investigation in order to uncover the layers and essence of the experience itself. This process helps by reducing the impact of peripheral elements surrounding the essence and meaning of the experience under enquiry. This is a simple but powerful method, focused on imagining other ways of seeing the experience under investigation, in order to
peel back the possible layers of experience to identify constancies that may be invariant and shared across a group of research participants for instance.

When reflecting on the nature of phenomenological studies, Langdridge comments that they “are usually, although not always, qualitative research projects designed to understand more about the experience of some phenomenon. Appropriate research questions are therefore open-ended questions seeking to understand more about a particular topic rather than attempt to explain or identify causes for phenomena.” (Langdridge, 2007) “Qualitative methods have increased in popularity in psychology in recent years, in part as a result of dissatisfaction with quantitative methods among sections of the psychological community.” (Langdridge, 2007) One such family of qualitative research approaches that have seen an increase in popularity are the methods and methodology surrounding phenomenological psychology.

Phenomenology’s epistemological position stems from the observation that “consciousness is always consciousness of something” (Cogan, 2012) and within that context “phenomenology is the study of human experience and the way in which things are perceived as they appear to human consciousness” (Langdridge, 2007, p. 10). Within phenomenological investigation the lifeworld (Lebenswelt) is the concrete lived experience of a person – the ‘real world’ as experienced by ones consciousness. Individuals may very easily glean completely different meanings from the same object of consciousness (whatever that object may be) because their subjective experience of the object of consciousness can be radically different, due to differing worldviews: philosophies, hopes, wants and beliefs. There are however “a number of existential givens of the lifeworld (e.g., selfhood, embodiment, temporality, spatiality)” (Langdridge, 2007, p. 55) and these givens of the lifeworld ensure significant concordance of existential experience across human subjects.
A phenomenological reduction involves ‘epoché’ – the bracketing off, or putting aside of one’s presumptions and preconceptions about the experience that is under investigation in the reduction. This “establishes the "phenomenological attitude or perspective from which experience is to be taken” (Ihde, 2007, p. 29). According to (Langdridge, 2007, p. 15 & 86) noema, that which is experienced and noesis, the way in which it is experienced become the focus of attention in the reduction. The reduction builds upon the start made by epoché, by providing a detailed description of the things (objects of consciousness) as they appear to consciousness itself.

Following epoché, one can then begin categorization and value giving in order to identify meaning from the reduction. This categorisation of the meaning of the experience described in the reduction can then be used in critical analysis, evaluation and interpretation of another’s lived experiences, such as those reported by a research participant for instance. A researchers interpretation of a participants lived experience is then subjected to rigorous verification by returning to the participant report to check the validity and sense of the researchers interpretation. Finally the detailed description of the experience under investigation is written up.

“If we believe in the real world then we need to employ a scientific epistemology (where we are “objective, detached and value free”) in order to discover truths about the world” “A phenomenological paradigm would, by contrast, have an epistemological focus on experience or narrative (rather than the real knowable world) and so require ways of capturing this that are subjective and involved” (Langdridge, 2007, p. 4).

To make clear the division between a phenomenological philosophical practice and the phenomenological research approach adopted in this research investigation; a phenomenological philosophical practice is a series of questions that interrogate the conceptual, and which lead to creative conceptual output along a line of argumentative enquiry. This is a practice of philosophy. By contrast, a phenomenological research approach is a structured approach to an enquiry located outside the philosophical realm,
which provides structure to that enquiry in response to the phenomenological philosophical enquiry, but which does not constitute a phenomenological philosophical enquiry in itself. This is a practice of methodological research, and does not need to slavishly follow the original philosophical enquiry, but rather is informed by it. The methods chosen for use during this investigation are firmly situated in, and informed by a methodological framework, which has Grounded Theory and Phenomenology at its centre.

3.4 A Comparative Analysis
To progress this investigation it was thought helpful to develop a deeper understanding of the context within which interactions between a human and the Articulated Head were taking place. The setting up of this context will first involve the selection of key concepts upon which to advance the investigation, by considering how interaction with a robot differs from interaction with another human being.

Human ↔ human compared to human ↔ robot communication
Communications that appeared to take place between the Articulated Head and a human interacting with it were a central feature of this human ↔ machine interaction. The word communication and a few other key terms will be considered in more detail as discussion in this section proceeds.

3.4.1 Communication Theory
A mathematical theory of communication (Shannon & Weaver, 1948, p. 380), presents a simple general communication system showing transmission of a message from sender to receiver. Many people subscribe to the view that communication is simply the transfer of messages between a sender and a receiver, where the receiver subsequently uses the information sent. That is, on this view, communication can be said to have taken place irrespective of whether the sender is aware of the receiver’s identity, and in most cases without a mechanism to check that the received message matches the intention of the sender.

Communication theory has various perspectives on the communication of information. The mechanistic perspective on communication would say that
information is simply passed from a sender to a receiver; where as other perspectives highlight the causes and effects of communications, not just the mechanism of information flow. From a psychologist’s perspective what might be interesting is the emotion a sender may be trying to impart to a receiver, this perspective might also find it interesting to know how a receiver interprets and acts in relation to this information. A sociological perspective would focus on how we communicate and how this initiates, promotes or evokes further communication, whereas the systemic perspective points to all communications being new, generated in the process of individual interpretation of previous information received. A critical perspective might propose that communication is simply the means by which a living entity conveys status and standing within the community and environment within which it exists.

Whichever perspective one subscribes to, information exchange, the examination and consideration of the causes and effects of the information transmitted or received are significant to this study and as such, this concept forms a core ingredient of the discourse and will be re-visited many times throughout this document.

3.4.2 The Universal Law of Communication

According to (Pillai, 2011, p. 58) S. F. Scudder (1900) posited The Universal Law of Communication that states; “all living entities communicate”. Everything living communicates as part of the rich interplay of concerns that can be glossed as “survival”.

Given the above stated Universal Law of Communication, it is important to note here that the virtual performer, the Articulated Head, was not a living entity and therefore did not communicate as such. However the Articulated Head did act as a carrier of information, a conduit for information exchange. The Articulated Head could be given agency with regard to the depositing and release of information between living entities through the use of programming, sensors, transducers, memory and storage.
Cultural conditioning, agency, embodiment and engagement become foci in the phenomenological reduction that follows. The findings that emerge from comparison in this reduction are core to the design, evaluation and validation of creative work proposed to help enhance audience engagement with the virtual performers performance. The rationale and details of the technical execution of the creative works are documented in section 6.

3.4.3 A quick review: What was the Articulated Head again?

Stelarc described it as an embodiment of the Prosthetic Head by the addition of the robotic arm. Indeed this was the case, the Articulated Head configuration did embody the Prosthetic Head within its design and that in turn led to an extended set of behaviours in terms of performance as is discussed below in relation to the Articulated Head’s abilities for seeing and hearing in the phenomenological reduction that follows - but embodiment stretches far beyond the obvious in this project as will be made clear in the discussion, conclusions and tentative findings from this section.

The Articulated Head was capable of spatial movement via the robotic arm, speech and a small variety of facial expressions. The Articulated Head was surrounded by a metal-framed glass enclosure of some form or another in all three of the exhibited spaces discussed in this thesis, NIME, SEAM and the Powerhouse Museum. The metal-framed glass enclosure in whatever form it has been implemented during the various exhibitions was designed to protect its audience from being hit by the robotic arm, a health and safety requirement.

A commercially developed software conversational agent called the A.L.I.C.E Chabot handled the input/out to the speech engine, and thus the avatar exchange of information with its audience. The Head was also furnished with some sensory capabilities both auditory and visual in nature, which allowed it to track various properties pertaining to its possible audience.
An attention model, programmed in MatLab by Dr Christian Kroos used the sensory information collected, applying probabilities, weightings and thresholds to the figures input from the sensory apparatus to make decisions relating to various behavioural aspects of the avatar, and to direct the industrial robotic arm’s movement within the interactive installation enclosure.
4. The Reduction

4.1 Scope of this Reduction
As is the accepted and the recommended tradition in the practice of a phenomenological reduction and epoché, this reduction should be seen as a simple, possibly naïve description of the things as they appear to my consciousness, absent of concern for previously established and accepted perspectives, academic or otherwise.

“We take for granted our bodies, the culture, gravity, our everyday language, logic and a myriad other facets of our existence. All of this together is present to every individual in every moment and makes up what Fink terms “human immanence”; everyone accepts it and this acceptance is what keeps us in captivity. The epoché is a procedure whereby we no longer accept it. Fink notes in Sixth Cartesian Meditation (Fink, 1995): “This self consciousness develops in that the onlooker that comes to himself in the epoché reduces ‘bracketed’ human immanence by explicit inquiry back behind the acceptednesses in self-apperception that hold regarding humanness, that is, regarding one’s belonging to the world; and thus he lays bare transcendental experiential life and the transcendental having of the world” (Cogan, 2012). If any links between the thoughts that are detailed in the following text regarding the senses and perception of things as they appear to my consciousness seem to correlate with the thoughts of others, these links are purely coincidental and born of a common interpretation of the lifeworld, they are not surmised by any direct reference to the writing or thoughts of those others.

To begin this reduction it should be noted that this is a critical evaluation of the forum and ways in which human communication takes place, focused on the phenomena that facilitate it. This evaluation is conducted, specifically in order to gain a deeper insight into the environment and means in which human ↔ robot communications take place, with a focus on how this differs from human ↔ human communication. The robot being referred to here is
specifically the Articulated Head, and although much of what is said probably does apply to other robots, there may well be cases where it does not. The purpose of this critical evaluation is an attempt to lay bare the realities of the situation under investigation, to get at ‘what is’ nothing more.

Human ↔ human communication has many facets that need consideration and human ↔ robot communication is likely to present another range of facets for consideration in comparative analysis. Human perception as a multisensory experience plays a critical role in this comparison. “After many years of using a modality-specific ‘sense-by-sense’ approach, researchers across different disciplines in neuroscience and psychology now recognize that perception is fundamentally a multisensory experience.” (Calvert, Spence, Stein, & MITCogNet, 2004)

Clearly, the robot end of human ↔ robot communication will lack the sophisticated multimodal senses that humans enjoy. As such, human ↔ robot communication will be different; from the human’s perspective it might involve the realisation that robot interlocutor is “not all there”.

It is this type of realization that I wish to capture. This datum is something that is primarily declarative, a thought that can be put into words. It is this end product of perception and thought that forms the reduction I am making (something opposite to the physical reductionism of chemistry or physics).

Notwithstanding the concept that both consciousness and perception have sedimented layers of experience, enculturation and acculturation present and stored in memory in any human brain, and that these sedimented layers will inevitably precondition both consciousness and perception themselves, the following reductive descriptions are made with a focus on the things as they appear to consciousness - rather than any preconceived or pre-researched perspective on the things presented and described henceforth. The descriptions given are not necessarily exactly the same as might be experienced by all other individuals but the majority of the detail is expected to find concomitance with most others, assuming unconstrained operation of
all their senses - hence the employ of the term ‘our’ in the descriptions that follow.

4.1.1 Consciousness

Consciousness is a state of mind in which a human or other biological being is said to be aware of, and responsive to their surroundings. Consciousness is also the state of mind within which communications between humans take place. Our senses are vehicles by which external communications are intercepted and transported to the brain for processing. Perception is the initial outcome of this processing prior to any further action initiated as a result of our perceptions. Put another way, perception is fundamentally interpretation of messages and communications, received by the brain via our senses - from both internal and external sources to the human body. Therefore, consciousness is essentially a multisensory experience and the sense fields that significantly contribute, if not constitute consciousness, within which communication takes place, are of particular interest here. So, it is probably helpful to try to describe the perceptual approach of things, as they appear to consciousness via the sense fields, including perceptual parameters, dimensions, boundaries, interdependencies and constituent characteristics as the first step in this reduction.

I note that many facets of communication are evanescent: variable in nature with a multitude of overlapping dimensions, which can impact and have ramifications which are diverse and wide ranging within the boundaries and context within which the communications are taking place. Secondly, I note that the variable facets of communication can have both instant, and delayed conditioning affects to subsequent elements communicated. For instance; when two humans are conversing, a facial expression could trigger a perception that one of the humans in the conversation was uninterested in the topic being discussed, this could trigger an instant change in the topic of conversation but might also trigger a shortening of the exchange over time. It may also affect future communication between these two individuals on this subject.
4.1.2 The senses

For the purposes of describing the sense fields and things as they appear to our consciousness, communication elements can be broadly defined as nonverbal and/or verbal in nature and can also be broadly described as emanating from outside the body and mind of the perceiver (external) or inside the boundaries of the body and mind of the perceiver (internal).

The field within which human ↔ human communication takes place is broad and may encompass elements supplied by any of our five traditionally accepted senses: sight, hearing, smell, taste and touch. Human communication also includes our sense of body position - kinesthesis and other aspects of the use of our senses: “In addition to sight, smell, taste, touch, and hearing, humans also have awareness of balance (equilibrrioception), pressure, temperature (thermoception), pain (nociception), and motion all of which may involve the coordinated use of multiple sensory organs.” (Zamora, 2013)

4.1.3 A spiritual dimension

Many humans would proclaim they have a spiritual dimension to their consciousness and senses, and that this spiritual dimension has direct impact on their perception of communications received, and would also have influence upon any subsequent actions taken. Therefore, even if one is completely non-spiritual and does not believe that this so called ‘sixth sense’ really exists, one does have to take account of this belief as if it exists within the minds, consciousness and perception of others, since such a belief influences the direction of their subsequent actions, that is, its influence exists and so is real and potentially significant, and must therefore also be a consideration within this reduction if, and where appropriate.

4.1.4 Communication or information exchange, initiations and sense fields

One could consider a communication between two humans to have begun in a non-verbal format even when these individuals are very considerable distance apart. For instance, a person may glimpse and make brief eye contact with another individual from the corner of his/her eye from several
hundred meters away, whilst walking along a busy city street. If the other knows this individual then one would consider that this distant eye contact had already been conditioned by previous communications between these individuals. On the other hand, if the other is not known, then one might say that this initial eye contact appears not to be preconditioned. However, beyond this point one has to consider that some form of conditioning has brought these individuals into communication. The spark that triggers these individuals to register the presence of one another, for whatever reason, even if it is just by chance, defines that communication has already begun. It might be a person’s looks and some aspect that promotes physical attraction, it maybe some other element of visual intrigue such as a shared idea of fashion or some other aspect of one individual that fosters perceived affiliation within the other, even if in reality the feeling of affiliation is not shared by both individuals. Conversely the non-verbal communication between these two individuals could be of a hostile nature, body language from one individual might foster aggressive tendencies within the other. It is also possible that the spark that triggered these individuals to register each other’s existence was not by the giving of any external sign, but could be the result of association with some other environmental object, something that has become the subject of one of the senses. There is an enormous range of possibilities present within this type of scenario, making it very difficult if not impossible for any third party to identify causes or even notice the onset of communication without the direct reporting of the individuals involved. However, what is clear is that the visual domain is very important in non-verbal communication.

4.1.5 The senses and range
At this juncture it becomes apparent that the sense field of sight might have a longer range than the other sense fields depending on the conditions of communication taking place. For the blind person to identify the existence of someone several hundred meters away on a city street seems somewhat unlikely. Taste and touch are not possibilities in this scenario and hearing would be neigh on impossible due to environmental noise effectively dissipating any chance of hearing a specific individual over this distance. The sense of smell is a possibility but even if the other individual were known
to the blind person, the wind was blowing in the right direction and this individual had a particularly pungent or unique aroma, one would expect that there would be so much other aromatic information mixed in the air between these individuals, such that any presence would effectively be masked from the blind man.

**Imaginative variations**

However, if we change the scenario through some imaginative variations, it becomes clear that one can recall events within our own sedimented memories of experience, whereby any individual sense can take precedence over another in terms of its identification-al range, and any sense can effectively have content masked from it by natural barriers. Given a human with no sense impairments, examples might include:

**Sight**

On a clear night one can see the light emanating from celestial bodies, some of which may not even exist anymore, it’s just that that light has taken so long to reach Earth that the demise of this distant solar system precedes the arrival of its previous glittering glory to Earth. In the broad, vision operates over straight lines and so long as there is no physical obstruction between a bright source of light and the eye, sight’s range is literally astronomical in comparison to the range of other senses (of course, although sight is exquisitely sensitive, detailed information is curtailed due to limitations in acuity).

**Hearing**

Whilst walking home from the Pub along high hedge rowed lanes and wooded surroundings on a warm summers evening in the English countryside, one hears the approach of a car winding around bends towards you long before any sight, smell, taste or touch of the vehicle can be made.

**Smell**

Whilst walking down the street, one can smell a Fish n’ Chip shop long before one can hear, see, taste or touch it!
Taste
One cannot necessarily tell that a fruit is bitter from the use of the senses of sight, hearing, smell or touch, only by tasting can we reveal this secret. Taste can reveal the hidden side of a multitude of herbs and spices present in food, which are not always revealed by smell, although sometimes smell can give clues which taste confirms.

Touch
Whilst engaged in a contact improvisation dance, the dancers involved can identify the direction and speed of movement of their co-dancer and can also identify the weight of leaning and needs of support between each other through the sense of touch. The amount and detail of information conveyed is only apparent to the sense of touch and cannot be accurately extracted by the sight, hearing, smell or taste senses of either dancer or an on looking audience.

4.1.6 Interdependencies and embedded enculturation
At this juncture it becomes clear that there are interdependencies between the senses and the hermeneutic interpretations formulated as a result of the communications received by them. There are also differences between the senses, they can have varying ranges and be more or less efficient and useful for identification purposes, dependent on the specific scenario and information that is the object of communication. Moreover, it becomes clear that our sedimented layers of experience allow associations and correlations between the senses that enable us to automatically identify, process and interpret sensory information that are so deeply embedded in our consciousness as to make description of objects that are the foci of any particular sense inseparable from the multisensory experience of consciousness and perception. Therefore we can only describe objects, which become the foci of consciousness and perception in terms of their apparent salience within any sensory conduit, given the relative influence imbued by layers of experience. We cannot completely dissect and separate any given phenomena from the experience (qualia) that also may have been the seed for the hermeneutic identification of the phenomena. That is, from our phenomenology, we may have come to know that an
event, object or stimulus exists but to all intent and purposes that phenomenon did not exist before we identified it, even though its, chemical, biological and ultimately physical antecedents did exist, since these ultimately gave rise to the events that underpinned identification.

4.1.7 Sense field pick up patterns
There is a distinction between the term sense, which is an instance of receiving information (input) and the subsequent processing of that information. The distinction between sense (input) and processing is true in both human ↔ human and human ↔ robot communication.

It is prudent then to note that the sense fields of human communication in our multisensory experience of the world are interactive and variable in range, and are conditioned by the scenario and object of information that is being communicated. Further, the senses can be sensitive to overlapping correlates amongst each other that help to solidify perceptions brought about by the processing of the sensory information received, the confirmation that taste brings to the smell of a particular herb, being just one of millions of possible examples.

As described by Don Ihde in his book, the Phenomenologies of Sound (Ihde, 2007, pp. 73-84) our sense fields display pick-up patterns of varying different shapes; that is, we can hear all around us, so our hearing displays a somewhat global pick up pattern, whereas, the pickup pattern of vision has a funnel shape (with maximal sensitivity in the fovea) and our brains stitch together a panoramic illusion of vision, which is what we perceive around us. Our external environment structures the pick-up pattern of our sense fields shaping them in conjunction with the barriers and noise present into a continuously morphing set of pattern shapes and resonances, which change in response to our movement and according to the physical changes within the external environment.

Although the above sensory vignettes and descriptions of capacities are fairly obvious, focusing on a description of the sense fields within which human communications takes place, may help to identify some of the key
differences and variations highlighted by comparisons in the reduction to phenomenology. Agency, experience and embodiment come to the fore, as it is in the description and comparison of these elements that the variations and differences between human ↔ human and human ↔ robot communication become most apparent.

4.1.8 The whole body sense

Touch (and the companion sense of proprioception), from a first glance appears to be the only overtly whole body specific sense. Sight, sound, smell and taste seem more head specific in terms of human design and structure, especially in relation to the location of the sense receptors positioning on the body. However, it is acknowledged here that a whole body chemical and electrical messaging system does exist and since the senses are connected to the central nervous system, all the senses do in fact have whole body influences. The minutiae of chemical and electrical messaging within the body is an area generally considered beyond the scope of this reduction but studies in these areas are worthy avenues of further scientific research in relation to the internal communications triggered by the senses and their associations.

The fact that four of the five senses considered so far, have sensors or receptors located in the head (a body part that can be oriented in diverse ways) may, or may not be important in terms of the phenomenological reduction of human ↔ human communication, it is not beyond the realms of possibility to consider that the evolutionary placement of the eyes, ears, nose and mouth of a human, being located so close to the brain, could, in part, be to allow for messages to reach the brain for processing as fast as possible. Another human sense apparatus located close to the brain is the vestibular system. The vestibular system located in the inner ear provides humans with their sense of balance, a sense that is essential for horizontal and vertical orientation (standing and walking) in the physical world. The Articulated Head was anchored to the ground and its physical movement hinged around those anchor points. Rotation and physical movement of the industrial robotic arm was defined by coordinates and driven by motors meaning that no sense of balance was necessary for its orientation.
Now, considering the brain itself, it definitely has the ability to learn and to commit information to memory. The brain has the capability to collect, analyse, evaluate and synthesize information. Many human beings believe they possess the power of free will, if so then the brain is the receptor of sensory information, a place of processing that information, and a place of genesis of will. A human’s sensory apparatus is the conduit, which carries messages to the receptor (the brain), and ‘will’ can be considered to be emanating from sources internal to the brain. Will may be the outcome of processing of sensory information and memory, in which case it is not entirely free but rather it is influenced - and so will becomes the decisions of a conscious mind. If will is simply the outcomes (decisions) of processing of sensory information and memory, then theoretically a robot designed with similar processing power, memory and sensory apparatus should be able to demonstrate similar abilities to those of the human brain such as collecting, analysing, evaluating and synthesizing information, it should also have the capacity to learn.

4.1.9 Genesis

With regard to the robot having the capabilities to be the place of genesis of free will; a human might find it easy to conclude that no such dimension of brain or mind could exist within the robot, so why would people hope, want and expect to be able to converse with the Articulated Head in a similar vein to that which they might experience when communicating with another human being? Just as one might conclude that the robot has no capacity for free will, one might also conclude that the robot has a lack of human capabilities and capacity related to any, and all of the senses too. The Thinking Head could not see, hear, smell, taste or touch in any meaningful way. Let us deal with each of these senses in order and consider what, if any aspect of these senses the robot could summon; to what extent the information received and the abilities of the robot to process this information compares to that possessed by humans, and to what extent the differences enable or inhibit the scope of communication or information exchange between the two. To do this we should consider in more detail what a communication is and in what form it may be transmitted or received.
4.1.10 The language of communication
Some time has already been spent discussing the senses, what has been left out is the language of communication. By language, I do not mean French or English; I refer more to the interactive interplay present in human ↔ human communications, the interception and hermeneutic interpretation of information captured by the senses providing conversational directives and foci for extension of the conversation-taking place. If one were to record a normal conversation between two individuals taking place on a bus or train heading to work, it is very likely that the dialogue would be laced with verbs, doing or being words and descriptive adjectives – you can imagine statements such as:

"I feel wonderful today"
"We had a brilliant night out at the theatre and the performance was magical, it has left me inspired to join an amateur dramatics group"

"I love driving over Sydney Harbor Bridge each day, the view reminds me of why I wanted to live here"
"It’s a shame we have to go to work, I could easily spend the whole day fishing on one of those boats"

These short statements can build affiliation with another individual we are communicating with by the associations that are invoked or they can elicit further comment, elaboration or the drawing of parallels extracted from our own experience and perceptions.

4.1.11 Non verbal communication
Humans have built a whole non-verbal “language” that communicates all sorts of information to those around them. The bus or train on the way to work is a very good place to see the semiotics of this communication in action. The signs of this communication can be as stereotyped and iconic as the waiving of a fist to signify annoyance or the rubbing of a thumb and finger to signify money - to the very subtle candour of a female in recognition and tentative acceptance of an attempt at wooing her. Even in the apparent disinterest of those surrounding you, you can see that, in the incessant stare
at the mobile communication tools and screens, the glow of laptops and iPads all over the place, people at first glance appear to be so wrapped in what they are doing as to be oblivious to the activity around them. However, when one takes a closer look, it becomes clearer that, at least in part, people are using these electronic communication tools to hide behind. You will notice them look up from time to time and if you watch people’s eyes closely it becomes apparent that many eye movements appear driven by interest in others and that many of us, in part, use electronic distractions to disguise this.

4.1.12 Sense by sense comparison

In what follows, I undertake a sense-by-sense comparison of human ↔ robot communications and relate this to the portrait of human ↔ human communications that I have already sketched.

Sight

The Articulated Head did have some simple capacity for visual processing. A stereo camera provided the ability for the Articulated Heads computer to track the co-ordinates and relative locations of several people in the immediate vicinity of the robot. It could also make use of this information. The THAMBS attention model helped to decide who or what the Articulated Head was going to prioritize in its processing. This information was then used to direct the industrial robotic arm to place an affordance on this location (i.e., if an action was to be undertaken it would occur in this space). Face recognition software was trialed with the Articulated Head but it proved to be unreliable and caused multiple and regular crashes, so it was not operational for the vast majority of the time in which this investigation took place. Nevertheless, face recognition software presents the possibility of extending behaviours beyond those that were possible with the Prosthetic Head. These behaviours included the gestural mimicry of a particular audience member and the possibility of remembering aspects of people who had previously interacted with the Articulated Head. Beyond this, the Articulated Head had no further capabilities for processing visual information. It could tell you that it was tracking something and where that something was but it could not tell you that it was a human it was tracking, and it could
not tell you anything related to shape, colour or any other visual attribute. Therefore one could sum up the Articulated Head’s visual capabilities and response mechanism as follows - “there was something in the monitored visual field and THAMBS instructed the robotic arm to follow this” So, the Articulated Head could not see as it had no capabilities for differentiation and identification of any visual attributes of anything within its visually monitored field.

**Hearing**

The Articulated Head was equipped with a stereo microphone, which was in theory capable of calculating the direction from which a sound source created in the sound field in front of it had emanated, assuming a low noise ambient environment. This in turn endowed the Head with the ability to track and also to follow, or “pay attention” (viz., allocate resources) to a sound source by moving the robotic arm and screen mounted upon it, to face the general direction from which the sound source was emanating, in practice, however, the accuracy and resolution of such a system in an unconstrained public exhibition space was questionable, indeed, where ambient noise was present this tracking system was rendered all but useless. This is the total extent of sensory processing that was available from the sound field of the Articulated Head. It was not able to process language or anything about the type of sounds that were produced within its microphones pick-up field, it only registered that a sound had emanated from a particular direction in the x plane. Accurate identification of sound generation in the y and z planes was not possible because only two microphones were used, mounted on the horizontal plane a short distance apart from each other. This system measured the time difference between a sound reaching one microphone and the other in order to establish the direction from which a sound was emanating. There was no provision for calculating the vertical position (angle) or distance of any sound source. Further, the Articulated Head did not have the ability to separate the different auditory information being collected (auditory streaming) and it certainly could not interpret the addition of multiple sounds emanating from various different places at the same time, so spatial auditory processing and interpretation of multiple data
streams simultaneously was a one-sided capacity that was restricted to the human in interaction with the Articulated Head.

However, with reference to the auditory domain, although not a specific sense, it must be noted here that the Articulated Head did have the ability to speak to its human audience, or perhaps more accurately, it had the ability to articulate predetermined and programmed narratives to a human audience through the mechanism of a text to speech engine. Although at a functional level the abilities of the Articulated Head appeared speak-like (they could be interpreted as speech), these head noises were at a deeper level not the same as speaking. Indeed, although some semantic analysis was borrowed in the head responses via the triggering of pre-packaged narratives, the Articulated Head was not capable of relating stored information in a symbolic fashion (it did not have a “semantic engine that preserved the truth-value of the proposition over which it might operate). However, even giving the appearance of speaking is a significant ability and it was this capacity that was probably the Articulated Head’s most significant ability during this study.

**Smell**

The Articulated Head had no capacity for the sense of smell; it had no olfactory sense organ and therefore no apparatus to capture aromas or the input patterns that give rise to the human perception of smell. It had no transducer or way to transport any of this type of information to any processing engine. Therefore we have to conclude that this sense did not exist for the Articulated Head and it could therefore not learn from this sense.

**Taste**

The Articulated Head had no capacity for the sense of taste; there was no receptor for this information and transducer to pass information for processing. Therefore the Articulated Head could not learn from this sense, and had no experience related to this phenomenon, other than that programmed into the conversational agent by a human. The Head could say “Oh yes I like chocolate!” but this was in fact just the re-iterated words of a human programmer – and as such the utterance “Oh yes I like chocolate!”
did not involve a thinking process on the part of the Articulated Head itself, this point though, does bring us closer to a more interesting discussion of embodiment and agency (which will unfold below).

**Touch**

Touch is an interesting sense because it is overtly and obviously distributed throughout the whole of the human structure – head and body. It is also interesting that it has the ability to be considered in the mental realm too: as one might say; “I was touched by the words of Nelson Mandela’s speech”, this statement indicates a stronger and more physical sensation of the thought processes being touched than the sense of hearing the words of Nelson Mandela’s speech would normally convey alone.

Touch is also interesting in that movement is the action and medium in which this sense takes place and one of its major constituents is to feel. That is, it has the ability to act as a transducer of two-way communication; it can just as easily transmit information as it can receive it! For example, one can touch the cooker and feel that it is hot therefore receiving information through touching and feeling. One can also touch and feel the hand of a loved one transmitting the information that the loved one is cared for, cherished. The vestibular system plays a role in the articulation of a human’s body movement in correlation with the ability to touch and feel.

With a deeper focus, it becomes clear that other senses too, or more specifically the receptive apparatus of each can also provide the facility for two-way communication. For example the eyes can both see - and can convey or reveal aspects of a story of the mind. The eyes can reveal happiness, sadness, anger and pain. In conjunction with facial expression, dress and composure, the eyes can reveal deeper details of the story of a long and troubled past for example. Many would say that the eyes are a window into the soul! Much can be gained from clues given by looking into another’s eyes. The two-way nature of these sense conduits and the handshake style information exchanges that take place over them is of considerable significance in relation to Articulated Head-to-human interaction because this is the way in which humans perceive agency.
To consider the ramifications of embodiment and how it affects the distribution of agency in humans, there is a type of dance called Contact Improvisation, which is a good example of a human ↔ human activity that demonstrates some of the effects that contact can have with regard to distribution of agency between dancers interacting. Touch, feeling and movement conspire to distribute and extend the sense of embodiment of each dancer to encompass parts of the other dancer’s body - and in this respect each dancer can experience the sense that the other dancer participating in the dance with them has some distributed agency within their body - therefore extending embodiment. Put simply, through touch you can feel the intention, leaning and the weight/balance of another individual in interaction and this in turn can influence your own movement. Although the Articulated Head could move, it could not meaningfully touch or feel - so couldn’t learn or convey any information through this sense. Furthermore because it could not feel, any attempt at exchanging information with the Articulated Head through touch was not received. This constraint was common to all the senses and the associated transduction of communication with the exception of sight and sound where the human could at least impart presence and approximate location but nothing else.

4.1.13 The revelation

The above analysis conflicts with how the Articulated Head was presented to the public, i.e., as a robot that could converse with you - and it was implied that the Articulated Head was capable of thinking and learning, indeed it was part of a larger research project labeled the “The Thinking Head Project”. So far, however, we find that the Articulated Head had no senses and was unable to gather any information about the humans in its vicinity other than their presence, approximate location in the x plane and information given in text input. Whilst information about someone interacting with the exhibit could be gathered from text input - and could be usefully stored in memory for recall and use in subsequent conversational discourse, this type of information gathering and processing does not really present convincing signs of thinking taking place in the machine to a human, largely because the human input the information into the machine in the first place. Human
beings understandably associate the phenomenon of thinking with consciousness and embodiment of their own brain - with its ability to gather (via sense apparatus), analyse, evaluate and synthesise information in order to stimulate original thought. Therefore, to present convincing signs of original thought taking place in another entity, a machine, to a human, requires that the thought appears to have had its genesis in that entity, the machine. Arbitrary presentation of preprogrammed thoughts from the Articulated Head to its interacting audience is unlikely to work convincingly because randomness is not really a feature of a thought process, rather the opposite, a thought process demonstrates rationale and/or logic. Since presenting a convincing demonstration of original thought taking place in a machine is inextricably linked to a humans perception that the thought originated in that other entity, the machine, presentation of the thought is likely to be most convincing via demonstration of an awareness of the multifaceted aspects and features of the immediate environment, including the human that the machine is suppose to be convincing.

The realisation that the Articulated Head had no capacity to sense aspects and features of the multifaceted environment was not a helpful place to be with reference to identifying new avenues for improvement of audience engagement with the machine, which was a primary directive of this research. If engagement is defined as passive observation of reaction to presence, then one can say that the Articulated Head did engage its audience. People did walk passed the exhibit, and in many cases did observe the Articulated Head’s reaction to their presence. Some did engage with the keyboard for text input, but most observed interactions with the exhibit were short and people very quickly became disengaged when the conversational agent made no sense or showed no awareness of the environment and features within it. Avenues for improvement of human engagement appeared very limited at this juncture. However, in terms of understanding the audience’s engagement in interaction with the Articulated Head, realisation of the limitations of the robot bring us to an important place with reference to an appreciation of the experience of avatar, conversational agent and audience interaction.
It is important to recognise here that there was one significant channel of information transfer, if not communication, that was available to the human to input information to the Articulated Head that has not yet been mentioned in this reduction, largely because its modality was not directly aligned with that of human ↔ human communications and the senses. A human could exchange information with the robot via typing text on a kiosk keyboard. This mechanism for exchanging information (texting) with the robot was not ideal but was a form of communication (= writing); speaking would have been far more conducive to the flow of information - as far at least one of the parties present in the interaction speech was the customary modality of exchange. However, in practice the keyboard was the way that the Articulated Head’s audience conversed with it over the duration of this study.

4.1.14 Conclusion to the reduction

In summary of the points made in the reduction above, the senses and their transduction apparatus feed consciousness with information. Perception of sensory information constitutes the nature and landscape within which communication and learning between a human and other living entities takes place. For two way communication of the senses to really take place such as has been discussed with the examples of touch and feel in relation to agency and embodiment, or eye contact in relation to it being a window into the soul, both parties present in the interaction have to share similar sense apparatus, receptors and transducers to conduct information to the brain, mind and consciousness in order to process and build an interpretation of the information received. The output of processing is perception, learning, the accumulation of our layers of experience, whether our perception is accurate or not, along with any actions subsequently instigated. Subsequent actions instigated can take a multitude of forms but will always include some other communication whether it is to modify our experience and memory as the result of something learned, or whether it is to take no action at all.
Finally if perception is fundamentally a multisensory experience, and the senses feed information to the conscious mind, within which processing of this information takes place – then it should be clear that the Articulated Head possessed little if any of the contributory phenomenological attributes of consciousness, perception or multisensory capabilities of a human because the Articulated Head did not have the ability to interpret multiple and multilayered data streams in parallel or simultaneously. Furthermore, since the Articulated Head couldn’t conceivably be given these uniquely human attributes of consciousness, perception or the multisensory capabilities to feed them, then one must conclude that (as was established previously as part of the universal law of communication theory (Pillai, 2011, p. 58) S. F. Scudder (1900) the avatar or robot was not 'in itself' a living entity and as such had no initiative, impulse or desire to initiate communication, survival was of no concern, the robot couldn’t see, hear, smell, taste, touch or more strictly – feel in any meaningful way. The robot did not have sense conduits, which allowed it to do anything in terms of collecting and interpreting information about its audience beyond the very basic auditory and visual capabilities previously imparted.

4.1.15 Tentative findings

So in the threads of the phenomenological reduction above, we find a vast void between human ↔ human as oppose to human ↔ robot communication - or perhaps more precisely we find that communication in terms of the above stated universal law, is only possible in one of the two conditions examined as both entities, sender and receiver, have to be living entities in any communication. Therefore, the Articulated Head or robot did not communicate anything! However, and most importantly, the Articulated Head was a conduit for the exchange of information, an agent between those who designed and programmed the Articulated Head and those who interacted with it. Just as the sense receptors capture external communications and act as a vehicle to transport the communications to the brain for processing, so to should the Articulated Head have been

---

8 The term ‘the conscious mind’ refers to the embodied human brain and the experience of existence as perceived by that brain through the sense apparatus that feed consciousness.
considered as an information interceptor and transport mechanism, and as such, it had the ability to display agency in communication transactions and also had the ability to encapsulate and display a multitude of ideas put into it by all involved, the designers, programmers and those interacting with it.

The Articulated Head was not just the embodiment of the original projected Prosthetic Head with the addition of the industrial robotic arm, moreover, it was the embodiment of a menagerie of ideas and embellishments that emanated from any and all those who contributed to its development. Therefore it stands to reason that any future propositions and developments designed to enhance engagement of an audience in interaction with the exhibits of this type should either:

1) seek to address the complete lack of multisensory capabilities identified in this reduction in one way or another - whether it be through addition of real sensory capabilities or through the enhancement of the avatars performance by conveyance of the illusion of sensory capabilities, perception and consciousness of the avatar to the interacting audience or

2) seek to bring a more integrated ergonomic system type approach to the audience in this interactive experience by providing more seamless linking of the assortment of ideas and embellishments that have emanated from any and all those who have contributed to its development, as opposed to the rather bolt on feel that a menagerie of ideas and embellishments can so easily bring to the design and feel of any interactive environment. This does not mean that the miscellany of ideas and embellishments were a mistake, it is just to say that their macro dynamic arrangement within the interactive environment needs careful consideration and alignment with the needs of the human in this interaction, if one is to communicate the feel of a cohesive engaging whole to the interactive experience, to the audience.

Having situated the project within a discussion of its limitations and constraints, in what follows I describe the methods by which data about human ↔ robot interactions was collected and the theories that underpinned these.
4.2 Practical application of the Video Cued Recall Method

Stage one of gathering Video Cued Recall Interview data was conducted at the NIME conference in 2010. The second phase of Video Cued Recall interview data collection was conducted at the SEAM Symposium held in Sydney in 2010. Stage three of the Video Cued Recall interview data collection was conducted at the Powerhouse Museum, Ultimo, Sydney, Australia.

4.2.1 Procedure

The person/audience interacting with the Articulated Head (a research participant) signed a consent form to agree to the Interaction and Video Cued Recall session being filmed. They also signed a release form at the end of the video recall to give permission for the audio/visual material and transcription related to their interaction to be published. This gave participants an opportunity to review their interaction before signing the release form. Finally participants were asked to answer a simple questionnaire, the results of which can be accessed via the Key Evidence Links Table Provided on CD ROM.

1. The consenting participant then interacted with the Articulated Head and the interaction was captured.

2. The audio and visual material was subsequently played back to the person who was interacting with the Articulated Head in a Video Cued Recall Interview, and they are asked to comment on what they were thinking and feeling during the interaction. This interview was also recorded.

A set up that allowed for the capture of a research participant’s interaction with the Articulated Head was required. A separate set-up to capture the subsequent Video Cued Recall Interview was also required. Since this investigation was embarked upon with little or no preconceptions about what were, and what were not important aspects of the interactions or interviews taking place at the time, it was decided to attempt to capture as much information as possible pertaining to both the interactions and
interviews, given the tools and technology available to the investigation. Therefore, the following set-ups were used.

**4.2.2 Interaction Capture Set-up**

The capture set-up in Figure 4-1 above utilises a Security Video Device that allowed for four separate video streams to be input to the device. The Security Video Device then allowed the user to switch output options between showing one of the four video input streams at full screen or showing all four video input streams on screen as one vision mixed data stream as shown in the diagram on the Video Cued Recall and Projects PC Monitor in Figure 4-1 above. The Articulated Heads text to speech audio stream could have been recorded directly, this would have resulted in excellent quality audio capture of what the Articulated Head was saying during interaction, but would not have captured any sound that the interacting audience may have made during interactions. After some initial experimentation with the possible audio streams, it was decided that the audio stream coming from the Firewire camera’s microphone captured a reasonable mixture of all the ambient sound sources providing a rich audio

![Diagram](image-url)
data stream that could provide information about audience and ambient sound conditions as well as the sound of the Articulated Heads voice. Whilst the Articulated Head’s voice was not so clearly defined in the audio stream captured, we also had text string data of the Articulated Head’s spoken words captured for comparative analysis. Interaction capture of the data fed into the computer via a video capture card was achieved using Adobe On Location (“On Location (software),” 2013)

4.2.3 Video Cued Recall Interview Set-up

To capture the interviewer and interviewee dialogue during the Video Cued Recall Interviews, a microphone present in the webcam in Figure 4-2 above was utilised. The software program iSpy (“iSpy (Software),” 2013), an open source camera security software, allowed for simultaneous display of the different data streams in one window. CamStudio software (CamStudio, 2013) was used for whole screen capture of the interview and playback data in one video file. The final audio stream for the CamStudio full screen capture, became a composite of the original interaction playback audio data stream, captured by the webcam microphone from the Video Cued Recall interview computer speakers, mixed with the interviewer and interviewee dialogue taking place in the interview. Although the mixed methods capture of audio data streams in both the Interaction Capture and
the Video Cued Recall Interviews was not ideal in terms of audio quality, the quality was sufficient to capture most of the richness of all contributory factors to the real auditory environments experienced, without expanding the technical complexity of the audio data capture set-up with the use of extra microphones and multichannel live mixing. Such an expansion of the capture set-up would have included the need for more equipment and possibly extra personnel as well. Furthermore the enhanced quality of the audio capture afforded, would not add greatly to the richness of the auditory detail apprehended in the data stream for analysis. Therefore the trade-off between actual audio quality as perceived by the listener and the quality of audio capture related to the richness of auditory detail apprehended in the audio data stream, was a fair and beneficial trade off for this investigation. A full list of the Video Cued Recall Interviews and other evidence collection conducted during this investigation is detailed in table 7-38, Video Cued Recall Interaction Captures.

4.3 Ethics
The completed National Ethics Application Form (NEAF) Approval Number H8543 for this research details the research participant consent procedures for participation in the Video Cued Recall Interviews and other aspects of this research investigation.

Research Participant Information Pack
The research participant information pack is included as E-Appendix 5.

Consent form
The research participant consent form is included as E-Appendix 6.

Release form
The research participant release form is included as E-Appendix 7.

Participant questionnaire
The research participant questionnaire is included as E-Appendix 8.
**Ethics amendment**

Ethics amendment approval was sought so as to include all text string data (made anonymous) if published, in data analysis. The ethics committee amendment approval is included as [E-Appendix 9](#).
5. Setting the Scene:
The Performance Arts Context and Nature of the Investigative Terrain

5.1 From the engineering - to performance arts perspective

The development of the Articulated Head project could so easily have progressed predominantly from a robotics and engineering perspective. Many aspects of the project did develop this way but the Thinking Head project owed its inception to Stelarc’s Prosthetic Head. Stelarc, who is first and foremost a performance artist, exerted his influence over the evolutionary stages of the Thinking Head Project from the original Prosthetic Head through to the Articulated Head. This influence both initiated and maintained a strong performance art dimension to the works, which might otherwise have been ignored. Indeed, the title and content of this thesis acknowledges the importance of the artistic dimension to these works, both in citing the fusion of science and technology within the arts in the title, and then presenting and testing creative new media propositions with the intention of addressing the big question: how can interaction between humans and machines be improved? (particularly from a performance arts perspective).

5.1.1 The positive engineering perspective

The big question can so easily be interpreted in terms of a massive engineering challenge. That is, how can we make the robot more sophisticated? Sophisticated enough that humans move beyond the perception of merely talking to a machine, to a perception of having an empathetic relationship with the conversational agent that inspires an awareness of the head as an agent, one that is both engaging and has agency. One can just imagine an engineer’s planning meeting wish list: give it more inputs, more sensors, more sensitive apparatus, finer movement mechanisms, more processing power, more memory, better, smaller, lighter and faster electronic and mechanical components - but this approach brings with it a multitude of complications. Teams of engineers and programmers would be required to tackle all of the developments present on the wish list. Each development installed would likely manifest in another
set of behaviors presented by the machine to an interacting audience. Some of the emergent behaviors may be accidental, unplanned and happily beneficial to the human ↔ machine interaction in some way or another, whereas other emergent behaviors may be less helpful and require further engineering to address and suppress these unhelpful manifestations. Most importantly, the above wish list (hypothetical or otherwise) implies that the machine is the thing that needs changing; that development of the machine is the key object of focus in order to answer the big question. That is, the suggestion is that through an exponentially expanding wish list and a sustained engineering approach, coupled with a trial and error development of the machines capabilities, improvements to human ↔ machine interaction will somehow eventuate.

5.1.2 Questioning the approach.

However, an engineering approach is unlikely to be the correct one that would satisfy all of the issues raised in this thesis in their entirety. For whilst it is fair to assume that modifying the machine might bring about improvements in the flow of human ↔ machine interaction, one must consider all aspects of the performance that is taking place in interaction, the human and the spatio-temporal environment in which the interaction is taking place as well. Therefore, the engineering view outlined above (positive though it maybe) is too narrow to capture the domain of human interaction; it overstates the mechanistic by seeing all problems as stemming from input/out limitations. In short it over-generalizes what can be solved by brute force.

"Too large a generalisation leads to mere barrenness. It is the large generalisation, limited by a happy particularity, which is the fruitful conception." (Whitehead, 1997, p. 176)

5.1.3 The negative perspective

The opposite perspective to the positive (but possibly over-general view above) might be a view that the Articulated Head will always be just a computer-controlled robot/avatar. That is, given that any convincing avatar performance is so drastically constrained and limited by the current state of technology in areas such as automatic speech recognition, and in how our
abilities to glean, analyse, store and represent useful data regarding aspects of the audience (e.g., a person’s sex, mood, dress, health or even altitudinal disposition) might be mimicked, and given that sensing technologies and databases are likely to remain under developed, then it could be concluded that these restrictions effectively presented insurmountable obstacles to “conditioning affect and engagement” so that the venture is unworthy of effort as it is hopelessly futile. This somewhat negative, some might say cynical (though some might also say realistic) view would definitely lead to “mere barrenness” for this investigation because not to attempt to establish how one might achieve enhancement of performance in this human ↔ machine interaction, simply fails to address the big question.

5.1.4 Of positive thought, critical evaluation and leaps of faith
Both the positive and negative views towards improvement of human ↔ machine interaction projected above held some merit in terms of fact and prognosis, though both have also ignored very important aspects of the context within which the big question is asked. For example, the positive view projected towards machine development completely ignores possible contextual development in the environment surrounding the robot and also fails to acknowledge the importance of the human in the interaction environment. The positive view simply represents a leap of faith that everything will be solved by technological developments; yet these developments may well lead to a snakes and ladders period of protracted work that ultimately fails to make progress because some of the new and intended improvements made to the machine may actually cause more problems than they solve in terms of performance (a case of one step forwards two steps back!). On the other hand, the negative view, whilst possibly projecting an accurate assessment of current technological achievements and the barriers to development, completely ignores the future technological horizons that may occur. Indeed, if one never makes the attempt then one will never know the precise nature of the problems that need to be overcome.

A critical surveyor of possible project developments provides a more tempered view of both the current status of the project and the scope for
future developments. This tempered view is strategically useful for formulating an overview for project management. Whilst a positive disposition might be attractive and make one popular amongst the project development team in the short term, and faith might occasionally lead to fruitful conceptions, it is much more likely that carefully considered critical evaluation will lead to successful outcomes and by so doing make more concrete steps forward in determining how human ↔ machine interaction can be improved in this and similar interactive environments.

5.1.5 A melioristic yet tempered view towards development

Broadly speaking, a melioristic view towards development of the machine and interactive environment in which the human plays a very major part offers a viable pathway to adornment, embellishment and agency along with the possibility that the machine’s performance may have a conditioning affect that will help interacting and engagement with its audience. This positive view emphasizes many small but developmental steps in the direction of audience engagement and so should ultimately enhance performance within the domain of this human ↔ machine interaction.

However, before charging ahead with even small developmental steps, it is important to first formulate a clear overview of the interactive environment under investigation and then subsequently determine which project developments might prove to be fruitful in terms of improving human ↔ machine interaction. This thesis is more concerned with trying to establish what developmental steps make sense and why, than it is in making the developmental steps and testing them, (i.e. the concern is to engender a process of emergent discovery rather than hypothesis testing).

Furthermore, this investigation is less concerned with giving credence to current technological barriers and constraints, choosing instead to place more focus on the development of arguments for determining which technological barriers must be overcome in order to achieve improvement of this human ↔ machine interaction. Thus, this investigation projects a positive view towards future developments in human ↔ machine interaction, based on a tempered critical analysis of human centered empirical Video
Cued Recall Interview evidence collected. The evidence then leads to a carefully considered and critically evaluated recommendation for future developments (including technological leaps of faith) that are expected to prove successful in terms of improving this human ↔ machine interaction. The emergent theories and ideas that stream from this investigative work are based on sensible constructed and substantiated arguments that are linked with happy peculiarities of the interactive environment under investigation and are expected to lead to fruitful conceptions when implementing the recommendations with similar interactive exhibits.

5.1.5.1 The happy particularity with regards to this investigation
The principal “happy particularity” (Whitehead, 1997, p. 176) with regards to this investigation, which was identified through analysis of empirical research data collected, presented and discussed in more detail in Section 7 under Observation 1 — The anthropomorphic stance, was that the audience interacting with the Articulated Head appeared (without exception) to be predisposed to adopting an anthropomorphic stance toward interactions, when first approaching the Articulated Head. The audience appeared predisposed to the perception of the avatar as a conscious entity, indeed they acted and engaged with the Articulated Head as if it was an existential being. I will elaborate on the anthropomorphic stance later, before this I will describe elements in the environment that might help condition this.

5.1.6 The influence of the spatio-temporal environment
To help illustrate the importance of a spatio-temporal environment and its influence in performance more clearly, I am asking the reader to consider the following scenario as something potentially parallel to the Articulated Head’s interactive performance:

5.2 The analogous performing arts scenario: The Stage
My colleague, a lecturer of Performing Arts and a thespian might play Long John Silver in Treasure Island on the theatre stage, where his convincing performance, the costumes and the stage set, touch the audience in such a way as to make them forget, for a short time, that what they are actually watching is really just an act – the performance draws the audience into the
story in such a way that *it is real – he is a pirate* for the time in which the performance is taking place, I argue that so too must the robotic exhibit’s performance, costume and staging touch its audience and seductively draw them in. Just as the audience might leave the theatre following the performance of Treasure Island, saying, the actor was a convincing pirate, so too must the performance of the Articulated Head elicit such a complimentary commentary. If I find such a reflection from the audience in the Video Triggered Recall Interviews after interaction with the robot has taken place, I would deem it a success.

For a human to assimilate the characteristics and narrative of another human’s story, and to project this convincingly on stage, takes considerable discipline and practice, along with attention to detail to the art of acting, costume and set design. To convince a human that the avatar, the Articulated Head, was a conscious entity, one which acted with autonomy and intent would be a similarly difficult illusion to create - and one especially difficult to sustain.

The goal of this study and accompanying creative work (see section 6), was not necessarily to make the avatar more humanlike, nor was it to make the Articulated Head overtly theatrical but to find ways in which it could produce the practice of agency so that, at the very least, it appeared to act with autonomy and intent. This is the crux upon which this study evolved, a pivot that led to both the direction and the research methodology employed as well as shaped the development of the creative additions that were proposed and tested.

An audience perception of the avatar as a conscious entity, which acted with autonomy and intent was conceptually wrapped in the necessity to persuade the audience that the avatar possessed consciousness.

Drawing parallels between the theatre stage and the performance of the Articulated Head was both intentional, and *not* insignificant to this project. However good an actor my colleague might be, the removal of his costume,
his wooden leg and the sound of its thumping on the stage floor, all the elements of the stage set, the boat, the island, sand, palm trees, the sounds of water and wind and the treasure chest itself, would likely weaken the pirate illusion and inhibit the audience’s propensity for suspension of disbelief in the character being presented to them by the actor.

That is, environmental auditory and visual contextualizing cues, when carefully constructed in relation to the performance of Long John Silver as detailed above, not only condition affect and engagement but also act as catalysis to the perception of a ‘real’ narrative consciousness within the actors portrayal of Long John Silver - the part and performance, as perceived within the minds and consciousness of the audience watching it.

The environmental cues, when carefully constructed in relation to the avatar, the Articulated Head, should similarly have been able to not only condition affect and engagement, but should also have been able to act as catalysis to the perception of a ‘real consciousness’ of the avatar within the minds and consciousness of its audience, in the proceeds of their interactions.

Careful attention is drawn to the critical conceptual point made above, as it represents a clear rationale and motive in relation to the proposed creative new media additions.

5.2.1 Illusion and Conscious Perception

Illusion and conscious perception are brought to the fore. After all, with reference to our ‘theatre stage scenario’ detailed above, it is very clear upon reflection that all the elements of the stage set, the boat, the island, sand, palm trees, the sounds of water and wind - are all essentially falsely situated within the wider context of our existence and the layers of our experience.

It is likely that most of the stage elements are not even ‘real’; the majority of the scene may be represented as a painted backdrop. This however does not stop these elements from exerting a contextualizing influence upon the act, or from being brought to bear upon the minds and consciousness of the
audience watching the show. Neither do these contextualizing elements detract from the centre of attention on the stage, being the actor in this case, Long John Silver. In fact, very much the opposite, they are the catalyst, which cause the catalysis of the perception of a real ‘other’ within the consciousness of the audience to be realised. Therefore, the contextualizing influence of auditory and visual cues surrounding the avatar and its audience should also have proved to be just as powerful a tool, without detracting from the star of the show, being the Articulated Head at the centre of the interactive installation

5.2.2 The importance of what is said and what is not
As previously stated, the importance of the human in the human ↔ machine interaction is of central concern throughout this investigation. It is the human that is the rationale for why this study prioritises human experience and asks the end user (humans) what they actually think and feel during interactions, what matters to them and indeed what does not. Furthermore, there is as much to be gleaned from what participants interacting with the Articulated Head did not say in Video Cued Recall Interviews as there is to be gleaned from what they did say. That is, if ‘something’, an observable characteristic present during interaction (a possible object of consciousness), was not mentioned as being a point of focus or raised in discussion of interactions, it may still play a role in the interpretive data analysis. Just because the participant did not mention it, that does not mean it was irrelevant. Indeed some unmentioned aspect of the interactive environment in which the interactions took place may be highly significant.

5.3 Investigative framework: Evidence, not hypothesis led
Any self-respecting and professional chief crime scene investigator would not enter into an inquiry with any assumptions about the perpetrator of the crime (even if there were some highly likely suspects) without first considering all the evidence and clues at their disposal very carefully, this investigation has similarly not been entered into with any specific hypothesis or assumptions about how human ↔ machine interaction is perceived or can be improved. The investigation, its findings and its conclusions are research data [evidence and clue] led, not researcher or hypothesis led. There was
no attempt to prove a predefined theory or belief. This investigation has been conducted from the outset with this, and only this particular case in mind, the human ↔ machine interaction that took place between the Articulated Head and an interacting audience in relatively unconstrained public exhibition spaces.

A section of text is included electronically as E-Appendix 10: The Analogous Road Traffic Accident Investigative Scenario. The E-Appendix contains some thoughts related to the nature of the investigative terrain encountered in this investigation. The treatment of circumstantial evidence and the impact that this might have on the investigations findings in relation to understanding and establishing how to improve the interactive environment is also considered.

5.4 Philosophical and practical justification of approach

Art influences science and technology just as science and technology influence art. The Articulated Head was an interactive installation artwork that incorporated a menagerie of technologies, many of which displayed functionality that was at the cutting edge of scientific technological development at the time. Popular perceptions and expectations of state-of-the-art science and technology can be exaggerated beyond reality, driven by ideas presented in science-fiction popular culture and other media.

5.4.1 The current state of play: Popular perceptions

In recent years, there has been an increasing focus in the media on the development of robots and robotic-type devices. For example, an article in the August 2011 copy of the National Geographic Magazine (NGM, 2013, p. 66) dedicated to the latest developments in humanoid robots, covered the work of some well-known roboticist such as Professor Hiroshi Ishiguro with his work on human-like robots at Osaka University, and the work of the LifeNaut project in Vermont with the robotic head called Bina48.
Figure 5-1 Professor Hiroshi Ishiguro (right) sits next to the Geminoid HI-2 robot

Professor Hiroshi Ishiguro (right) sits next to the Geminoid HI-2 robot, a tele-operated robot that looks exactly like himself in Kyoto, Japan. Ishiguro is director of the Intelligent Robotics Laboratory at Osaka University.

Picture: Reporters / Barcroft Media

Figure 5-2 Bina48 an early demonstration of “mindfile” transfer at Lifeanaut.com.

The LifeNaut project and its social robot Bina48 are featured in a Vermont local newspaper. This is an excellent article about “mindfiles” and the purpose of the Lifenaut Project:
(http://www.addisonindependent.com/201202robot-blurs-biological-boundaries)
The National Geographic article mentioned projects a widely held popular public view about the current state of robotic development. The Magazine article opens with the statement:

“Robots are being created that can think, act and relate to humans. Are we ready?” (NGM, 2013, p. 66)

These are indeed bold proclamations, which seem to be ubiquitous amongst sci-fi fans and robotics researchers alike. One might somewhat cynically suggest that the increase in the frequency of such statements in these two groups is no accident, with the latter group being a subset of the former. On this view, media narrative and hyperbole, have led to an outlook on the likely progress of science and technology in relation to robotics that is ridiculously over optimistic (especially when viewed in relation to actual achievements). The latter of the two groups, appears to exaggerate the current prospective forecast of humanoid robot capabilities in relation to concepts such as thinking and building relationships with humans, simply because it makes a lot of sense to paint a very rosy picture, especially when trying to secure research grant funds from those who are probably less well informed about robots and are inclined to believe the publicity.

5.4.2 The role of the Arts in Robotics and vice versa

The above is not to say that the active imaginations of sci-fi fans, TV and film scriptwriters and alike have had only negative impact upon robotics development. Indeed, much the opposite may be the case; the Arts can and do lead scientists and technologists towards new visions and aspirations that are sometimes realized. Certainly the direction of scientific and technological developments is influenced and guided by artist’s aspirations and portrayals. There are numerous examples of artist aspirations and portrayals influencing scientific and technological developments from history, in Leonardo Da Vinci’s depiction of a vertical flying machine in the 1480s, which many now attribute as the first documented design of the modern day helicopter; -
- to more recent artistic influences on modern scientists and technologists such as the inspirational influence that Star Trek episodes (CBS-Entertainment, 2012) are said to have had upon NASA (Wilson, (n.d.)) scientists investigating rocket propulsion and space travel, and the statements of people like Martin Cooper (“Martin Cooper - History of Cell Phone,” 2011) who is attributed as the inventor of the cell phone. Cooper says that his inspiration for invention of the cell phone came directly from Star Trek Communicators.

5.4.3 The Tryst: The Science/Technology/Arts Nexus

It is here, that the Fusion of Science, Technology within the Arts proliferates. It is also here, in this interdisciplinary tryst, that strong philosophical debates
surrounding constructionism, determinism, reductionism, interpretivism, epistemology, methodology, phenomenology, ontology, anthropology – (the list of ‘isms’ and ‘ology’s’ could be extended almost ad infinitum) transpires.

5.4.3.1 The contrasting methods debate

Escalation of philosophical debate in this Arts/Science/Technology nexus comes as no surprise because the seemingly opposing philosophical worldviews of practitioners and researchers in these domains become convergent – thus leaving scholars no choice but to contemplate the relevance and currency of converging philosophical and methodological viewpoints that are normally unceremoniously rejected by their own doctrine.

The point of the above observation is to emphasize the potential interdisciplinary nature of the Science/Technology/Arts weltanschauung, and to acknowledge the tensions between accepted methods of investigation in these differing domains. The purpose of this acknowledgement is to indicate how consideration of the differing worldviews and the differences between the scope of questions relate to human ↔ machine interaction. Included in this is a consideration of the methods used to get at these questions. These points are relevant to the research detailed herein because of the scope of the big question: How can interaction between humans and machines be improved?

5.4.3.2 Human ↔ machine interaction: Micro versus Macro

It is my view that the study of human ↔ machine interaction (HMI) interaction, or human ↔ computer interaction (HCI) as it is often referred to, fundamentally takes place at an intersection occupied by areas of Science, Technology and the Arts. Computer Sciences, Usability, Interactive Design, Electronic Design, Sound and Vision, Psychology and Behavioural Science, and a range of other fields of study all have a role to play in this intersection. Psychology clearly falls on the human side of the human ↔ machine relationship whereas Computer Science clearly falls on the machine side of this relationship. However, there are several fields that traverse both sides of
this relationship, Interactive Design, Sound and Vision, are three such examples that feature strongly in this thesis.

With regard to the methods adopted by some of the study area groups mentioned above in relation to the human ↔ machine interaction under investigation here:

- Computer Science might ask questions such as: What impact does expansion of the conversational agents AIML coded vocabulary have? Does expansion of the conversational agents vocabulary correspond with a direct improvement in human ↔ machine interaction? They might then select a specific number of words or text string responses to increase the Chatbot vocabulary by - and then set out to test the effectiveness of the expanded vocabularies introduction on the human ↔ machine interaction in collaboration with an experimental psychologist by using an experimental testing paradigm of one form or another.

- The Experimental Psychologist might ask a question such as; has the new code improved human ↔ machine interaction by extending the scope of dialogue exchange taking place between the Articulated Head and its audience? To find evidence of this they might test participant’s interaction, first without the new code - and then with, to see if factors such as conversational longevity in the interaction had been increased - or they may use targeted participant questionnaires asking participants to rate their engagement against a scale after interaction, to see if an increased level of engagement was reported by research participants.

Whilst each of the exercises mentioned above would probably contribute something useful towards knowledge about one particular micro dynamic aspect of the human ↔ machine interaction and how it might incrementally be improved, that contribution would be very small in relation to the enormity of the big question. Analysis of the data collected may very well
appear significant when represented in isolation but might actually be relatively insignificant or possibly (in some cases) irrelevant once adjustments for addressing some concern related to some macro dynamic aspects of the interactive environment has been considered.

This research project was conducted broadly speaking from an interactive designer perspective that closely resembles the view below.

- The interactive designer by contrast to the approaches adopted by the computer scientist and experimental psychologist mentioned above, might ask questions such as: what are the overarching features (macro dynamics) governing the way in which interaction between the Articulated Head and its human audience take place? How do these features contribute positively or otherwise to the flow of interactions? And how can the interactive design be rearranged to improve the flow of human ↔ machine interaction?

5.4.3.3 Emergent Phenomena: Context and Causality: Micro versus Macro.

For the purposes of clarifying a particular point in this thesis - it seems best to declare my personal view that there is no clear divide between the domains of the Arts and Sciences other than (perhaps) a predisposition to be bias in favour of cause over effect or vice versa. I feel it necessary to declare this stance because I approached this research project with a strong performance, media arts and technology background - but with much less experience in the social and behavioural sciences such as psychology. The institute from within which this project was managed has a very strong experimental psychology academic tradition and I have encountered many experimental psychologists as a result.

Experimental psychologists in particular, subscribe predominantly to a reductionist methodology and as such often question other philosophical and methodological approaches such as those used in this investigation. In light of the practical considerations governing the methodological approach to this research to those who would question it, although some
phenomena can be explained by reducing them to a more fundamental phenomena and breaking it into constituent parts (especially when these phenomena are perceived as emanating from a non-biological complex system), subscription to the philosophical viewpoint that any complex system is simply the sum of its parts, especially with reference to biological beings seems somewhat harder to justify. Emergent phenomena do arise that can transcend the components revealed by any reductionist investigation - such that even if these phenomena do not escape an explanation due to a scientist’s determination to prove their cause, often that explanation seems superficial in relation to the wider context within which the phenomena transpired. That is not to say that causality is not important, it is – but this investigation is not about cause - but is about effect; it is focused on how human ↔ machine interaction can be improved.

The minutiae of interactions are as transient, dynamic, versatile and spontaneous as the human mind that participates in them, and whilst the experimental psychologist might be predisposed to try and explain how the human mind works by employing reductionist methods, this investigation is not trying to do that. Trying to explain the cause of phenomena by reducing them down into constituent parts (what the scientist sees as the more fundamental phenomena that explain the epiphenomena) and then using an experimental psychology testing paradigm to find out facts in relation to the minutiae of interactions taking place in this investigation, would require a vast team of researchers and a protracted period of time, possibly several years, in order to collect and analyse sufficient data to draw sustainable conclusions. Therefore a more practical and strategic approach was adopted, one that had some chance of finding useful leads to possible answers to the questions posed in this investigation.

Focusing on the macro dynamics of interactions taking place within this context of human ↔ machine interaction in the complex system under investigation, makes much more sense because the system and the interactive environment in all its complexity, seems to be irreducible (especially in the timeframe afforded this study) - and is inherently irreducible
due to the human component, the relatively unconstrained and unencumbered environment in which the interactions were taking place, and variations between the human subjects involved in these interactions.

The human ↔ machine interaction under investigation here, and the cause but (perhaps) not the effect of all the emergent phenomena in these interactions are enormously complex and possibly beyond the current scope of mankind to explain accurately in all their micro dynamically psychological, interpretative, physical, biological, chemical, and electrically detailed processes by reductionist methods. Furthermore, even if the reductionist-testing paradigm were able to describe every aspect of the minutiae of emergent phenomena in these interactions over a protracted period of time, its employ would have been inappropriate at this stage of the investigation into how one can improve human ↔ machine interaction.

Why? Well, if epiphenomena are explainable by more fundamental constituent phenomena, as subscribers to the reductionist philosophy would claim, then it is reasonable to infer that epiphenomena wield no unpremeditated agency upon the phenomena that explain them. Similarly, it is reasonable to suggest that investment in understanding the macro dynamic arrangements of this human ↔ machine interaction should take precedence over micro dynamic explanations of causality of phenomena. It is strategically wise to focus upon the macro dynamic design of interactive environments first - especially as any move to make macro dynamic rearrangements to the interactive environment in order to improve the flow of interactions, might easily render investigation of some aspect of the micro dynamic interaction superfluous to the investigation in the changes wake.
Thus a micro over macro dynamic focus to this investigation would represent a simple, yet fairly clear case of ‘can’t see the wood for the trees’.

Section 6 that follows, imparts the rationale, artistic and technical execution details for the creative new media propositions. Section 7 presents the research data and findings in relation to a theme based analysis. Section 8 then examines key concepts that help to explain findings from the theme based analysis and inform design refinements in relation to the key research questioning strands. Conclusions to this investigation are presented in Section 9 with a blueprint of emergent recommendations as to how human ↔ machine interaction can be improved. Section 10 concludes this thesis with further discussion and a view of the future surrounding the Science, Technology and Arts Nexus in relation to Robotics and human centered Interactive Design.
6. The Creative New Media Propositions

It was proposed by the Thinking Head performance team that the development of some creative new media work to augment the Articulated Heads auditory visual environment should take place and be evaluated in parallel with the collection of Video Cued Recall Interview data during this investigation. A lot of technical and programming work went into the creative new media additions to the Articulated Head exhibit; some of this work is cited in the text that follows. I would not want the breadth of this work to go unnoticed. However, the technical execution of much of this creative new media work is in many ways (perhaps) peripheral to the core research detailed herein, therefore a separate document, which outlines more of the technical execution of these creative new media additions, is included electronically as: E-Appendix 11: Creative New Media Additions Documentation. It should be noted that implementation of these creative new media projects, and the findings from evaluation of their effectiveness, has provided a convincing endorsement of conclusions (see section 7 & 9) emerging from the core research activities. Indeed the implementation of these new media projects, and more specifically their evaluation, helped to identify the key barriers to interactive engagement in relation to the human ↔ Articulated Head interaction that was under investigation.

6.1 The Additions

The creative new media propositions were conceived in general alignment with one of the objectives of the original grant proposal: E-Appendix 2: From Talking Heads to Thinking Heads being; to allow researchers to examine key questions in individual research areas, which could not have been addressed without the implementation and development of the Thinking Head.

The key question/s related to an individual research area that were of particular interest were:

1. Investigation of modes of auditory/visual media delivery in a three dimensional spherical interactive environment including an;
• Investigation of tools and techniques for the presentation of contextualising visual media in interactive environments.
• Investigation of multichannel spatial audio tools and techniques for the delivery of contextualising three-dimensional sound representation to audiences in interactive environments.

With the above study aims in mind two creative new media augmentations to the Articulated Head’s interactive environment were proposed:

1. Projection of text – the project was given the title ‘The Thought Clouds’
2. A Spatial Auditory System

Section 3 presented a reduction comparative analysis of human ↔ human, compared to human to robot communication, which gave a deeper insight into the existing nature of interactions that were taking place between the Articulated Head and its audience. Conditioning, agency, embodiment and engagement in relation to the senses became foci in the reduction. It was established that there was an absence of apparatus afforded the Articulated Head in terms of its ability to capture and analyse information in the sense-domains of smell, taste and touch. There was some possible, but limited scope for provision of apparatus to affect audience engagement through stimulation of their sense-domains of smell, taste and touch in future human ↔ machine interaction, more of which is discussed in Section 9 – Senses Related.

However, the reductions quickly established that the success, or otherwise of the avatars virtual performance, was ostensively entwined within the auditory and visual domains. As such, presentation of contextualising media in the auditory visual domain made sense as a primary target for enhancement of audience engagement. Very much the same as was identified in the analogous performing arts scenario: The Stage presented earlier in Section 5.3, the conditioning affect of the stage set, props and sounds surrounding the virtual performer (the Articulated Head), should have been able to act as a catalyst for catalysis of the audiences experience of ‘other’ (an existential being) present within the interactive environment. The
stage set and props (artificial or otherwise) should have been able to make the virtual performer’s performance more believable.

The initial expectation was that implementation of the propositions would manifest in a deeper more immersive and engaging interactive experience being reported by the audience interacting with the Articulated Head. It was identified early on in the brainstorming of these project ideas that the real challenge with regard to achieving the desired outcomes (illusions) from media delivery in the auditory/visual environment surrounding this avatar ↔ audience interaction, would not necessarily be dependent on achieving high accuracy, realism or speed of media delivery in these domains, but should be far more focused on achieving the desired effect - greater audience engagement.

Returning to The analogous performing arts scenario: The Stage once more, the stage set and props do not have to be real in order to achieve catalysis of the projected narrative in portrayal of the performance. In contrast to programming languages and their associated technical, mathematical and scientific goals, which drove the operating system of the Articulated Head, the challenges presented here were artistic in nature with agency, affect and engagement considered more important than accuracy and perfection in the details of technical and creative design related to these projects. I.e., if a desired aesthetic goal can be reached by simple means, without specific attention to technological detail and accuracy then those simple means should be utilised.

6.2 Developmental Stages and Logistical Impingements
The Articulated Heads auditory and visual augmentations were implemented in developmental stages in parallel with the collection of Video Cued Recall Interviews, which were conducted across the three exhibition spaces that the Articulated Head was exhibited in during this study: The New Instruments for Musical Expression (NIME) (“NIME International Conference,” 2014) exhibition in June/July 2010, The Somatic Embodiment, Agency & Mediation in Digital Mediated Environments (SEAM) exhibition (“SEAM2010,” 2012) in the Seymour Centre (“The Seymour Centre,” 2010),

This study was conducted between April 19th 2010 and April 18th 2013. The first of the three exhibitions that the Articulated Head was exhibited at during this study, the NIME exhibition, took place in June 2010 before any ideas for new media augmentations to the Articulated Head’s interactive environment were solidified. Initial findings based on observation of the Articulated Head in operation and Video Cued Recall Interviews collected at the NIME exhibition (prior to any detailed qualitative data analysis and node coding was conducted) contributed to the development of the ideas development for the augmentations. The SEAM exhibition was held in October 2010 and by this time the concept for both the projections and the spatial auditory system had been discussed, but the development of both projects were still very much in their infancy. The equipment required to implement the projects was being purchased at the time. A prototype version of the initial projection idea as detailed in Section 6.9 The Projections were tested at the SEAM exhibition but no Spatial Audio System had been developed at this point. Therefore, most of the development and testing of these creative new media projects took place with the Articulated Head exhibited in the Powerhouse Museum.

6.3 Cost Implications and Technical Constraints

There were cost implications, aesthetic artistic considerations and a range of practical and technological limitations, which both constrained the Articulated Head’s performance capabilities and restricted the number of developmental avenues that were likely to prove fruitful in terms of “conditioning affect and engagement” within the timeframe of this study, and more particularly the Articulated Head’s residency at the Powerhouse Museum ("Powerhouse Museum | Science + Design | Sydney Australia," 2012), because this was where most of the developments to the auditory visual environment took place.
6.3.1 Outline of the cost implications
A full breakdown of the equipment and costs is presented in Appendix 3 [see Table 12.1.] The cost implications related to the creative new media developments proposed, and the concerns they may have raised, were partially dispelled by the plug and play nature of the creative additions, and the adaptability and transferability of the resources required to facilitate the audio/visual work. If all the hardware and software resources required for the auditory visual additions *other than some of the wiring* were retrievable from the Articulated Head installation at the Powerhouse Museum once the exhibit was dismantled, the equipment would subsequently be reusable for other applications within The MARCS Institute (MARCS, 2012); therefore the unrecoverable technology and equipment costs of the creative new media projects to the Institute would in fact be relatively small.

Some aesthetic artistic considerations, and practical or technological constraints proved to be persistent barriers to improvements of the Articulated Head’s virtual performance that were, at least to some degree, irresolvable within the timeframe of this study. Examples include;

- The space available for a projection area around the exhibit
- The constraints on speaker mounting positions for the Spatial Audio System
- The support for testing and implementation, accuracy and usability of automatic speech recognition technology in exhibition environments
- The problem of deciding what the Articulated Head appears to pay attention to when surrounded by crowds of people

However, many of these aspects are important, and do clearly feature in section 9; *The Blueprint of Emergent Recommendations* that has been derived from the findings from this investigation.

6.4 Project links to agency, affect and engagement
The following table was created in the planning stage of these projects and is provided to make explicit how the creative project outcomes were intended to link directly to agency, affect and engagement.
### Project 1 – Textual projections

**Agency**
The sprites, which display text, should change according to User and/or Chatbot text string input to the projection system, derived from the current conversational foci of interactions taking place. The word associations relayed in the projections should demonstrate some semantic correlation with user input and thus promote suggestive evidence of thought and consciousness of the avatar, the result of which should be perceived agency within the eyes and minds of the audience interacting with it.

**Affect**
The affect expected from this projection initiative was that the audience interacting with the Articulated Head would experience a heightened sense that the Articulated Head showed evidence of thought and cognitive processing related to its responses to user input during interactions.

**Engagement**
This projection initiative was expected to provide enhancement of plausible cognitive links between user input and Chatbot output, therefore increasing engagement of the audience by providing a more cohesive conversational experience.

### Project 2 – Spatial Auditory Cues

**Agency**
Spatial audio should be used in two ways here: (1) the system should be used to make the voice of the avatar follow the position of the face on the screen on the end of the robotic arm. (2) THAMBS moods/modes and co-ordinates for centre’s of attention should be used to control and affect spatial auditory information giving the Articulated Head’s audio output agency in relation to its attention model.

**Affect**
The affect expected from these auditory initiatives was that when an audience interacted with the Articulated Head, they should have experienced enhanced attention to their presence as the Articulated Head’s voice and other spatial auditory cues would be spatiotemporally and contextually aligned to their immediate experience.

**Engagement**
It was expected that an audience interacting with the Articulated Head would report enhanced engagement, because the Articulated Head’s voice and other spatial auditory cues would be spatiotemporally and contextually aligned to their immediate experience.
To summarise the above table: the creative projects, designed to embrace the ways in which the avatar already had agency, were multisensory and could be multi-present. Furthermore, the projects sought to enhance the ways in which characteristics of the virtual performer condition affect and engagement, therefore directly advancing propositions for evaluation through the qualitative data analysis methods of investigation adopted during this study. The evaluation of these projects is detailed under Theme 12: Auditory Visual Interactive Environment.

With reference to placement of the robots voice in the spatial auditory system as cited in the table above, one might ask, why not mount a loudspeaker at the screen position on the end of the Articulated Head’s industrial robotic arm but a loudspeaker large enough to project the voice at high quality to the audience would have been heavy and put extra strain on the robotic arm. Furthermore because the Articulated Head was surrounded by a high walled glass enclosure the internal acoustic nature of the enclosure would render the results of said speaker placement muffled and unclear to an audience present on the other side of the glass enclosure due to reflected sound waves, standing waves and the associated deterioration of the sonic image presented.
6.5 Projected data-flow diagram

The flow diagram in Figure 6.1 below shows the projected plan for data flow between the Articulated Head and the new media augmentations proposed.

![Suggested Flow Diagram of data communications for the AH event manager and implementation of text clouds and spatial audio projects](image)

**Figure 6-1 Data flow diagram for Creative Media Augmentations**

Much of the creative additions technical work documented in this section was developed using Max/MSP/Jitter, now known as simply Max. Max is a computer based, cross platform, object oriented programming environment, which allows the user to link objects. Objects have inputs and outputs and linking is achieved using virtual cables. Linking the objects is executed in what are known as patcher windows on computer screen. A patcher window is the place where patch development starts within the Max environment. A patch may contain many objects and can also contain embedded patches, which open in their own patcher window. When discussing Max objects in the following description of the technical development of this creative new media work *(including any linked documentation)*, any Max object name present in the text will be enclosed in *[name]* brackets. Development descriptions that follow do not include a detailed explanation of every object present within a patch, or every
patcher embedded within any other patch; rather the development
 descriptions are aimed at presenting an overview of the functionality that
 Max and the other software technologies utilised in the development of
 these works, delivered to the projects.

6.6 Overview of the Hardware Equipment Set-up
A simplified overview of the equipment set up for the creative new media
augmentations is given in Figure 6-2 below.

Figure 6-2 Simplified Augmentations Equipment Set-up

Figure 6-2 above belies the complexity of the set-up for the auditory visual
augmentations installed and implemented in the Powerhouse Museum. In
reality there was a large number of soldered connections and wiring to inlay
and install alongside the set-up and construction of the rest of the exhibit.
Construction of the exhibit enclosure and installation of all the equipment,
and the Articulated Head itself, took a team of people approximately three
weeks to complete. A small short low quality iPhone slide show of images
displaying some of the stages of the exhibit construction is included as an
electronic document for reference as E-Appendix 12.
The following two images help to illustrate the complexity of just one part of the set-up; the wiring up of the 19 inch rack mounted equipment in the left-hand cupboard located inside the Video Cued Recall Interview Laboratory area as indicated on the plan view diagram of the exhibit, Section 12.1.1 Appendix 1.

![Installation of wiring: Left hand lab cupboard prior to hardware installation](image-url)
6.7 Introduction to the Software Technologies Used

A detailed description of all of the programming of the software technologies utilised during the development of these creative new media works is not necessary to understand the outcomes. So where possible, flow diagrams and illustration have been used to impart the overview expediently. However, the following developer sourced descriptions of the software technologies used, along with a brief description of the technologies contribution to this work are to help describe the Projection and Spatial Audio System outcomes in E-Appendix 11: Creative New Media Additions Documentation.

6.7.1 Max/MSP

Max/MSP is now known as simply Max with the release of Max5. “Max has been used by performers, artists, and composers to make cutting-edge work by connecting basic functional blocks together into unique applications. Max gives you the parts to make your own music, sound, video, and
interactive media applications. Simply add objects to a visual canvas and connect them together to make noise, experiment, and play”. (Cycling 74, 2012). Max has been used extensively in the creative new media projects providing interfacing and text string management for the projections, WiiMote and iPhone interface control using the C74 object and OSC protocol.

**Max’s main contribution to this work**
Max Patches have played a central role in both the Spatial Audio and the Projection projects that have been implemented and critically evaluated within the context of the Articulated Head’s interactive environment.

### 6.7.2 Jamoma

“Jamoma, a platform for Interactive Arts-based research and performance. Jamoma is a community-based project, dedicated to the development of several toolsets and frameworks for artistic creation through digital means. The project is led by an international team of academic researchers, artists and developers, and supported by several institutions” ("Jamoma.org," 2012).

**Jamoma’s main contribution to this work**
Jamoma was utilised for the creation of a Vector Based Amplitude Panning Spatial Auditory System that was implemented as a prototype on the opening evening of the Articulated Head’s exhibit in the Powerhouse Museum. The Vector Based Amplitude Panning prototype was later upgraded to a Distance Based Amplitude Panning model using the SPAT objects ("Spatialisateur," 2012).

### 6.7.3 Ircam SPAT objects for Max

Ircam SPAT ("Forumnet," 2012) provides a number of objects for extension of the Max-programming environment, which extend facilities for implementation of a range of different spatial auditory configurations.

**Ircam SPAT's main contribution to this work**
Ircam SPAT provided the main tools for creation of the Distance Based Amplitude Panning configuration for the Spatial Audio System implemented in the Powerhouse Museum.
6.7.4 Vizzie

“A new set of modules called VIZZIE to help you create your own unique video programs right away. VIZZIE makes putting it together fun and gets you from start to finish in record time” (Cycling, 2010).

Vizzie’s main contribution to this work

Vizzie modules were utilised in this work for the purposes of creating/mixing various visual effects within the projections.

6.7.5 Osculator

Osculator provides a software link between devices such as a WiiMote, iPhone and several other hardware controllers to video and audio software on a computer. Osculator supports the Open Sound Control Protocol, which makes it easy to interface and send messages between devices and software parameters for control purposes.

Osculator’s main contribution to this work

Osculator worked as the interface between the WiiMote’s and Max.

6.7.6 Open Sound Control

Open Sound Control is a simple network control protocol. “This simple yet powerful protocol provides everything needed for real-time control of sound and other media processing while remaining flexible and easy to implement. Open Sound Control (OSC) is a protocol for communication among computers, sound synthesizers, and other multimedia devices that is optimized for modern networking technology. Bringing the benefits of modern networking technology to the world of electronic musical instruments, OSC’s advantages include interoperability, accuracy, flexibility, and enhanced organization and documentation. There are dozens of implementations of OSC, including real-time sound and media processing environments, web interactivity tools, software synthesizers, a large variety programming languages, and hardware devices for sensor measurement. OSC has achieved wide use in fields including computer-based new interfaces for musical expression, wide-area and local-area networked distributed music systems, inter-process communication, and even within a single application. (Osc, 2012)
Open sound Control’s main contribution to this work
OSC worked as the protocol used to communicate between Osculator and Max for the “WiiMote Patch” and also between Max and Director over a UDP connection.

6.7.7 The CNMAT Library
The CNMAT Library ("Center for New Music & Audio Technologies (CNMAT),” 2013) is a group of Max objects developed by the Centre for New Music and Audio Technologies, University of California, Berkley, which expands the Max programming environments capabilities with extended support for Open Sound Control amongst other things.

CNMAT Libraries main contribution to this work
CNMAT objects have been utilised in various Max patches throughout this creative new media work. The OSC CNMAT objects were a notable contribution to this work.

6.7.8 The [c74] Object
The c74 object is one of a number of solutions that allows you to dynamically create user interface objects on your iOS device from within Max via a networked connection, this is very useful because these user interface objects can pass control information to and from Max allowing for remote control of Max patch parameters on one side of the network whilst also allowing for dynamic control of user interfaces on the iOS device of your choice. Furthermore utilization of the OSC protocol allows Max control of a vast array of other networked hardware and software parameters over Bluetooth, WiFi, and Ethernet (TCP/IP).

[c74] Max object’s CNMAT Libraries main contribution to this work
The [c74] object was used to interface an iPhone remote control surface with Max to control variable parameters related to the creative additions.

6.7.9 Processing
"Processing (Processing, 2012) is an open source programming language and environment for people who want to create images, animations, and interactions. Initially developed to serve as a software sketchbook and to teach fundamentals of computer programming within a visual context,
Processing also has evolved into a tool for generating finished professional work. Today, there are tens of thousands of students, artists, designers, researchers, and hobbyists who use Processing for learning, prototyping, and production (Processing, 2012)

**Processing’s main contribution to this work**
Processing was utilised to create the main Thought Cloud projections.

### 6.7.10 Director

You can “explore exciting new dimensions in the award-winning multimedia authoring environment of Adobe® Director® software and Adobe Shockwave® Player. Create engaging, interactive games, applications, simulations and more for the desktop, kiosks, DVDs, CDs, and the web. Robust, flexible authoring and a streamlined workflow deliver a greater return on your creativity” (“On Location (software),” 2013).

**Director’s main contribution to this work**
Director was programmed to display the sentence and words prototype projections used at the SEAM Symposium and was also utilised for the dripping text projections shown at the Powerhouse Museum.

### 6.7.11 TeamViewer

“TeamViewer - the All-In-One Software for Remote Support and Online Meetings. Remote control any computer or Mac over the Internet within seconds or use TeamViewer for online meetings. Find out why more than 100 million users trust TeamViewer” (“TeamViewer,” 2012).

**TeamViewer main contribution to this work**
TeamViewer was an invaluable tool during this project allowing for remote management and maintenance of the Spatial Audio System and projections. Furthermore it allowed for complete remote control of the computers - so was useful for remote connection to the Articulated Head’s computers by the engineering team and it also allowed for a visual check of the exhibits operation remotely by accessing the data streams sent into the computers from the video cameras.
6.7.12 Alchemy Player

The Alchemy Player ("Alchemy Player," 2013) VST plugin was utilised for creation of an ambient morphing soundscape, generated from within the Spatial Auditory System in Max. A range of other robotic sounding samples including motors, electrical sounds and other electro mechanical samples were used as sound bites delivered to the audience in various perceptual positions within the Distance Based Amplitude Panning spatial audio field using the aforementioned Ircam SPAT ("Spatialisateur," 2012) objects in Max.

6.8 The Creative New Media Projects Development

E-Appendix 11: Creative New Media Additions Documentation displays images related to technical Max Patch implementations. Each image presented includes a brief description of the operation of the patch shown and its contribution to the operation of the projects. A short commentary relating the creative additions follows under the next few headings.

6.8.1 The Projections

Existentialism - the ability of the Articulated Head to act at will or at the very least, appear to act at will to a person interacting with it was considered at the outset of these projects to be essential to that persons sense of other rather than object, robot or it. I.e., perceived acts of will (whether false or otherwise) equal ‘Other’. The turn of phrase from ‘it to him’ captured in a research participant’s Video Cued Recall interview dialogue at the NIME conference (see; Key Evidence Links Table), was a very clear example of a person interacting with the Articulated Head who appears to have subconsciously attributed it with the status as other rather than object, robot or it.

6.8.2 Rationale

The projections aimed to develop the avatar so as to give its audience an enhanced sense that it possessed consciousness, that it exhibited some evidence of thought and memory related to the current conversational foci taking place during the interaction, so as to give the audience a sense of the presence of an - ‘other’ - the presence of a conscious existential being as an active partner in the interaction. It was hoped that projection of words,
either behind or in front of the Articulated Head, displayed on a mounted projection screen of some form, probably using frosted projection film would help to achieve the above aim.

Text was to be gathered from both user input and Chatbot output with the intention that Chatbot output to the voice synthesis engine, which allowed the Articulated Head to speak words out loud, would sometimes be delayed momentarily whilst the previous Chatbot output and subsequent User input text strings were displayed on the projection screen. In practice the insertion of the specified delay proved to be unnecessary because some endogenous variable latency of the Articulated Head’s system operation was already present. It was expected that this simple projection idea would give the audience the impression that the Articulated Head had thought about what it was going to say before saying it.

A possible extension to the simple idea above was that text strings from current conversational foci could be used to derive similes, synonyms, antonyms and other related words for extraction and display from a database - this would allow for the projection of thought clouds - displaying text with relational connections to current conversational foci. It was expected that this would provide stimuli for the audience – food for thought so to speak. It was intended that these stimuli would trigger identification of plausible cognitive links between user input and Chatbot output strings within the audiences’ mind. That is, the audience would imagine variations of possible links between what had been input and the Chatbot’s output strings, based on perceived relational connections between the words displayed on the projection screen, therefore enhancing their perception of connections between conversational foci. The hope was that this trick would enhance engagement of the audience by providing a tangibly more cohesive conversational interactive experience.

In order to achieve the above ideas, consideration of the average audience vocabulary came into question. A paper in a Journal of Literacy Research (Zechmeister, Chronis, Cull, D’Anna, & Healy, 1995) indicates that the
receptive size of a college-educated native English speaker is about 17,000 word families, about 40% more than first year college students, who know about 12,000 word families.

Given the relatively small vocabulary used by most people, the fact that only some of the words they use would provide satisfactory relational word output in projections, and the fact that conversation with the Articulated Head was dramatically constrained by interaction time and the context of engagement, it was thought that a fairly limited database thesaurus of relational words would be sufficient. When considering accessible online database options for production of relational word strings, on the advice of Professor Chris Davis word.net (Princeton, 2012) in particular became of interest. However, it was also considered possible that a faster and more effective database of similes, synonyms, antonyms and the like could be constructed manually from a recorded set of user interaction text strings. This possibility although provisionally identified, was not in fact employed in the construction of the thought clouds in the end. Nevertheless, there are some interesting observations emergent from section 7.1 Text String Input/Output Analysis that lend considerable weight to the idea of employing a variation of a simplified, manually constructed text string database. The suggestion for this variation is carried through into the blueprint of emergent recommendations from this investigation presented in Section 9.1.12 The Auditory Visual Interactive Environment.

The projections could have been developed aesthetically in a number of different ways, but initially the project was constructed using Cycling74 Max and Adobe Director. It was recognized that networking information between Max and Processing could yield more adventurous and pleasing artistic and aesthetic possibilities and so further developments were planned in the Processing (Processing, 2012) direction.

The Projections at SEAM
At the SEAM exhibition the first prototype of the text projection idea was developed, using Max as the software to interface text information between
the Articulated Head’s Event Manager and the projections. Max passed the text strings to Adobe Director sprites for display. Open Sound Control messages were sent from Max to Adobe Director to control the sprites positions in the projections. Figure 6-5 below shows the projections and the Articulated Head set-up at SEAM.

In the Figure 6-5 above, you can clearly see three glass panels (approx. 1.25m wide by 2m high). There were eight Aluminium framed glass panels in total, which together formed an octagonal protective surround for the Articulated Head. The enclosure shown here was the same enclosure that was used at the NIME exhibition. The only difference here is that you can see three panels to the front of the enclosure displaying the presence of frosted projection film on the lower half of the panels. The film displays some multicoloured words (sentences), derived from user input or Chatbot output text strings. The projectors (three of them), are mounted on the ground inside the enclosure below the extended reach of the Articulated Heads arm so as to avoid damage.
Execution of the prototype Text Projections

Simplicity was a necessity with the projections because initially there was only a little time for programming them before the SEAM exhibition. The Articulated Head’s Event Manager required modifying in order to make provision for passing text strings out to Max. The format for passing text strings out was agreed quickly and the prefix-delineators of a 1 to identify User input strings and a 2 to identify Chatbot output strings was implemented. Development of the Event Manager/Max interface was limited to passing text strings out from the Event Manager at this point in time, no provision had yet been made for the passing of text strings or any other data from Max back into the Event Manager.

The Articulated Head enclosure as seen in Figure 6-5 utilised one of its panels (to the rear in the picture shown) as a door. The door was normally locked. If the door was opened then a switch was tripped to automatically stop the Articulated Head’s robotic arm from moving for health and safety reasons. Once the trip switch had been activated then a full restart of the system, which took several minutes, was required to get the Articulated Head up and running again. Naturally it was desirable to keep the door locked, as extended periods of downtime in an exhibition environment is embarrassing.

Diagram 6-6 below indicates the flow of text string data from the Event Manager through to Max via an Ethernet cable. The Data was then passed from Max to Director via the Director OSCxtra (“Adobe (Software),” 2012). The director stage was designed to be 2400 pixels long by 600 pixels high so that the stage could be spread evenly across the entire display area of three projectors (800 by 600 pixel displays each). Figure 6-7 indicates the extended screen layout that the KVM switch module facilitated for the projections. The Director file was designed to display sixteen text sprites placed in various positions on the stage so that the average sentence lengths gathered from the event manager filled the stage area appropriately. The text sprites would randomly display the last sixteen text strings derived from either user input or the Chatbot output. Various display parameters of the text sprites on stage could be controlled, such as colour, size and rotation.
The Articulated Head’s computers including the machines running the Event Manager and the projectors were mounted inside the Articulated Head’s enclosure. Max and Director were running on a separate laptop. It was necessary to connect the laptop to the Event Manager machine via an Ethernet cable to get reliable data transfer, and it therefore made sense to locate the laptop inside the enclosure to keep cabling extending outside of the enclosure to a minimum.
Some experimentation with the size, color and placement of text in the projections to give an aesthetically pleasing result was required. Therefore the design of some sort of remote control system for the prototype projection was needed. To do this an OSC Max patch called the “WiiMote Patch” designed by Kim Cascone, December 2007 (kim@anechoicmedia.com) sourced from my principal supervisor Assoc. Prof Garth Paine was utilised for the purpose of passing control information from the WiiMote through Osculator to the WiiMote Patch. The WiiMote patch was given [send] objects to pass the WiiMote parameters on to the Projections Patch. The Director OSCXtra collected data from the Projection Patch to control sprite parameters. Figure 6-6 a & b and Table 6-3 below indicate the data flow and connection protocols employed:

![Diagram](image)

**Figure 6-8 Director projection control**

a) Projections Data Flow and Protocols
b) [Send] objects (WiiMote Patch ↔ Projections Patch)
End-to-end connections were set up with WiiMote controls as follows:

<table>
<thead>
<tr>
<th>WiiMote 1</th>
<th>WiiMote 2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Controller</strong></td>
<td><strong>Parameter</strong></td>
</tr>
<tr>
<td>Up</td>
<td>Text size up</td>
</tr>
<tr>
<td>Down</td>
<td>Text size down</td>
</tr>
<tr>
<td>Left</td>
<td>Rotate left</td>
</tr>
<tr>
<td>Right</td>
<td>Rotate right</td>
</tr>
<tr>
<td>A</td>
<td>Set straight</td>
</tr>
<tr>
<td>B</td>
<td>Set yellow</td>
</tr>
<tr>
<td>Minus</td>
<td>Multi Color down</td>
</tr>
<tr>
<td>Plus</td>
<td>Multi Color up</td>
</tr>
<tr>
<td>Home</td>
<td>Start fading</td>
</tr>
<tr>
<td>1</td>
<td>Same Color Up</td>
</tr>
<tr>
<td>2</td>
<td>Same Color Down</td>
</tr>
</tbody>
</table>

Table 6-2 - The WiiMote Interface

Figure 6-9 above, displays the locked Projection Max Patch for control of automation of colour, rotation and opacity of text in the projections and for
the text storage and release in the dripping text projections. Figure 6-10 below shows the same patch unlocked.

![Figure 6-10 Director projection embedded patcher window](image)

6.8.3 The Creative Additions at the Powerhouse Museum

Prior to installation in the Powerhouse Museum, the Articulated Head made only rudimentary use of spatio-temporal auditory visual contextualising cues for the enhancement of its interactive performance. The aforementioned spatial auditory systems, and the projections titled the Thought Clouds by Associate Professor Garth Paine, were implemented in the Powerhouse Museum.

**The Spatial Auditory System**

Initially, Genlec 6010A bi-amplified active loudspeakers ("Genelec 6010A is an Extremely Compact Two-Way Active Loudspeaker," 2012) provided monosonification of the Articulated Head’s phonetic audio sample library to the interacting audience. Expansion to a three dimensional spatial auditory system was proposed for three explicit reasons: 1) For psychoacoustic placement of the robot’s voice in the interactive environment, with the intention that the voice of the robot would appear to emanate from the mouth of the head present on the screen, regardless of the screen position.
within the interactive environment. The rationale for this was that engagement would increase with a sense that the voice and the visualisation of the head were co-located, as is the nature of a real human being moving around in the interactive space. 2) For the presentation of other contextually relevant spatial auditory cues, which link visual movement to features of auditory output. 3) For the sonification of a complimentary soundscape, the intention being to create a more intimate auditory environmental experience for patrons of the Powerhouse Museum in the surroundings of the exhibit. To give the exhibit its own individual and separated, yet still contextually linked auditory space in relation to the wider museum environment surrounding it.

The plan for presenting contextually relevant spatial auditory cues was to map specific features of synthesized auditory material such as the embodied conversational agent’s voice or parameters of a particular sound patch such as the cut off and resonance filters and/or a panning control to robotic arm position co-ordinates and movement. Coordinates were to be provided to the spatial audio system by the Event Manager/Max interface, which was specifically programmed for the task of passing data out from the Event Manager to Max by Zhengzhi Zhang. Linking physical or visual movement to correlating complementary auditory output is known to be effective in solidifying events into one unified percept (Lewald, Ehrenstein, & Guski, 2001).

The plan for the soundscape was to contextually link the themes of robots, films and space. These themes were chosen, as they were prevalent subjects of interest arising in audience discussion and information exchange in relation to, and in interaction with the Articulated Head. Robots, films and space also featured strongly in many other exhibits throughout the Powerhouse Museum. Use of sounds such as those generated by the robotic droids R2D2 and C3PO in the film Star Wars (Lucasfilm, 2012) mixed with synthesized sounds similar to those used in science fiction films such as Star Trek (CBS-Entertainment, 2012) and mechanical sounds, such as those generated by clockwork winders, small Meccano (Meccano, 2012) motors and alike, were employed within the spatial audio composition, with the express intention of promoting a sense of intertextuality (Kristeva, 2002).
between these themes within the minds of patrons visiting the exhibit. Intertextuality means the intermingling or the shaping of one text’s meaning with another text and it applies equally to associations made between features of audiovisual content. Parenthetically Star Trek is a favourite of Stelarc’s and the Articulated Head often referred to Star Trek when asked “what do you like to watch on TV?” or “what is your favourite film?”

**The Thought Cloud Textual Projections**

The Articulated Head was promoted as a Thinking Head and not all of its responses to user input were immediately logical or made proper sense within the context of the conversation taking place, therefore the employ of intertextuality within the visual projection domain was considered. Intertextuality can be intentional or accidental, designed to convey a specific message to an audience, who perceive the message as intended - or a message perceived by an audience, which may not necessarily be the express intention of the writer but is more a construct of self interpretation. Therefore, it was proposed that visualisation of contextualising cues in the form of similes, synonyms, antonyms and other words, related to either user input, or the responses of the robot's embodied conversational agent, would further enhance the illusion that the Articulated Head was indeed thinking, that a thought process of deduction was actually taking place.

The proposition was that relationally linked words would make for a more cohesive interaction by extending the number of stimuli for audience perception. The plan was to present the words from a cloud of jumbled text characters, hence the name “Thought Clouds”. NextText ("NextText," 2012) is a Java library for building applications to display dynamic and interactive text-based applications and NextText for Processing (Processing, 2012) is a port of the library for the Processing development environment. An example file provided with the library called “WhatTheySpeakWhenTheySpeakToMe” - is an interactive artwork by Jason Lewis and Elie Zananiri, with the Processing port by Elie Zananiri | Obx Labs | June 2007. This code formed the starting point for the programming of the Thought Clouds. Dr Richard Leibbrandt
from Flinders University, Adelaide, South Australia took on the task of programming the Thought Clouds.

Some alternative projections were required to work as a placeholder in the Powerhouse until the Thought Clouds were ready. The plan was to extend the SEAM projections using Adobe Director "Adobe (Software)" 2012) Lingo, an interactive multimedia programming language and development environment. The functionality of these projections required the same input and output system from the Event Manager/Max interface as the Thought Clouds, but was intended to display a simpler use of the I/O word data. The plan was to have dripping text characters, bearing an unmistakable resemblance to the display of green text characters synonymous with the opening and closing sequences of the film, The Matrix. (Warner Bros, 2012) Word extracts from user input or embodied conversational agent output would appear from within the dripping text. The intention here was to accentuate the robot theme, drawing on popular culture influences including, films, space and encouraging intertextuality within the placeholder projections.

The dripping text projections and a Jamoma vector based amplitude panning spatial audio system were up and running at the opening of the Articulated Head exhibit in the Powerhouse Museum at the beginning of February 2011. The SPAT ("Forumnet," 2012) distance based amplitude-panning system and the Processing (Processing, 2012) Thought Clouds were developed and implemented over the first few of months of the exhibits display.

Evaluation of the impact and effectiveness of these creative new media additions to the Articulated Head’s interactive environment are detailed in Theme 12: Auditory Visual Interactive Environment. A projected design refinement targeted at similar auditory visual interactive environments and exhibits, based on the findings from the evaluation presented in Theme 12 is imparted in section 9.1.12 The Auditory Visual Interactive Environment. For more details regarding the technical execution of these projects please see
the E-Appendix 11: Creative New Media Additions Documentation, which is supplied as an electronic appendix. Section 7 that follows presents the research data and findings.
7. Presentation of Research Data and Findings

This section presents research data from text string analysis and from analysis of dialogue and events that took place during video cued recall interviews. Details of how the data was qualitatively and/or quantitatively analysed, and the emergent themes from this investigation are given. The themes and supporting evidence are presented with a narrative emanating from the research data analysis and coding processes. After presentation of the themes, Section 8 considers key concepts that help explain the findings and inform the blueprint that follows. Section 9 then presents the blueprint of emergent ideas related to both the new media augmentation projects and the wider research questions related to agency, performance and engagement in general. The blueprint puts together a set of developmental recommendations for the improvement of human ↔ machine interaction based on the emergent themes and findings from this investigation and the key concepts that help to explain them.

Please note, the interpretive account of the narrative emanating from the research data collected during this investigation, will present you with hyperlinks in the text that will allow you to review textual or auditory visual supporting and substantiating evidence, which bolsters the validity of the interpretive narrative. Microsoft Word, Excel and Adobe Acrobat Reader are required to review textual evidence interactively. Windows Media Player (Windows) or QuickTime (Mac) - with browser support for Windows Media Video files (.wmv) will be required in order to view the auditory visual data, which is highly recommended. Windows Media Video files (.wmv) support for Mackintosh computers can be installed using Flip4Mac (Telestream, 2013). Node textual data is too expansive to be supplied as a hard copy appendix but supporting textual evidence is supplied through the Key Evidence Links Table electronically. All the textual and auditory visual content is supplied on an install DVD. The anonymity of participants and the public are maintained throughout the evidence presented. However media release forms were collected from participants because; although their names are not
projected in the textual evidence, they may of course be present in the videos.

7.1 Text String Input/Output Analysis
Text input by the User interacting with the Articulated Head’s A.L.I.C.E Chabot and the Chatbot’s response strings were stored into text files by a Max patch every 30 minutes, for the creation of a data log. This logging process resulted in many thousands of text files for analysis. User input strings were prefixed with a ‘1’ and Chabot output strings prefixed with a ‘2’. The data log spans June 2011 to November 2012 - 17 months. The log files do not include every single interaction that took place during the exhibits time in the Powerhouse Museum, as logs before June 2011 were not captured and the logging process was interrupted by various crashes and shutdowns. However the logs do include the vast majority of all interactions that took place during the time span stated, and therefore, they are considered to be highly reflective of conversational content taking place between Users and the Chatbot. Furthermore because the text data logs are not restricted to the research participant frame, within which all other concrete evidential data collected in this investigation is constricted, it was thought that if a strong correlation exists between the text analysis and the rest of the research participant data collected, then it would strengthen and ratify the themed interpretive phenomenological analysis of data within the research participant frame considerably.

This text string input/output analysis acts as a sort of quality control mechanism in this investigation by providing a view of the conversational interaction that took place, which is largely independent of any transient environmental or operational conditions of the Articulated Head at the time. The data is also largely independent of any influences and preconceptions that might be present within the participant frame. It should be noted that node coding but not theme identification of the data collected within the research participant frame, was in fact conducted prior to this text input/output analysis, even though they are presented in the reverse order in this section of the thesis. In line with a Grounded Theory approach, the text input/output analysis served as purposeful sampling across a different data
set, facilitating the crosschecking of the fit of theory to the data. The reason for presenting the text input/output analysis first is that it helps cement the validity and relevance of the emergent themes from nodal analysis. The emerging pattern from the text input/output analysis does indeed fit and substantiate the identification of critical elements at work in the interaction. There is a very strong correlation with two particular themes that had already been identified from Nodal analysis within the frame of research participant data. The two themes of interest are titled: Theme 5: Movement, Tracking and Control & Theme 6: Dialogue. The text input/output analysis also lends some weight to the other themes.

The analysis of text data that follows in this Sub Section focuses primarily on User input strings and Chatbot output strings as separate entities. Theme 6: Dialogue; then brings the findings from this input/output string analysis to compare it with the nodal data analysis of conversational dialogue that preceded it. The nodal analysis of conversational dialogue looks at User input strings and Chatbot output strings together, to consider such issues as nonsensical input from the User, or vice versa, nonsensical output responses from the Chatbot.

User input text had an inordinate amount of spelling errors, general typo’s, mobile phone speak, slang, swearing and other spurious or erroneous data input, including colloquialisms and characters entered by small children fiddling with the keyboard, therefore an accurate spelling mistake count was not practical. However, it is clear from review of the Nodal data presented throughout Section 7 that data input did have many spelling errors. Chatbot output strings are more consistent than user input - but also have some erroneous data strings present, including strange spellings and some words sewn together, an example of which is present in table 7.35 under the Chatbot phrase query lists that follow in Section 7.1.2. I believe the programmers executed the words sewn together because the text to speech engine sounded better that way. Furthermore to the above anomalies in input/output textual data, a significant number of Chatbot responses do not make sense in relation to the User inputs that herald them,
see Theme 6: Dialogue for more details. Therefore, the text input/output data is considered something of a quagmire in the first place. For this reason all text analysis is based on near approximations and only clearly indicative numerical string search counts are used in input/output text data analysis, to minimize the chances of misinterpretation of the data. Any specific words and/or phrases used from the textual data for queries are referred to by placing them between inverted commas in the following text, like so; ‘word or phrase’. In total the text log files generated a word count of approximately 3 million words or 6 thousand pages of data. However it should be noted that these figures are not very useful and somewhat vague because they count all spurious as well as useful input/output.

To establish whether the textual data could reveal anything important about the conversational interaction-taking place, the following approach was adopted: The text files (approx. – 1488 files per month) were organized into monthly data folders. The files that were stored outside Museum opening times had little or no data in. An Automator - Apple script workflow was built to merge monthly text data files into one long text file for each month, which was saved, resulting in 17 monthly long text files. Another Automator - Apple script workflow was built to separate User input strings from Chatbot output strings using the saved monthly long text file. The script then saved the separated data into two separate text files one for the User input strings and one for the Chatbot output strings. This process was repeated on the long text files for each month resulting in 34 files in total, 17 monthly files for the User input and 17 for the Chatbot output. All the resulting files from the processing described above were imported into NVIVO, which is a Computer Assisted Data Analysis Software program for analysis.
A word frequency search was conducted on the User input string files to find the top 10 words present in User input strings with the following result:

<table>
<thead>
<tr>
<th>Word</th>
<th>Length</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>You</td>
<td>3</td>
<td>50892</td>
</tr>
<tr>
<td>Is</td>
<td>2</td>
<td>24864</td>
</tr>
<tr>
<td>Are</td>
<td>3</td>
<td>24388</td>
</tr>
<tr>
<td>What</td>
<td>4</td>
<td>23702</td>
</tr>
<tr>
<td>Do</td>
<td>2</td>
<td>17214</td>
</tr>
<tr>
<td>Your</td>
<td>4</td>
<td>16338</td>
</tr>
<tr>
<td>How</td>
<td>3</td>
<td>14414</td>
</tr>
<tr>
<td>Name</td>
<td>4</td>
<td>13765</td>
</tr>
<tr>
<td>I</td>
<td>1</td>
<td>11112</td>
</tr>
<tr>
<td>A</td>
<td>1</td>
<td>9713</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>206402</td>
</tr>
</tbody>
</table>

Table 7-1 - User input strings - top 10 words

The same word frequency search was conducted on the Chatbot output string files to find the top 10 words present in Chatbot output strings with the following result:

<table>
<thead>
<tr>
<th>Word</th>
<th>Length</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>1</td>
<td>80189</td>
</tr>
<tr>
<td>You</td>
<td>3</td>
<td>64706</td>
</tr>
<tr>
<td>Is</td>
<td>2</td>
<td>37883</td>
</tr>
<tr>
<td>A</td>
<td>1</td>
<td>36821</td>
</tr>
<tr>
<td>To</td>
<td>2</td>
<td>32620</td>
</tr>
<tr>
<td>The</td>
<td>3</td>
<td>28722</td>
</tr>
<tr>
<td>Are</td>
<td>3</td>
<td>25249</td>
</tr>
<tr>
<td>Do</td>
<td>2</td>
<td>25044</td>
</tr>
<tr>
<td>Me</td>
<td>2</td>
<td>21703</td>
</tr>
<tr>
<td>That</td>
<td>4</td>
<td>21581</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>374518</td>
</tr>
</tbody>
</table>

Table 7-2 - Chatbot output strings - top 10 words

The results from these word searches are not surprising with short prepositions featuring strongly. The appearance of the words ‘what’ and ‘how’ in the Users top ten list is consistent with the User asking the Articulated Head questions, which was of course its intended function. Not much can be established from these searches other than to suggest perhaps that the User
input, having ‘you’ as its number one most used word, indicates that the User was more interested in the Chatbot than in themselves with ‘I’ coming in at number ten. Whereas the Chatbot, having ‘I’ as it’s number one most used word, appears to have been perhaps more interested in itself than the User as the word ‘you’ is used approximately 15000 times less over the data span. The other thing that appears fairly obvious from this initial word frequency search is that, the reciprocity of conversation was lopsided; the Chatbot appears to have had roughly twice as much to say as the User based on the total count of word instances in both tables with the ratio at 1.81:1. This point of observation based on the two tables above is indicative of the mode of information exchange (text in – speech out) representing difficulties of data string input for the User, as is detailed and evidenced in the Themes that follow later in Theme 4: Mode of Interaction.

The next step for text input/output text string analysis was to expand the word frequency search to the top 100 words and then drill down on a group of selected nouns, verbs and adjectives to see if any emergent patterns in conversational foci could be established.

The top 100 words in User input strings are shown as E-Appendix 13: Top 100 User Input String Words and E-Appendix 14: Top 100 Chatbot Output String Words respectively. When viewing these tables you will find the selected words highlighted by the background color Red. The chosen groups of words are shown in the two tables on the following page;
### Table 7-3 User Input (left) Chatbot Output (right) Selected Search Words

<table>
<thead>
<tr>
<th>Position</th>
<th>Word</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Name</td>
<td>13765</td>
</tr>
<tr>
<td>2</td>
<td>Like</td>
<td>6156</td>
</tr>
<tr>
<td>3</td>
<td>Old</td>
<td>6015</td>
</tr>
<tr>
<td>4</td>
<td>Say</td>
<td>3041</td>
</tr>
<tr>
<td>5</td>
<td>Sing</td>
<td>2007</td>
</tr>
<tr>
<td>6</td>
<td>Favourite</td>
<td>1510</td>
</tr>
<tr>
<td>7</td>
<td>Love</td>
<td>1170</td>
</tr>
<tr>
<td>8</td>
<td>Dance</td>
<td>1107</td>
</tr>
<tr>
<td>9</td>
<td>Colour</td>
<td>1075</td>
</tr>
<tr>
<td>10</td>
<td>Life</td>
<td>1001</td>
</tr>
<tr>
<td>11</td>
<td>Cool</td>
<td>983</td>
</tr>
<tr>
<td>12</td>
<td>Robot</td>
<td>906</td>
</tr>
<tr>
<td>13</td>
<td>Harry</td>
<td>871</td>
</tr>
<tr>
<td>14</td>
<td>Live</td>
<td>834</td>
</tr>
<tr>
<td>15</td>
<td>Eat</td>
<td>827</td>
</tr>
<tr>
<td>16</td>
<td>Happy</td>
<td>824</td>
</tr>
<tr>
<td>17</td>
<td>Time</td>
<td>756</td>
</tr>
<tr>
<td>18</td>
<td>Speak</td>
<td>665</td>
</tr>
<tr>
<td>19</td>
<td>Think</td>
<td>665</td>
</tr>
</tbody>
</table>

The tables above do not show direct correlation between the words chosen from the User input Top 100 list in comparison with the Chatbot output Top 100 list, simply because the use of vocabulary of the User did not correlate with the Chatbot vocabulary. A few more words have been selected from the User input Top 100 list over the Chatbot output Top 100 list. Words chosen for further examination were picked on the basis that they present clear subjects of relevance. The aim was to avoid the ambiguity that some more frequent but less prominent words in the top 100 lists might introduce into the analysis, because they have vast arrays of possible conversational uses in phraseology. Choice of appropriate words from the two lists was more important than the number of words chosen. Given this investigation focus
on the human experience of interaction with the Articulated Head, a slightly stronger emphasis on User input as oppose to Chatbot output in this analysis might be more appropriate and desirable anyway.

Each of the selected words from the top 100 User input and Chatbot output strings were entered into a string search (query) and the resulting references were reviewed to try to find the most common group of words surrounding the search string word. For example; if the search word was ‘favorite’ a surrounding conversational foci string might read ‘what is your favorite song’ or ‘my favourite time of day is’. Queries were constructed in NVIVO to include both English and American spellings of words where appropriate, such as ‘colour & color’ or ‘favourite & favorite’ E-Appendix 15 - Favourite - April2012-Results Preview for an example of a search result).

The aim was to catch as many instances of phrases with the same meaning as possible. Once a recurring common group of words surrounding a search string word was identified, the recurring group of words, normally limited to two or three because of the speed of data processing, then formulated the search string for the next query directed at the data. This process of phrase identification and query construction was conducted on both User input and Chatbot output strings. The results of this exercise were initially entered into two Excel spreadsheets in the order in which the identified phrases emerged from the text data E-Appendix 16: - User Input String Phrase Table and E-Appendix 17: Chatbot Output String Phrase Table.

An attempt to get 70% coverage of the initial search word count, by subsequent phrase searches was made. The choice of 70% was made on the basis that the vast majority of recurring common groups of words were picked up in the first 70% in initial searches and above this figure the commonality depleted rapidly and erroneous data began to appear in the phrases. The data was extrapolated from the linked tables above and assembled into the lists presented in the tables in subsection 7.1.1 and 7.1.2 below. Each list was either topped up to 70% coverage, or a reason is identified below the table that explains why this was not possible. All lists, and
the phrases presented in them are in the most > least significant word or phrase order, based on the frequency of that word or phrase. A short statement of observations emerging from the information follows each table. Subsection 7.1.3 then looks at key observations made from examination of the table data in more detail.
7.1.1 User Input String Lists

<table>
<thead>
<tr>
<th>Name</th>
<th>13765</th>
</tr>
</thead>
<tbody>
<tr>
<td>what is your name (5574)</td>
<td>8030</td>
</tr>
<tr>
<td>my name is</td>
<td>3871</td>
</tr>
<tr>
<td>your name is</td>
<td>85</td>
</tr>
</tbody>
</table>

**Table 7-4 - Word ‘name’ phrase query list**

The word name as the Users most common word is not really surprising and with a very significant proportion of instances grouped with the phrase: ‘what is your name’, it can be said that this is likely to be one of the most common phrases input by Users because name is the most common word in the user input strings. Although there is no reliable delineator of the start and stop of individual interactions in the text data, it is likely that the word ‘name’ is used by the User or the Chatbot at the outset of many interactions. The frequency of the use of the phrase ‘my name is’ suggests that the Chatbot regularly asks Users what their name is.

<table>
<thead>
<tr>
<th>Like</th>
<th>6156</th>
</tr>
</thead>
<tbody>
<tr>
<td>(do 3696) you like (misc * aligns favourite)</td>
<td>4017</td>
</tr>
<tr>
<td>I like</td>
<td>817</td>
</tr>
</tbody>
</table>

**Table 7-5 - Word ‘like’ phrase query list**

The phrase ‘you like’ has hundreds of appendages in the data with examples including rag dolls, kittens, space travel, cheese, chocolate, surfing and so on. Instances of these appendages are diverse and seem to be based on the likes/dislikes of the human inquisitor. Based on the frequency of the “I like” phrase, it appears that the User is much more inclined to question the Articulated Head’s preferences rather than inform it of their own. The inclination towards establishing the Articulated Heads preferences based on the likes/dislikes of the human inquisitor is discussed
more in Observation 1 – The anthropomorphic stance & in Theme 1: Anthropomorphism.

<table>
<thead>
<tr>
<th>Old</th>
<th>6015</th>
</tr>
</thead>
<tbody>
<tr>
<td>(how) old are you</td>
<td>4331</td>
</tr>
<tr>
<td>Coverage</td>
<td>72%</td>
</tr>
</tbody>
</table>

Table 7-6 - Word ‘old’ phrase query list

The phrase ‘(how) old are you’ is prefixed with ‘how’ almost exclusively in the text data. The only reason for the leaving the word ‘how’ out of the phrase search string is to speed up data processing of the query. It is clear that ‘what’s your name’ and ‘how old are you’ are very common phrases used by the general public during interaction.

<table>
<thead>
<tr>
<th>Say</th>
<th>3041</th>
</tr>
</thead>
<tbody>
<tr>
<td>Erroneous for phrase search</td>
<td></td>
</tr>
</tbody>
</table>

Table 7-7 - Word ‘say’ phrase query list

The reason that the word ‘say’ is erroneous for a phrase search is that the public interacting with the Articulated Head discovered that it was possible to make the robot say anything by prefixing the sentence typed in with the word. For example; if one enters ‘say Cornflakes’ into the kiosk keyboard, the Articulated Head will say Cornflakes. This control word became a point of great interest, especially with school kids resulting in a number of expletives and other questionable or rude statements as well as many more benign uses present in the text data E-Appendix 18: [Example 1]. It is a bit of a mystery how so many people got to know of the control word because new visitors to the museum appeared to know that they could use it – perhaps the school kids grapevine, the bush telegraph, is more powerful than I had realised? This control word anomaly is discussed in more detail under Theme 5: Movement, Tracking and Control.
Table 7-8 - Word ‘sing’ phrase query list

<table>
<thead>
<tr>
<th>Sing</th>
<th>2007</th>
</tr>
</thead>
<tbody>
<tr>
<td>can you sing</td>
<td>450</td>
</tr>
<tr>
<td>sing vincero</td>
<td>256</td>
</tr>
<tr>
<td>sing happy birthday</td>
<td>450</td>
</tr>
<tr>
<td>sing a song</td>
<td>142</td>
</tr>
<tr>
<td>sing je crois</td>
<td>102</td>
</tr>
<tr>
<td>sing for me</td>
<td>34</td>
</tr>
<tr>
<td>sing for us</td>
<td>26</td>
</tr>
<tr>
<td>sing something</td>
<td>20</td>
</tr>
</tbody>
</table>

Coverage 61%

Note: lots of solo instances of ‘sing’

The ability for the Articulated Head to simulate singing was introduced during its time at the Powerhouse Museum and so most, if not all instances of the word ‘sing’ related to its time there. Getting the coverage of the word ‘sing’ to over 70% was not possible because of the vast number of solo instances of the word. The reason for this is probably because the public was aware that the robot could sing as a result of museum staff regularly typing in the trigger strings, but the public did not know the trigger strings so tried the word ‘sing’ regularly. In fact ‘sing vincero’, ‘sing je crois’ or typing ‘it’s my birthday’ were the triggers, which explains their high occurrences in the list. The museum volunteer guides and permanent staff were aware of the trigger words and used them often when passing the exhibit. The other phrase instances in the table attempt to establish whether the robot can sing. How many of the phrase instances are attempts to trigger the robot to sing is unclear. However, as is the case with the word ‘say’ above, the control/triggering of a response from the Articulated Head was also at work with the word ‘sing’ and control words or phrases have implications for engagement in interaction as discussed in Theme 5: Movement, Tracking and Control.
The word ‘favourite’ and surrounding phrases shows a strong correlation with the word ‘like’ and its surrounding phrases. Links are also made to the words ‘colour’ and ‘love’ in the tables below. Both words are linked to evidence that suggests the Users are testing the robots preferences in a specific way, this is discussed in more detail after these tables, under Observation 1 – The anthropomorphic stance & in Theme 1: Anthropomorphism.

<table>
<thead>
<tr>
<th>Favourite</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>favourite colour</td>
<td>643</td>
</tr>
<tr>
<td>favourite food</td>
<td>324</td>
</tr>
<tr>
<td>favourite movie</td>
<td>84</td>
</tr>
<tr>
<td>favourite sport</td>
<td>56</td>
</tr>
<tr>
<td>favourite animal</td>
<td>49</td>
</tr>
<tr>
<td>favourite song</td>
<td>43</td>
</tr>
<tr>
<td>favourite thing (&quot;to do&quot; etc)</td>
<td>40</td>
</tr>
<tr>
<td>favourite game</td>
<td>34</td>
</tr>
<tr>
<td>favourite book</td>
<td>20</td>
</tr>
<tr>
<td>favourite country</td>
<td>9</td>
</tr>
<tr>
<td>favourite city</td>
<td>6</td>
</tr>
<tr>
<td><strong>Coverage</strong></td>
<td>70%</td>
</tr>
</tbody>
</table>

**Table 7-9 - Word ‘favourite’ phrase query list**

The word ‘colour’ is clearly linked to ‘favourite’ and is also discussed under Observation 1 – The anthropomorphic stance & in Theme 1: Anthropomorphism.

<table>
<thead>
<tr>
<th>Colour</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>favourite colour</td>
<td>643</td>
</tr>
<tr>
<td>what colour</td>
<td>359</td>
</tr>
<tr>
<td><strong>Coverage</strong></td>
<td>83%</td>
</tr>
</tbody>
</table>

**Table 7-10 - Word ‘colour’ phrase query list**
Table 7-11 - Word ‘love’ phrase query list

<table>
<thead>
<tr>
<th>Phrase</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>I (351) love you</td>
<td>387</td>
</tr>
<tr>
<td>you (117) love me</td>
<td>139</td>
</tr>
<tr>
<td>you love</td>
<td>117</td>
</tr>
<tr>
<td>is love</td>
<td>30</td>
</tr>
<tr>
<td>love it</td>
<td>18</td>
</tr>
<tr>
<td>in love</td>
<td>17</td>
</tr>
<tr>
<td>feel love</td>
<td>7</td>
</tr>
<tr>
<td>your love</td>
<td>6</td>
</tr>
<tr>
<td>to love</td>
<td>5</td>
</tr>
<tr>
<td>love God</td>
<td>5</td>
</tr>
<tr>
<td>love Jesus</td>
<td>2</td>
</tr>
</tbody>
</table>

Note: multitude of misc; chocolate, cats, movies, sport – shows alignment with favourite and like

A [Sample] of ‘Loves’ search results was exported for reference as it was hard to get the coverage figure up to high percentage. The word ‘you’ almost always follows ‘I love’, although there are some instances of ‘food’ ‘movies,’ Whereas “you love” is predominantly preceded by ‘do’ followed by a range of subjects ‘food (types chocolate, cheese, sushi)’, ‘movies’, ‘tv’ ‘Justin Beiber’.

The word ‘love’ has a large number of diverse appendages in the data, making it very difficult to attain 70% coverage. However, it is linked to the word ‘favourite’ because of an identified correlation in subjects appended to the word such as ‘chocolate’, ‘cats’, ‘movies’ and ‘sport’. This link is also discussed under Observation 1 – The anthropomorphic stance & in Theme 1: Anthropomorphism.
‘Dance’ is linked with the control words ‘say’ and ‘sing’ because it appears that Users are either attempting to establish whether the robot can dance, or are trying to trigger it to do so. Control words are discussed in more detail under Theme 5: Movement, Tracking and Control. The Articulated Head did have a few preprogrammed movements installed for a performance that took place in the museum but they were not displayed in normal operation. The museum did have a dancing robotic arm exhibit and it is thought that this may have had an influence on Users inclinations to check the Articulated Head’s capabilities for dance.

The high occurrence of the phrase ‘meaning of life’ was predominantly prefixed with the words ‘what is’ and sometimes appended with ‘the universe and everything’, this phrase is synonymous with the Douglas Adam’s third science fiction book in the series titled: Hitchhikers Guide to the Galaxy (Adams, 1984). The Users question in this context is very apt given the nature and context within which the Articulated Head was presented. Notably the
Articulated Head did not answer 42, which is the answer to the question in the book.

<table>
<thead>
<tr>
<th>Cool</th>
<th>983</th>
</tr>
</thead>
<tbody>
<tr>
<td>is cool</td>
<td>76</td>
</tr>
<tr>
<td>are cool</td>
<td>65</td>
</tr>
<tr>
<td>that’s cool</td>
<td>20</td>
</tr>
</tbody>
</table>

**Coverage**: hundreds of solo instances of 'cool'

**Table 7-14 - Word ‘cool’ phrase query list**

The word ‘cool’ turned out to be unhelpful for phrase searches, mainly because of the high frequency of solo instances. However, the solo instances are indicative of an aspect of human ↔ human conversation that doesn’t work so well in human ↔ robot conversation. Humans often say ‘cool’ as single word responses during conversation with someone to show that they understand, agree and/or approve of what the other person is saying but when the Chatbot was confronted with this single word input conversational flow appeared to falters.

<table>
<thead>
<tr>
<th>Robot</th>
<th>906</th>
</tr>
</thead>
<tbody>
<tr>
<td>(‘like being’, ‘you are’, ‘not’) a robot</td>
<td>533</td>
</tr>
<tr>
<td>hello robot</td>
<td>24</td>
</tr>
<tr>
<td>hi robot</td>
<td>18</td>
</tr>
<tr>
<td>robot friends</td>
<td>12</td>
</tr>
<tr>
<td>stupid robot</td>
<td>12</td>
</tr>
<tr>
<td>real robot</td>
<td>7</td>
</tr>
<tr>
<td>goodbye robot</td>
<td>6</td>
</tr>
<tr>
<td>mr robot</td>
<td>5</td>
</tr>
<tr>
<td>robot arm</td>
<td>5</td>
</tr>
<tr>
<td>cool robot</td>
<td>4</td>
</tr>
<tr>
<td>sexy robot</td>
<td>4</td>
</tr>
<tr>
<td>speak robot</td>
<td>3</td>
</tr>
</tbody>
</table>

**Coverage**: 71%

**Table 7-15 - Word ‘robot’ phrase query list**
The word robot’ was used in lots of different ways in the data but the most common instance was ‘do you like being a robot’, this was again the human User checking on the robot preferences. Further discussion regarding this preference checking is present in Observation 1 – The anthropomorphic stance & in Theme 1: Anthropomorphism.

<table>
<thead>
<tr>
<th>Harry</th>
<th>871</th>
</tr>
</thead>
<tbody>
<tr>
<td>Harry Potter</td>
<td>675</td>
</tr>
<tr>
<td>Coverage</td>
<td>77.50%</td>
</tr>
</tbody>
</table>

Table 7-16 - Word ‘Harry’ phrase query list

The high occurrence of the name ‘Harry’ being followed by ‘Potter’ was expected because the museum had a Harry Potter exhibition running for part of the Articulated Head’s time there. This table clearly helps to illustrate that the environment within which the exhibit and human User are situated, can influence conversational foci and be reflected in conversational references. This is discussed more under Theme 6: Dialogue.

<table>
<thead>
<tr>
<th>Live</th>
<th>834</th>
</tr>
</thead>
<tbody>
<tr>
<td>(where 407) do you live</td>
<td>483</td>
</tr>
<tr>
<td>live in (country)</td>
<td>164</td>
</tr>
<tr>
<td>Coverage</td>
<td>78%</td>
</tr>
</tbody>
</table>

Table 7-17 - Word ‘live’ phrase query list

The User questioning the robot about where it lived is interesting because it was essentially a fixed, situated machine with no legs, wheels or means of movement from the base. Furthermore it was enclosed in a metal-framed glass surround, which would suggest that it lived where it was. Therefore the word ‘live’ adds some weight to the anthropomorphic stance detailed in Observation 1 – The anthropomorphic stance and in Theme 1: Anthropomorphism.
The word ‘eat’ was used frequently to test robot preferences with the phrases ‘what do you eat’ and ‘what do you like to eat’ easily making up the 70% coverage threshold. Again, this is discussed further in Observation 1 – The anthropomorphic stance & in Theme 1: Anthropomorphism.

The phrase ‘are you happy’ once again is checking the robots status using an anthropomorphic stance as discussed in Observation 1 – The anthropomorphic stance & in Theme 1: Anthropomorphism. The other significant phrase ‘happy birthday’ is linked with the control word ‘sing’ and is clearly indicative of Users trying to get the robot to sing happy birthday.

‘What time is it’ as a phrase could be the User testing the robots knowledge but equally, it could be the User wanting to know the time and expecting the robot to recite it. Either way, the presumption of the User that the robot
knows the time is implicit in the input. User expectations are discussed more under Theme 2: Expectations and enculturation.

<table>
<thead>
<tr>
<th>Speak</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>can you speak (languages)</td>
<td>221</td>
</tr>
<tr>
<td>do you speak (languages)</td>
<td>227</td>
</tr>
<tr>
<td>speak to</td>
<td>12</td>
</tr>
<tr>
<td>speak louder</td>
<td>4</td>
</tr>
<tr>
<td>only speak</td>
<td>7</td>
</tr>
<tr>
<td><strong>Coverage</strong></td>
<td><strong>71%</strong></td>
</tr>
</tbody>
</table>

Table 7-21 - Word ‘speak’ phrase query list

Users were very interested in whether the robot could speak other languages, Spanish, German, French, Italian, Japanese and a long list of other languages were cited. Clearly this unmistakable wish of Users presents one very obvious way in which the Chatbot’s capabilities could be extended to improve engagement in interaction.

<table>
<thead>
<tr>
<th>Think</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>(what) do you think (about 46)</td>
<td>370</td>
</tr>
<tr>
<td>(misc)</td>
<td></td>
</tr>
<tr>
<td>I think you (are 26)</td>
<td>50</td>
</tr>
<tr>
<td>you think I am</td>
<td>36</td>
</tr>
<tr>
<td>think so</td>
<td>23</td>
</tr>
<tr>
<td><strong>Coverage</strong></td>
<td><strong>72%</strong></td>
</tr>
</tbody>
</table>

Table 7-22 - Word ‘think’ phrase query list

The phrase ‘do you think’ was frequently prepended by the word ‘what’ and had many diverse appendages. The appendages predominantly appeared to be related to specific interests of the User. The diversity of subjects did not contribute to the identification of any particular pattern.
7.1.2 Chatbot Output String Lists

<table>
<thead>
<tr>
<th>Name</th>
<th>Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>(my) name is Stelarc</td>
<td>79%</td>
</tr>
<tr>
<td>(what is 2838) your name</td>
<td></td>
</tr>
</tbody>
</table>

Table 7-23 - Word ‘name’ phrase query list

The word ‘name’ appearing as the number one high frequency word in Chatbot output strings marries with its appearance at position number one in the User input strings. The interesting point here is that the robot again presents some indicative evidence that it was more interested in itself than its audience, because the prepended word ‘my’ comes in at a higher frequency than ‘your’.

<table>
<thead>
<tr>
<th>Stelarc</th>
<th>Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>(my) name is Stelarc</td>
<td>75%</td>
</tr>
</tbody>
</table>

Table 7-24 - Word ‘Stelarc’ phrase query list

The phrase ‘my name is Stelarc’ was used by the Chatbot regularly. This broadly, although not exactly, aligns with the frequency of the question ‘what is your name’ used in User input. The table data does not contribute very strongly to any identified pattern or theme.

<table>
<thead>
<tr>
<th>Think</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>you think it is</td>
<td>1042</td>
</tr>
<tr>
<td>I dont think</td>
<td>596</td>
</tr>
<tr>
<td>let me think</td>
<td>438</td>
</tr>
<tr>
<td>you think artificial</td>
<td>386</td>
</tr>
<tr>
<td>(intelligence ...)</td>
<td></td>
</tr>
<tr>
<td>I think you mean its</td>
<td>355</td>
</tr>
<tr>
<td>I like to think</td>
<td>352</td>
</tr>
<tr>
<td>(Do you think) I</td>
<td>288</td>
</tr>
</tbody>
</table>
like to think in (black …) 238
I can think 183
think very many people 174
I think it was 139
I think so 122
think you already know 121
couldn't think 95
you don’t think so 34
did you think about 14

Coverage 71%

Table 7-25 - Word ‘think’ phrase query list

The word think is important in the respect that it is synonymous with the title of The Thinking Head Project. The Articulated Head was marketed as a thinking head. The list of contributory phrases required to get the coverage figure above 70% is quite long because the word ‘think’ was used in a number of differing but overlapping ways in the text data. Since the robot was projected as a thinking head, the number of ways that the Chatbot introduced the word ‘think’ into the conversational flow was probably encouraging Users to contemplate the possibility that the robot was thinking. If so, this was good thing. There is a healthy balance between the suggested robot’s thinking and recognition of User thinking in the Chatbot’s use of the word.

<table>
<thead>
<tr>
<th>Favourite</th>
<th>4694</th>
</tr>
</thead>
<tbody>
<tr>
<td>favorite food</td>
<td>1112</td>
</tr>
<tr>
<td>favorite past times</td>
<td>951</td>
</tr>
<tr>
<td>favorite movie</td>
<td>481</td>
</tr>
<tr>
<td>favorite Science Fiction (author)</td>
<td>415</td>
</tr>
<tr>
<td>favorite color</td>
<td>265</td>
</tr>
<tr>
<td>favorite song</td>
<td>165</td>
</tr>
</tbody>
</table>

Coverage 72%

Table 7-26 - Word ‘favourite’ phrase query list

The word ‘favorite’ in Chatbot output is directly linked to User input and the checking of the robots preferences, which is discussed in more detail under
Sub Section 7.1.3.1 Observation 1 – The anthropomorphic stance, and in 7.2.4 Theme 1: Anthropomorphism.

### Table 7-27 - Word ‘think’ phrase query list

The phrase ‘were talking about’ frequently has ‘we’ as a prefix. The phrase is interesting because it suggests that the robot possessed memory. There is evidence of this in the text data because the robot recalls a previously entered User string. There appears to be some randomness about which string it chooses to recall as the subject that they were previously talking about because sometimes the string makes sense in context and sometimes it does not. This is discussed in more detail in Theme 10: Memory Related.

<table>
<thead>
<tr>
<th>Talking</th>
<th>4566</th>
</tr>
</thead>
<tbody>
<tr>
<td>were talking about</td>
<td>1350</td>
</tr>
<tr>
<td>still talking about</td>
<td>772</td>
</tr>
<tr>
<td>were we talking about</td>
<td>399</td>
</tr>
<tr>
<td>Ill stop talking now</td>
<td>393</td>
</tr>
<tr>
<td>you like talking to me</td>
<td>385</td>
</tr>
<tr>
<td>I am talking about</td>
<td>367</td>
</tr>
<tr>
<td><strong>Coverage</strong></td>
<td><strong>80%</strong></td>
</tr>
</tbody>
</table>

### Table 7-28 - Word ‘people’ phrase query list

The Chatbot used the word ‘people’ in a number of different ways, which suggested to the User that the robot chatted with lots of people and was
aware of what other people thought and said. There was the inference that
the robot could hear people and that the robot could compare previously
expressed opinions with those being expressed by the current User. The
phrase ‘mind if I tell other people’ is directly linked with a research
candidate exchange, which included a secret between the User and the
robot.

<table>
<thead>
<tr>
<th>Time</th>
<th>4125</th>
</tr>
</thead>
<tbody>
<tr>
<td>asked all the time</td>
<td>1426</td>
</tr>
<tr>
<td>what time is it</td>
<td>394</td>
</tr>
<tr>
<td>spare time</td>
<td>388</td>
</tr>
<tr>
<td>percent of the time</td>
<td>355</td>
</tr>
<tr>
<td>point in time</td>
<td>340</td>
</tr>
<tr>
<td>killing time</td>
<td>136</td>
</tr>
<tr>
<td>Time has no</td>
<td>84</td>
</tr>
</tbody>
</table>

**Table 7-29 - Word ‘time’ phrase query list**

The Chatbot high frequency use of the phrase ‘asked all the time’ is
clearly reflected in the high frequency of the phrases ‘what’s your
name’ and ‘how old are you’ in User Input. These phrases might have made
the robot appear more human to Users. The ‘spare time’ phrase is normally
prefixed with ‘what do you do in’, this phrase does encourage User input
and also indicates that the robot was interested in the User. Attention is a
subject of interest under **Theme 7: Engagement**.

<table>
<thead>
<tr>
<th>Hear</th>
<th>3885</th>
</tr>
</thead>
<tbody>
<tr>
<td>want to hear a (joke 704 + misc)</td>
<td>1375</td>
</tr>
<tr>
<td>(question) I dont hear everyday</td>
<td>782</td>
</tr>
<tr>
<td>only hear that type (of response)</td>
<td>355</td>
</tr>
<tr>
<td>hear about Stelarc</td>
<td>334</td>
</tr>
<tr>
<td>glad to hear that</td>
<td>144</td>
</tr>
<tr>
<td>glad to hear it</td>
<td>103</td>
</tr>
</tbody>
</table>

**Table 7-30 - Word ‘hear’ phrase query list**
The high frequency of the phrase ‘want to hear a joke’ was in fact linked to a recurring crash loop that the Chatbot got stuck in from time to time. The log files only captured the first occurrence of this Chatbot output string in the loop subsequent repeats were not passed to the log file system by the event manager. The Chatbot repeated this phrase many thousands of times more than is indicated in the table above. This crash loop was a persistent problem that the engineers never really managed to solve completely. Why it happened with this particular phrase is a mystery. There was no filtering of text string data in the log file system – this was the only anomaly when recording text string data. The key point with reference to the Chatbot’s use of the word ‘hear’ in the other five instances in the table above below the phrase ‘want to hear a joke’, is that they infer to the User that that the Articulated Head can hear, which links with Theme 9: Senses – Related and Theme 10: Memory Related. The phrases ‘question I don’t hear everyday’ and ‘only hear that type’ both infer memory related to previous conversations and are also linked to Theme 10: Memory Related.

<table>
<thead>
<tr>
<th>Talk</th>
<th>3775</th>
</tr>
</thead>
<tbody>
<tr>
<td>lets talk (about (my dress) etc.)</td>
<td>1295</td>
</tr>
<tr>
<td>wanting to talk to</td>
<td>892</td>
</tr>
<tr>
<td>(I like the) way you talk</td>
<td>390</td>
</tr>
<tr>
<td>(I dont want to) talk about that now</td>
<td>253</td>
</tr>
<tr>
<td>Coverage</td>
<td>75%</td>
</tr>
</tbody>
</table>

Table 7-31 - Word ‘talk’ phrase query list

The Chatbot said ‘lets talk about’, and then prepended many different things. Sometimes ‘my dress’ was prepended and although I think Stelarc meant this to be discussion about clothes, observation of the public laughing and talking about the phrase indicated that they thought it would be very funny to see the robot in a dress. Generally the Chatbot’s use of the word ‘talk’ was positive towards the User and encouraging conversation. However the phrase ‘I don’t want to talk about that now’ was mentioned by several people that interacted with the exhibit in the museum and came across to at least one participant interacting with the Articulated Head like an angry
parent or other person making a terse, aloof comment that could be perceived as arrogant or rude. Rudeness appears under Theme 6: Dialogue.

<table>
<thead>
<tr>
<th>Understand</th>
<th>3330</th>
</tr>
</thead>
<tbody>
<tr>
<td>do not understand</td>
<td>1732</td>
</tr>
<tr>
<td>I understand</td>
<td>1115</td>
</tr>
<tr>
<td>Coverage</td>
<td>86%</td>
</tr>
</tbody>
</table>

**Table 7-32 - Word ‘understand’ phrase query list**

The Chatbot said it does not understand 67% of the time; this is too simplistic but in many ways it was also true as substantiated via Nodal data in Theme 6: Dialogue.

<table>
<thead>
<tr>
<th>Western</th>
<th>3226</th>
</tr>
</thead>
<tbody>
<tr>
<td>Country and Western</td>
<td>1810</td>
</tr>
<tr>
<td>Western Sydney(s)</td>
<td>1363</td>
</tr>
<tr>
<td>Western Bulldogs</td>
<td>35</td>
</tr>
<tr>
<td>Coverage</td>
<td>99%</td>
</tr>
</tbody>
</table>

**Table 7-33 - Word ‘western’ phrase query list**

The Chatbot, having been programmed by some American programmers in consultation with Stelarc, who provided most of the preprogrammed responses to User input, clearly reflects the fact that Stelarc enjoys country and western music. This is a clear example of where Stelarc’s preferences showed in the Articulated Head’s performance and it is also a clear example of where Stelarc’s contribution to the programming of the Chatbot responses gave the Articulated Head agency. More examples of where Stelarc’s contribution to the programming presents Stelarc’s personality through in the performance of the Articulated Head are present in Theme 6: Dialogue.
Table 7-34 - Word ‘human’ phrase query list

‘Human years’ and ‘old’ are clearly linked in Chatbot strings because the Chatbot said ‘I am 62 of your human years’ frequently in response to the User questioning its age. This Chatbot phrase does infer to any User hearing it that the Articulated Head is not impersonating a human, this however does not preclude the inference that the Articulated Head was capable of human abilities such as seeing hearing and thinking. Many of the uses of the word ‘human’ in the phrases above are a little impersonal and clearly set the robot apart from humans in that respect. Possibly a more personal approach might increase engagement.
**Table 7-35 - Word ‘old’ phrase query list**

Probably the two most consistently common phrases to appear in both User input and Chatbot output strings are ‘what’s your name’ and how old are you’; more discussion of this takes place in **Theme 6: Dialogue**.

<table>
<thead>
<tr>
<th>Old</th>
<th>3035</th>
</tr>
</thead>
<tbody>
<tr>
<td>how old are you</td>
<td>1225</td>
</tr>
<tr>
<td>62 years old</td>
<td>716</td>
</tr>
<tr>
<td>(many) years old my guess</td>
<td>157</td>
</tr>
<tr>
<td>how many years old</td>
<td>87</td>
</tr>
<tr>
<td>Coverage</td>
<td>72%</td>
</tr>
</tbody>
</table>

**Table 7-36 - Word ‘color’ phrase query list**

The word ‘color’ and its spelling are linked to the word ‘favourite’ and ‘favorite’ and discussed under **Observation 2 – Programming additions** that follows Observation 1 below.

<table>
<thead>
<tr>
<th>Color</th>
<th>1120</th>
</tr>
</thead>
<tbody>
<tr>
<td>the color of (misc)</td>
<td>418</td>
</tr>
<tr>
<td>What color are your eyes</td>
<td>361</td>
</tr>
<tr>
<td>favorite color is I like to think</td>
<td>234</td>
</tr>
<tr>
<td>Coverage</td>
<td>90%</td>
</tr>
</tbody>
</table>

**7.1.3 Observations made from review of the phrase query lists.**

A number of observations, both qualitative and quantitative in nature were made as a result of the word search queries and review of the phrase search table data, which does reveal something of a pattern emerging in the conversational foci.

**7.1.3.1 Observation 1 – The anthropomorphic stance**

There is sufficient evidence in the word frequency and phrase table searches alone to suggest that users interacting with the Articulated Head, appeared predisposed to question the Articulated Head about its preferences, with phrases such as; ‘what is your favorite colour, food, game, sport, song’ E-Appendix 19: Example 2, or ‘do you like food, cats, sport, movies?’, ‘do you like cheese’, ‘do you like chocolate’ E-Appendix 20: Example 3 proliferating
in the text input/output data. Further substantiating evidence of this predisposition proliferates throughout the Nodal data that is presented with the themes that follow shortly.

The words ‘favorite’, ‘like’ and ‘love’ are all linked by a correlation between the subjects checked by User questioning of the robot. Implicit, in the human users questioning of the preferences of the Articulated Head is what I have termed the anthropomorphic stance. That is; a high occurrence of the questions a human User asked of the Articulated Head, at the very least test and probe the robot by making comparisons based on the human’s existential experience of their lifeworld, but furthermore, the questions regularly appear to have the implicit attribution of human capabilities afforded the robot. If one has a favourite colour or reads books and watches movies, then one can see. If one has a favourite song then presumably one has heard this song - so it stands to reason that one can hear. If one likes cheese or chocolate, then presumably one can taste, and probably smell as well. It was noted that there was much less evidence in the text data related to touch or feeling in particular, this observation will be discussed under Theme 9: Senses – Related. However emotional feelings do feature in the text and Video Cued Recall data, this will be discussed under Theme 8: Emotions.

The anthropomorphic stance taken by humans in approach to interaction with the Articulated Head is unmistakable and present in data across all conditions, both inside and outside the research participant data frame. Human User adoption of the anthropomorphic stance appears to be instinctive and is a pervasive, if not a universal approach to this particular interaction as supported by the research data collected in both video cued recall interviews and the text string analysis. There are comments in participant reports to suggest that the human-like face of the Articulated Head was the catalyst that engendered adoption of the anthropomorphic stance, but there is not enough evidence in the data to confirm that this was in fact the case. The suggestion that the human-like face was the catalyst is probably only part of the story. The human-like face being a contributory factor sounds completely plausible but other elements of the presentation of
the Articulated Head such as movement, speech and marketing must also have played their part.

The human User’s comparison of the robot’s preferences to their own, based on their own existential experience and lifeworld, and the correlation that this comparison has with the senses, is both obvious and remarkable given that the human must have known that the robot was not a biological existential being and was not endowed with the ability to see, hear, smell, taste, or touch/feel. Perhaps the anthropomorphic stance was adopted precisely for this reason, just to confirm their suspicions that the robot was not a real-live being. The anthropomorphic stance taken, draws some poignant parallels with the phenomenological reduction in Section 3.1.4.15 Sense by sense comparison, and is particularly interesting because, if one can predict with any significant degree of certainty, what subjects are the most important to the human in conversational interaction, then clearly the prediction provides some leverage for the business of a targeted approach to expansion of Chatbot phrase vocabulary, in order to enhance human engagement in this interaction. This opportunity is discussed in Section 9: The Blueprint of Emergent Recommendations.

7.1.3.2 Observation 2 – Programming additions

The words ‘favourite’ and ‘colour’ did stand out as very prevalent words with a range of surrounding topics in interaction strings E-Appendix 21: Example 4, so a little further investigation was called for. The fact that both these chosen top 100 search words from User input and Chatbot output strings can be spelt with English and/or American spellings becomes a point of interest in what follows. The word ‘colour’ appears in the User input strings top 100 words, but not the Chatbot output strings top 100 words. – Upon a search of the E-Appendix 22: Chatbot’s top 1000 words ‘color’, with the American spelling does appear at position 340, so I decided to add it to the E-Appendix 17: Chatbot Output String Phrase Table (with orange highlighting).

It was initially thought that perhaps the Chatbot might only use the American spelling. If so, then the word spelt with a ‘u’ in it, would indicate User input in every case within the data set. However upon further investigation I found the following; User input shows 130 instances of ‘color’ and 1075 instances of
‘colour’, indicating a very strong tendency towards the English spelling of the word in the demographic of the visiting public.

This is not surprising as the data was collected in Australia, which is English speaking and has very strong current and historical links with the United Kingdom. Chatbot output shows 687 instances of ‘color’ and 433 of ‘colour’, this was surprising as it was expected the American spelling would be the only spelling used because the Chatbot was programmed in America by American programmers in consultation with Stelarc. So queries were set up to look for only the English, and only the American spelling. Upon viewing all the Chatbot text references for ‘color’ the Chatbot phrases made sense as stand alone statements. However, upon viewing all the Chatbot text references for ‘colour’ I discovered that most instances of its use were where the robot utterances were nonsensical -‘the colour of infinity’, the colour of motion’, ‘the colour of transparency’, ‘the colour of knowledge’ E-Appendix 23: Evidence 5. There was also a couple of other high frequency uses, which made a little more sense as standalone statements, such as; ‘the colour of the pacific’ and ‘the colour of the train behind me’. The last of these instances very strongly suggests that someone, possibly an engineer, carried out this programming and expansion of the Chatbot’s phrase vocabulary after the Articulated Head was situated in the Powerhouse Museum. This observation is made on the basis that there was a train behind the Articulated Head in the Powerhouse Museum, and it is unlikely that this was the case anywhere else. It certainly was not the case anywhere else during the course of this investigation. This anomaly only affected a minute fraction of the overall data collected so this was not considered a major issue.

Whoever chose to add these nonsensical additions to the Chatbot’s output strings, had a preference for the English rather than American spelling of the word ‘colour’. Clearly entering nonsensical preprogrammed responses to User input, no matter what the triggers are, is hardly going to increase the Users engagement in the conversational interaction, and although there are instances of the User questioning the robot about these statements, more
often than not, the Users next input to these statements from the Chatbot was to move away from the current topic and try something new.

On the basis of the observations imparted above, the phrase counts listed in the Phrase Tables with specific reference to Table 7-9 - Word ‘favourite’ phrase query list and Table 7-10 - Word ‘colour’ phrase query list, where it is clear that the user is normally raising the questions rather than responding to Chatbot output, and taking into account the imbalance between User input and Chatbot output word counts, it can be said that the high prevalence of the word ‘colour’ as a conversational topic of focus in human input, was primarily human initiated rather than Chatbot driven. This is significant, as some frequently repeated pre-programmed strings, could otherwise have been driving the prevalence of this topic, which they were not. It can also be said that there appears to be a division in the ‘giver of agency’ to the Chatbot, where the switch from the American to English spelling indicates strongly that the Chatbot’s programming had at least two separate human contributors at the source of data input to the Chatbot’s phrase vocabulary. It is clear, is that the humans present on all sides of this information exchange felt the word colour was important. I too believe that the word colour is important, because it relates directly to the sense of seeing, which was one of the most prevalent probes of the human User in questioning the robot. Since the word ‘favourite’ also has the same English/American spelling anomaly as the word ‘colour’, I checked the data to see if there was a mixture of English and American spelling instances of this word in the Chatbot output text strings, there was not. The American spelling ‘favorite’ was used exclusively, lending more weight to the observation of that the word ‘colour’ was utilised for extending the Chatbot’s phrase vocabulary.

7.1.3.3 Observation 3 – The Input/output word ratio imbalance

The 1.81:1 Chatbot Output/User Input word ratio imbalance identified earlier under the top 10 word search expanded to nearly 2.54:1 in the top 100 word searches: User – 374,015 words, Chatbot – 948,230 words over the data set. This expansion of ratio indicated a possible growth curve, so a word frequency search for the E-Appendix 24: Top 1000 User Input String Words and E-Appendix 22: Chatbot’s top 1000 words was conducted for the
purposes of checking this ratio expansion over a broader span of the data. This also served to make sure that there was not any other obvious pattern emerging from the list (other than the observations already made) that could readily be substantiated with clear evidence. The ratio of Chatbot Output/User input words, when totaled from the Top 1,000 table linked above is 2.92:1, Chatbot – 1,436,945 words, User 492,000 words. No other specific patterns were identified. This clearly indicates that the rate of ratio expansion slows with the increased number of words.

Increasing the Top 1000 to the E-Appendix 25: Top 10,000 User Input String Words and E-Appendix 26: Top 10,000 Chatbot Output String Words did identify that a very large proportion of occurrences of input strings and output strings did not make sense as proper words. The further down the list of the Top 10,000 input/output strings you go, the more examples you can find of erroneous input/output. It is not clear how many of the questionable spellings and sewn together words of the Chatbot output strings had been specifically constructed because they sounded better that way. It is also not clear whether Users interacting with the robot understood these unusually constructed strings. However, what is clear is that the deeper you dig, the more noise there is in the data, making it harder and harder to extrapolate any useful information or make sustainable observations.

7.1.3.4 Observation 4 – Conversational diversity across the data set
Humans appeared to display less conversational diversity or vocabulary bandwidth than the Articulated Head based on the length of phrase lists presented earlier. The Top 10,000 lists also indicate a wider vocabulary being used by the Chatbot output as oppose to the human User input. However, these observations are somewhat misleading because of aspects discussed under Theme 4: Mode of Interaction.

7.1.3.5 Conclusions from the text string analysis
The string analysis has dealt with high frequency phrases and treated input and output strings as separate entities to a large extent. The nodal analysis that follows looks at the sense of dialogue exchange between the User and the Chatbot under Theme 6: Dialogue. The most significant observations
extrapolated from the textual input/output data are that; the anthropomorphic stance to conversational interaction is clearly evident; that the Chatbot says considerably more words than the User inputs; and that there is significant noise present on both sides of the input/output data flow, but User input is especially noisy.

**7.2 The NVIVO Coding Process and Emergent Themes**

A coding process was conducted in the computer assisted qualitative data analysis software called NVIVO. The coding process was used to sort out themes in research data collected. Before describing the coding process and how it was refined, it is important to note here the nature and presentation of data and evidence that follows in this section and how the presented evidence should be taken in context. You will be presented with a description of how the coding was done and how the main themes were sorted from the coding.

It is important to note that the textual data you view will in many cases display interspersion of qualitative researcher commentary and interpretation of the content of the interview, along with the Participants interview dialogue. The text may also show User Input and Chatbot output text from the interaction dialogue. This interspersion of data is appropriate because it captures the content, context, possible meanings and interpretations in the same place together, at source. Whilst attempts have been made to maintain spelling in the qualitative commentary and interview dialogue the NVIVO database software has no spell checker incorporated. Every attempt has been made to retain the meaning and context of participants interview commentary. The following bracketed indicators (………..) or (????), have been used to signify where something was not clear, and (XXXX) where someone’s name or identity has been hidden for ethical reasons. No attempt whatsoever has been made to change or correct spelling or typos in the interaction dialogue, because it is what it is, and it should be retained that way for the purposes of analysis. Therefore the evidence you view, which contains text from interaction input/output in particular, will be littered with typo’s and spelling mistakes. However, this in
itself is indicative of various points drawn out in the interpretive narrative that emerges from the data analysis.

It is also important to note that the NVIVO database and audiovisual content runs into several hundred gigabytes of data. To keep this thesis to a size practical for delivery and assessment on DVD, only specific auditory visual content is presented as substantiating evidence. This content has been exported from specific nodes in the database as interactive web pages, where only the relevant coded sections of video content are displayed at a smaller size (320 by 240) than the original capture size of (1024 by 768). Audio quality varies in the content considerably dependent on the ambient noise and acoustic environmental conditions under which the capture was conducted. Therefore good amplification and listening conditions are recommended when wishing to review the finer details of the audio content.

The content chosen for presentation as interactive web pages have been specifically selected as substantiating evidence to support major points emergent from the data, where it was felt that the textual evidence supplied from nodes alone may not be sufficient to substantiate the validity of the interpretive narrative, or where an emergent and pivotal observation from the data has profound ramifications on the findings of this investigation.

The following table outlines the interaction time stamps of Video Cued Recall interviews including the creative auditory visual project conditions.

<table>
<thead>
<tr>
<th>Name</th>
<th>Date</th>
<th>Time</th>
<th>Tot Duration</th>
<th>Notes</th>
<th>AV Conditions</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>16/6/10 16/6/10</td>
<td>13:47:40</td>
<td>9 mins 34 sec</td>
<td>No Audio JV</td>
<td></td>
</tr>
<tr>
<td>A2</td>
<td>16/6/10 16/6/10</td>
<td>13:57:14</td>
<td>No Audio JV</td>
<td>No projections</td>
<td></td>
</tr>
<tr>
<td>K1</td>
<td>18/6/10 18/6/10</td>
<td>11:00:07</td>
<td>15 mins 45 sec</td>
<td>No Audio JV</td>
<td></td>
</tr>
<tr>
<td>K 2</td>
<td>18/6/10 18/6/10</td>
<td>11:15:52</td>
<td>No Audio JV</td>
<td>No projections</td>
<td></td>
</tr>
<tr>
<td>O</td>
<td>16/6/10 16:19:00</td>
<td>16/6/10 16:25:40</td>
<td>6 min 34 sec</td>
<td>No Audio JV</td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>15/10/10 16:24:53</td>
<td>16:28:56</td>
<td>4 mins approx lost video - Interview</td>
<td>No Audio JV</td>
<td>DPF</td>
</tr>
<tr>
<td>C Interview</td>
<td>15/10/10 16:10:29</td>
<td>16:25:40</td>
<td>30 sec video cap + lost second half 10 min Approx</td>
<td>No Audio JV</td>
<td>DPF</td>
</tr>
<tr>
<td>D</td>
<td>15/10/10 14:30:02</td>
<td>14:32:23</td>
<td>2 mins 21 sec - Some lost</td>
<td>No audio JV DPF</td>
<td></td>
</tr>
<tr>
<td>K</td>
<td>15/10/10 16:00:19</td>
<td>6 mins 54 sec</td>
<td>No Audio JV</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Sub Section 7.2 – The NVIVO Coding Process and Emergent Themes

15/10/10 16:07:13 DPF
T 16/10/10 09:30:04 5 mins 19 sec No audio JV
16/10/10 09:35:23 DPF
T Complete video loss N/A
L Audio loss N/A
A1 4/8/11 12:19:53 5 mins 50 sec Audio SPAT VO VM +
A4 4/8/11 12:25:43 CP
B1 4/8/11 12:38:58 16 mins 57 sec Audio SPAT VO VM +
BB1 12/8/11 14:02:21 Audio SPAT VM +
BB4 12/8/11 14:13:17 Sw DP CP
F1 13/3/11 10:01:59 3 mins 41 sec Audio SPAT VO
F1 13/3/11 10:04:40 DP
GG1 12/8/11 13:27:24 9 mins 36 sec Audio SPAT VM +
GG3 12/8/11 13:37:00 CP
L1 5/7/11 14:52:50 4 mins 48 sec Audio SPAT VO +
L3 5/7/11 14:57:38 Sw DP CP
P1 4/8/11 15:51:57 21 mins 41 sec Audio SPAT Sw VO VM + Sw
P9 4/8/11 16:13:38 DP CP
S1 2/9/11 13:11:09 5 mins 38 sec Audio SPAT Sw VO VM + DP
S3 2/9/11 13:16:47
S10 19/10/11 13:51:49 13 mins 11 sec Audio SPAT VM
S15 19/10/11 14:05:00 DP CP
ST1 26/10/11 14:15:03 5 mins 59 sec Audio SPAT Sw VO VM + CP
ST3 26/10/11 14:21:02
Vint1 4/5/11 11:30:46 16 mins 46 sec Audio SPAT VO
Vint9 4/5/11 11:47:32 DP CP (early)
VR21 15/6/11 11:00:56 46 mins 36 sec Audio SPAT Sw VO + DP CP
VR214 15/6/11 11:47:32

Other evidence collection conducted during this investigation

<table>
<thead>
<tr>
<th>Textual Data</th>
<th>Video Interview</th>
</tr>
</thead>
<tbody>
<tr>
<td>June 2011</td>
<td>Sept 2012</td>
</tr>
<tr>
<td>Nov 2012</td>
<td>10 mins each</td>
</tr>
<tr>
<td>24 hours a day</td>
<td>17 full months of interactions</td>
</tr>
<tr>
<td>Data collected under any of the conditions below</td>
<td>N/A</td>
</tr>
</tbody>
</table>

Table 7-37 Video Cued Recall InteractionCaptures

Auditory Visual Conditions Key
Sw = Switched during interaction
Mx = Mixed during interaction
SPAT = spatial audio (spatial auditory composition).
SPAT_F = spatial audio (spatial auditory composition + affected synth parameters).
JV = Just voice (Stereo - stationary).
VO = Omnipresent voice.
VM = Voice movement (Robot voice controlled by co-ordinates).
DPF = Director projections - front.
DP = Director projections.
CP = Cloud Projections
JP = Jamoma Projections
Summary of data collection

In summary of the above table, there is 3 Hours 46 minutes of clear captured and coded Video Cued Recall Interview data, with 23 contributing participant reports related to the Articulated Head Exhibit in total.

In mid/late 2011 a very large construction and refurbishment project commenced in the Powerhouse Museum and collection of interactions and Video Cued Recall interviews became impractical as the project commenced. Environmental sound conditions, pneumatic drills and hammering made audio data unusable. The Articulated Head was shut down and covered by sheets on several occasions due to problems associated with construction dust. Furthermore problems with torrential rain and a leaking roof in the museum also resulted in numerous shutdowns of the system. From October 2011 onwards data collection slowed dramatically and the coding process was started. The construction project-taking place in the museum was an enormous undertaking and was still progressing in November 2012 when the Articulated Head was decommissioned.

The changes in auditory visual conditions detailed in table 7-38 above, did not have any dramatic influences over the interactions that took place for the reasons detailed under 7.2.3 The Narrative Emanating from the Coding. Discussion of the mediating considerations and constraints, which help to explicate confounding and compromising aspects of the experimental auditory visual designs and their presentation to the interacting audience is delivered under Theme 12: Auditory Visual.

There is a relative uniformity of participant reporting contributions to each of the emergent themes of this investigation, as well as consistency of participant spread across node coding that is relative to the duration of their interaction timestamps. Participants that have contributed with longer interaction timestamps, appear proportionately more often in the nodal data presented.
Most of the conditions changed in the interactive environment over the data set were very minor, and are not considered to have been significant to participant interactions. The reasons for this view are detailed under section 7.2.3 The Narrative Emanating from the Coding.

The individual differences of the participants involved in these interactions across the data set are evident in review of auditory visual data and questionnaire outcomes presented in the Key Evidence Links Table, but are not really relevant to the major findings of this investigation, because the consistency and spread of participant reporting on the themes picked up in data analysis, is relatively uniform across the data set. In the nodal analysis that follows, the omnipresence of participant contribution to the emergent themes, speaks clearly of the fit of analysis and interpretation to the research participant data set. More specifically, this omnipresence points to a fundamental aspect of the interaction and interactive environment, as being pivotal to all subsequent aspects of the interaction-taking place. For the most part (physical and mental disabilities excepted), the major findings from this investigation do not hinge on individual differences, and do relate to the vast majority of all possible human Users.

7.2.1 What was done?

Video Cued Recall Interviews were imported into NVIVO for qualitative data analysis. Watching and listening through the interviews started the analysis process. Whilst watching and listening the video was annotated in short sections. Selecting what appeared to be natural divisions of inactivity, sometimes bracketed by silence, or natural divisions in conversational foci-taking place during the Video Cued Recall Interviews, helped to choose each section for coding. Breaking video content and annotations up into small sections, allowed for interspersion of qualitative commentary with interpretation of the content of the interview, laced with transcription of the participants interview dialogue, and User Input/Chatbot output text. The image below shows a section of video with annotations in table rows to the right hand side of the video image. Each table row corresponds to a short section of video as indicated in the Timespan column. Each consecutive row
details commentary and annotations related to the video content in chronological order.

This process of annotation included simple coarse coding to nodes in NVIVO. Coarse coding of nodes means that sections of video related to a table row would be linked to newly created nodes, which would be titled to broadly describe the content of that table row. For example the node might be called Memory if the interview dialogue and annotations were related to the Articulated Heads memory. This process of annotation and coarse node coding was conducted for all the auditory visual data collected from participant interviews resulting in approximately 30 initial nodes.

Once the annotation and coarse coding was completed, a second and third pass of all the data was made as the node coding process was refined and intensified to include creating new nodes for any newly emergent subjects of interest, whether they emerged from User input or Chatbot output text, Video Cued Recall interview dialogue or qualitative commentary and interpretation of the auditory visual evidence collected. Node creation was completed with little or no attention paid to whether a node being created

Figure 7-1 Video Cued Recall Annotations, dialogue Transcription and Commentary
would belong to a particular family of nodes or not, as the theme identification process was expected to perform this service later anyway. As a result of this, some nodes were created with very similar names and some nodes relate to very similar content. However, a few parent nodes were created with families of child nodes, where it was obvious that the children inherited a particular aspect of the parent node in their coded content.

The process of revisiting the data and annotations resulted in a total of 175 nodes of interest being created, and well over 2400 referenced sections of data to these nodes. Whilst conducting the node coding process three separate nodes were created named ‘positive, negative and neutral’ for the purpose of linking clear statements about interaction from participants as coding progressed. This was set up to provide a way of having a control check node. I.e., these nodes would provide a simple way to check if the positive, neutral and negative nodes broadly align with the emergent narrative from the data analysis.

Some nodes have overlapping themes and any auditory visual or textual data can be coded to more than one node at the same time. It is also possible that sections of data traverse nodes, which overlap but the content coded does not necessarily extend to both ends of the each of the nodes coded spans. The distinct advantages of the researcher conducting the Video Cued Recall Interview dialogue transcription, qualitative annotation and interpretation alongside the video data itself is that; 1) Visual interpretation to confirm the meaning of auditory data is not lost - as would be the case if normal audio transcription had been conducted by a third party transcription company without any visual cues. 2) The researcher is brought much closer to the data itself, leading to a deeper more immersive understanding and qualitative interpretation of meanings and messages emanating from the data.

**Coding saturation**

The node creation coding process was continued until the auditory visual and textual data stopped throwing out new ideas for node names so that when the data was revisited again, new sections of interest were consistently
being coded or recoded against existing nodes. At this point coding saturation had been reached. It is possible to revisit the data over and over again creating ever more deeply layered and finer sub node coding of data already coded to broader nodes, but this process is unnecessary where a satisfactory understanding of the content of the node can be gained from review of the node data in reference view, which was the case for most of the 175 nodes created in this project.

Some nodes have many more references coded to them than others, primarily because the references occur much more often in the data. This does not automatically mean that these nodes are more significant than those with fewer references linked. However the multiple occurrences of specific or similar events in the interactions or the Video Cued Recall Interviews, along with associated annotations in any one particular group of nodes in the coding tree, is significant in that it speaks loudly of these events, and is likely to indicate a theme.

There are however some much quieter but very significant voices to be heard, or messages to be observed from the data, some of which are all but silent in the node coding, but endemic in the auditory visual material collected. A specific example is that all research participant interaction videos show that an inordinate proportion of the participants time during interactions was being spent looking down at the kiosk keyboard, and then another significant but smaller proportion of interaction time looking directly at the face of the Articulated Head, waiting for a response. Review of video data presented for the emergent themes that follow shortly will substantiate this observation. This fact is not surprising at all because the exhibit’s interactive design effectively forced this to happen. The time and attention-space left for the human to take in any other environmental information was greatly constrained by this single fact. Rather like fitting a horse with blinkers to keep its gaze focused on what is directly ahead, the human beings present in these interactions were similarly, effectively blinkered.
The above point is highly significant, because it dictated the mode of engagement. Some might say that it kept the human engaged in a specific mode, which indeed it did to a large extent, and human engagement in this interaction, or improvement thereof, was a goal of this investigation. However, the overarching purpose of this research investigation is to consider how human ↔ machine interaction can be improved within the context of the interactions that have taken place between the Articulated Head and humans. With this question in mind, one must consider whether this mode of engagement was the right one for nurturing human ↔ machine interaction in this context? A major conclusion of this investigation is that it was not! This Section of the thesis and particularly Theme 4: Mode of Interaction presents themed evidence that confirms and substantiates the reasons for reaching this conclusion and therefore, what follows should adequately convey this.

**Theme sorting**

The next stage of the nodal analysis was to sort out the main emergent themes from the node data. This process is in effect the business of trying to work out what narrative the cumulative data analysis is actually purveying. The question is of course, where does one start?

The text input output analysis at the beginning of this Section highlighted the high occurrence of what I have termed ‘the anthropomorphic stance’ adopted by many, if not all humans in conversational interaction with the Articulated Head. Therefore anthropomorphism was taken to be a likely theme and nodes were reviewed by title first, and then by content to see if they could be relevant to this particular theme. It very quickly became clear that several nodes were relevant to this theme but it also became clear that, just as nodes have overlapping and intersected data references, so to would the overarching themes and narrative that explicates them. The problem with this situation is that you do not want to repeat yourself (or presentation of evidence) too often when imparting a narrative in text, for fear of losing the reader’s interest and attention.
So, a list of all the node names was printed onto paper and they were cut up so that each node name was present on a separate slip. During the paper cutting session the names were reviewed again and a list of possible titles to themes were written on separate large labels. The individual node name slips were then sorted into lists by placing them beside the theme label that they most likely related to, any which could not initially be placed under a theme title were left aside for the moment.

After the process of sorting the node title slips under the titled theme labels, I was delighted to find that all but about 20 of the 175 slips had been fairly reliably identified as relating to one of the specific theme titles over and above any of the others, even if some of the data references coded at that particular node do have significance in other themes, as has already been discussed.

A new theme title ‘Outliers’ was created for the 20 or so node slips that could not yet be placed under a theme title, and then the list of node titles were arranged under theme titles in a word document list. This word document then became the Interactive Evidence List table, which was then distilled to create the Key Evidence Links Table supplied on DVD.

The Key Evidence Links Table lists theme titles with links to some key textual and audiovisual substantiating evidence. The links table is provided as electronic documentation to the reader to act as a quick access panel. The Key Evidence Links Table can, and should be used as a standalone link document, whilst reading the narrative that follows.

The next step was to review the content of the nodes to check that it made sense to consider that evidence under that particular theme title and if not, then move that node to an appropriate title.

The content of nodes under the Outlier theme title was reviewed to see if those nodes could be reliably and helpfully placed under an appropriate
theme title. This sorting and reviewing process sounds simple, but in fact, it was very involved and took a protracted period of time.

Theme titles were sorted into an order that appeared to make some sense for imparting the narrative emanating from the data. Then an attempt was made to sort the node titles under each theme into a sensible order. The initial intention was to try and address each node’s data in list order in the narrative. However, upon beginning to write the narrative, it very quickly became clear that sorting the node order under theme titles was somewhat superfluous, because, in order to make proper sense of it all, the narrative was unlikely to address each node in the list order systematically. Furthermore the narrative would not necessarily present the themes systematically in list order either, because many nodes traverse several themes. However, for the readers benefit, an attempt has been made to be as systematic as possible when imparting the themed narrative.

### 7.2.2 Revisiting the Knot

The node based data presented through the Key Evidence Links Table, is effectively the untangled knotted issues and strands of the Knot mentioned in section 2.14.2. The coded nodes are the strands that were knotted; the themes that follow are the families they should have be platted with. The Outlier theme and associated nodes are the strands that appear to be loose ends. The interpretative narrative that follows proceeds by identifying barriers in terms of nurturing flow in human ↔ machine interaction in this interactive environment. The themes themselves represent the brushed groups of strands that are to be re-platted and The Blueprint presented in section 9 is the effective re-platting of these knotted issues and strands into a vision, which represents a group of outcomes and related ideas emergent from the investigation that show how interaction between humans and machines can be improved within the context of this, and similar interactive environments.

When reviewing nodal text data from the Key Evidence Links Table, poignant node references have been highlighted.
7.2.3 The Narrative Emanating from the Coding

Scope
The Articulated Head was exhibited in three separate places, the NIME, SEAM and the Powerhouse Museum exhibition as previously described. The node coding and interpretive narrative emanating from the coding spans data collected from all three exhibitions. Various environmental conditions changed over the duration of data collection including development and installation of the auditory visual additions as described previously. At one stage (only for a couple of days) during the Powerhouse museum exhibition, the face and voice of the Articulated Head was changed to a female. Notably the Chatbot programming remained the same. The Articulated Head’s performance capabilities were expanded during the Powerhouse Museum exhibition with the ability for it to sing a few songs. However the core business of the Articulated Head, talking to humans interacting with it – primarily via the kiosk keyboard, did not change over the period of data collection. Furthermore the presentation of the Articulated Head, the Face shown on a screen, mounted on the end effector of the industrial robotic arm with the exhibit presented in a metal-framed glass enclosure did not change to any great degree in any of the three exhibitions. All three exhibitions were held in noisy environmental conditions in public exhibition spaces. The tracking system and all main capabilities of the Articulated Head remained the same over the duration of data collection. There were some very small variations to the preprogrammed vocabulary of the Articulated Head’s A.L.I.C.E Chatbot as discussed previously under Observation 2 – Programming additions, but for the most part; all the variances discussed above had only minor reported impact upon the overall experience of Participants interactions.

Wherever any variance of the exhibit has affected the data collected, and therefore influenced the interpretive narrative that follows, its affect and any possible significance is discussed in the narrative. Observable impacts of the variations to the wider public interactions that took place are also discussed in the theme interpretive narrative where appropriate.
7.3 The Emergent Themes

7.3.1 Theme 1: Anthropomorphism

This theme is directly linked with sub section 9.2.1 Anthropomorphism.

As has already been identified and discussed in relation to the word ‘favourite’ under Table 7-26 - Word ‘favourite’ phrase query list and Observation 1 – The anthropomorphic stance; the anthropomorphic stance taken by humans approaching and interacting with the Articulated Head is strongly evident in the conversational text data logs collected. Examples of evidence show cordial politeness and manners being displayed by the interacting audience in conversation with the robot. Furthermore, this anthropomorphic stance and a human communications approach to interactions, is very strongly evident throughout the interactions captured in the Video Cued Recall data with research participants.

Indeed, the propensity for the probing and testing of the robots preferences and human-like capabilities, with the interacting human clearly making direct comparison of attributes based on their own existential experience, is ubiquitous throughout the entire data set. This anthropomorphic stance is clearly laced with the humans’ wants, hopes and beliefs when interacting with the Articulated Head. The empirical data set collected during this research, and more specifically the linked textual and auditory visual data presented as electronic appendices to this document via the Key Evidence Links Table, cumulatively, with numerous instances present across the data set, in all conditions, and in all three exhibitions, substantiates that humans do want, hope and in a number of cases, believe that the Articulated Head does in fact have human-like inclinations and capabilities such as the inclination to flirt for E-Appendix 27 Example 6, or the ability to E-Appendix 28: See, hear, think and even feel emotions or perceive the intentions of others.

The hopes and the wants feeding the beliefs of the humans displaying this anthropomorphic stance towards the Articulated Head, is considered in the conclusion of this investigation to provide pivotal leverage with regard to how one can improve human ↔ machine interaction. Put simply, catering
for the hopes and wants of the human in this interaction nurtures belief, even if that catered entity is in fact an untruth, a false statement.

Before proceeding with the themes, it should be noted here that various mitigating circumstances surrounding project development and data collection are cited under the following theme headings. One of the specific aims of this project was to conduct the research in ecologically valid environments, public exhibition spaces with a project team and all the trials and tribulations that this scenario might incur. Therefore circumstances are cited to give a clear overview of what actually happened as this research project progressed. A positive view has been taken in relation to overcoming all the challenges and barriers encountered, resulting in the projection of ideas, theories and recommendations for circumnavigation and negation of these challenges and barriers in future projects in public exhibition spaces in section 9.

7.3.2 Theme 2: Expectations and enculturation
This theme is directly linked with sub section 9.2.2 Expectations and enculturation.

The initial impressions and expectations of people when first approaching the Articulated Head were mixed. For example: one participant reported expecting the kiosk screen to be a touch screen. Many members of the public, and particularly children, frequently assumed that someone behind the scenes was controlling the robot. The Powerhouse Museum has a very large number of schools visits and I was regularly questioned by groups of school children coming around to the laboratory at the rear of the projection screen saying, “You’re controlling that robot!” - referring to either myself or sometimes others present in the lab area at the time. The response given (sometimes many times a day) was to move away from the computer keyboard and exclaim, “look no hands” – then wait for the robot to speak. The children were amazed, and then a little perplexed by the idea that the robot might actually have a mind of its own, and was able to speak without me typing. Once convinced that I was not controlling the robot, in general the children showed credulity, appearing quick to accept the possibility that
the robot might be alive, whereas adults approaching the robot were considerably more skeptical and testing, probing the robots capabilities. In general the research participant’s initial impressions and expectations of the robot were most strongly stated by the first few actions during their interactions or the first few statements made during their Video Cued Recall interviews. For example one participant had the visual expectation of seeing Stelarc’s face as the face of the robot, as all reporting suggested that this would be the case. The participant was surprised because the female face mentioned earlier was being displayed on the screen during her interaction. Her first comment in the interview was:

“OK so my first thought was why on earth is it not Stelarc”

The research participants had many pre-conceptions and expectations of the robots abilities. Thought, memory, emotions, sight, hearing, controlled movements, dance, singing and a whole range of other expectations can be identified by reviewing the data collected. Some participant’s expectations were very high and they reported being disappointed by the interaction, which they expected more from, but others reported surprise when encountering flow in the interaction after lowering their expectations, because of their initial impression. This is an interesting point because it highlights a tangible link between the level of expectation in a performance and the perception of pleasant surprise or disappointment reported from the audience in this interactive performance.

Some participant expectations were the result of information received prior to the interaction, such as the expectation of the Articulated Heads ability to think or retain information in memory, whereas others were more likely the result of approaching the interaction, such as the belief that the Articulated Head could see or hear E-Appendix 29: Example 7; E-Appendix 30: Example 8. However, there are two resoundingly strong overarching expectations emanating from the research participant data that appear consistently held; 1) that the robot is expected to possess human-like capabilities; memory, environmental awareness, sight, hearing. - and 2) that these capabilities
should evoke responses that reflect those of a human in behavioural patterns, such as; politeness, courtesy, expression. For example: The robot should possess memory, be able to remember ones name once told, and be able to say goodbye politely using that name, especially if the robot has asked for the name in the first place, otherwise it appears impolite to the human user. For another example: The expectation that the robot should pay attention to the person who they are conversing with, or that the robot should smile in congenial response to human facial expression, or when exchanging a joke. Other examples include participants expecting the robot to know about other aspects of the museum exhibits and have knowledge of all sorts of things including, but not limited to, visual and auditory awareness of its immediate environment and the people within it.

These overarching macro dynamic preconceptions and expectations of humans in this human ↔ machine interaction are confirmation once again, of the strong anthropomorphic stance adopted by humans to this interaction. But what happens when these expectations are not met?

Enculturation describes a process that people go through when learning the requirements and behaviours expected of a surrounding culture. The Articulated Head and its interactive environment engender enculturation of those humans who spend more time with it. One participant in particular did spend a long time interacting with the Articulated Head. This participant also had a second interaction recorded and did a second Video Cued Recall interview. Aspects of enculturation do come up in this participant’s reports with reference to two specific aspects of conversational foci: 1) She became very aware that even a small mistake in typing could have a significant effect upon the Articulated Head’s response on conversation, this made her much more careful of her data input accuracy. 2) The participant found that she was able to make the Chatbot repeat itself with specific input:

“to get him to repeat something I would ask the question again and he would always give me the same answer”
The repetition of output from the Chatbot was also noticeable in observations of interactions with the general public. For example; when the Chatbot was asked ‘what kind of food do you like?’ – the answer invariably included Sushi and Capsicum.

7.3.3 Theme 3: Physical Presentation and Operational Aspects

This theme is directly linked with sub section 9.2.3 Physical Presentation and Operational Aspects and also has a very strong correlation with features discussed under the next two themes: Theme 4: Mode of Interaction and Theme 5: Movement, Tracking and Control.

There are a number of elements pertaining to the physical presentation and operational aspects of the Articulated Head that were picked up on through observation, by research participants, and in textual input/output analysis that had an effect upon engagement over the duration of this study. Most of the operational aspects identified were very practical in nature and normally noticed as aspects of the operation that were not working properly.

The operational status of the Data Logger, Sonar Proximity Client, Audio localizer and the Face Tracking Software were only partially operational, or nonoperational for the majority of the duration of this study; as mentioned under section 2.8 Software component based architecture. The reasons for partial or nonoperation are as follows;

**Data Logger**

The data logger was implemented by engineers at the outset of the Articulated Head’s debut in the Powerhouse Museum, and was working properly for February and March of 2011. However, an upgrade to Articulated Head system, which I understand included a new computer, was implemented in March/April 2011 and unfortunately an oversight in the data transfer process meant that the data logger’s functionality was not installed with the new implementation. It was not until June 2011 when I asked the engineers to do a log file backup that I discovered that this functionality was no longer operational. Furthermore the old files were not available to the
engineers because the old machine had been wiped for another purpose. I was told that a new logging system would be implemented in due course but it was not made clear to me when.

At this juncture I decided to implement my own logging system, hence the text input/output analysis starting in June 2011. The logging system I created was not as comprehensive as the original engineers system, because I was not given open access to the Articulated Head’s central systems or the data strings required to implement a fully featured data logger. At the time, only the User input and Chatbot output strings were available to be passed from the Articulated Heads Event Manager to the Max interface for my use in the Max-programming environment. Expansion of the Event Manager/Max interface to accommodate the passing of a little more data was tentatively achieved in June 2011 and the system was stable by August 2011, courtesy of Zhenghi Zhang who was most helpful in assisting me to interface with the Articulated Head’s Event Manager to facilitate the passing of the robotic arm’s coordinates into the Max programming environment. This made it possible to calculate the position of the robot’s display screen and to place its voice in correlation with the position of the face, in the spatial auditory field.

Non-implementation of the original engineers data logger, which logged far more information related to the Articulated Head’s operational status than my restricted version did, was not such a big problem because, as has been stated in sub section 7.1 the Text String Input/Output Analysis, the files collected between June 2011 and November 2012 are thought to be highly reflective of the conversational exchanges that transpired between the Articulated Head and its human interacting audience. However, as will become clear in further discussion of the functions of the sonar proximity sensor, audio localizer, face tracking software, and other possible related and linked functionalities of the Articulated Head detailed in section 9: The Blueprint of Emergent Recommendations, these extensions of the exhibits capabilities in conjunction with other extensions and a macro dynamic rearrangement of the interactive environment, are found to be critical to
addressing the barriers to fluent interaction that have been identified through text input/output data, research participant Video Cued Recall interview reports, and the interpretive analysis thereof.

**Sonar Proximity Client**

The sonar proximity client was implemented as a way of helping to delineate between different Users interacting with the robot. The sensor would register when a person was standing in front of the kiosk and this data was registered in the original engineers implementation of the data logger. Although this registering would have been a little hit and miss, because people who did not interact with the exhibit still walk passed the kiosk, it would have helped to identify where text input from a User started and stopped.

The sonar proximity sensor was mounted inside the kiosk above the screen and was functional for a very short time at the beginning of the Articulated Head’s time in the Powerhouse Museum. Unfortunately the sensor became a feature of interest for young children, and it was not very long before kids poking their fingers into the mount point hole destroyed the sensor. A grill was fitted but this was not sufficient to stop the problem. Furthermore resurrection of the sensor was not useful after the engineer’s data logger became non-operational because the relevant strings were not available to me.

**Audio localizer**

The concept of the audio localiser establishing the position of a person in the audience with the use of a matching-pair microphone set-up, which measures the phase difference between the pick-up signals of each microphone and calculates a horizontal angle of incidence from the data, makes some sense. However, because of the ambient noise present in the public exhibition spaces where the Articulated Head was exhibited, this systems accuracy has to be questioned at the very least.

The audio localiser was said to make a contribution to confidence in the systems identification of the location of a person in the audience, by comparing the figures generated with those generated by the stereo
camera tracking system. Under certain conditions this localiser data might help to increase confidence in the audience location identification, but not under the conditions that were in place during this study, and most certainly not the conditions that were implemented in the Powerhouse Museum. The mounting positions of the matching pair microphone set-up, which were placed above the projection screen to the rear of the exhibit in the museum [Appendix 1: Plan view diagram], and the glass enclosure between the microphones and the interacting audience, compounded with the ambient noise present in the exhibition environment, sound from the lab area, which was much closer to the microphones than the audience, and loud noises from other exhibits such as the train behind the Articulated Head, must have rendered the audio localisation system and data collected all but completely useless, providing nothing but erroneous and misleading information for comparison with the stereo camera tracking system coordinates. Erroneous Audio Localiser data would help to explain the erratic movements of the Articulated Head’s robotic arm, which are linked to participant complaints about the robot not paying attention to them. This is discussed under Theme 7: Engagement. Further discussion of a macro dynamic realignment of the stereo microphone system is discussed in section 9.2.3 Physical Presentation and Operational Aspects.

**Face Tracking Software**

The face tracking software was working at the NIME exhibition; this is evident from the mimicking reported by participants in Video Cued Recall interviews and evident in interaction material. Theme 1: Anthropomorphism video links in the Key Evidence Links Table have already presented evidence of this, so I will not present it again but this direct link between evidence and themes does help to amplify the crosspollination permeating across the data set. This crosspollination is evident when viewing audiovisual data. To detail the minutiae of every crosspollination is beyond the scope of this document, but the macro dynamic links in this data are of the greatest interest to this investigation and formulate the main findings brought out in this themes analysis that are then re-platted in section 9.2.
The face tracking software was also operational for short periods of time, in both the SEAM and Powerhouse exhibitions. The software presented the engineers with a number of difficult challenges, by frequently causing crashes of the Articulated Head’s systems. There were also crashes caused by other things. Each time a crash happened, restarting the systems took some time. The engineers eventually set up auto-restart routines to circumnavigate the work required restarting the systems after various crash scenarios. These auto-start routines significantly reduced downtime. However, after many attempts to get the face tracking software to work properly with the rest of the systems, the engineers abandoned this feature because it was causing too many persistent problems.

As the software was causing problems at the time, it was important to make sure the exhibit was perceived to be operational whilst on public display; the face tracking software function and expansion of the use of information that could be stored in a database presented opportunities for enhancement of engagement in this interaction that are important - and do address issues raised in research participant reports. More discussion of this subject is offered under 9.2.3 Physical Presentation and Operational Aspects.

The problem with failure to implement and expand upon the above functionalities and operational aspects of the Articulated Head’s systems, was that these functions, regardless of the technical difficulties that they presented, were crucial to addressing interactive engagement. Without prolonged resilient research, development and implementation of these and similar features and functionalities of the Articulated Head or any other similar interactive system or exhibit, the quest for improvement of human machine interaction is essentially thwarted.

**Crash/Restart**

Other examples of the crash and restart problems are manifest in the text input/output analysis. The high occurrence of the phrase ‘do you want to hear a joke’ identified in text input/output analysis in Table 7-30 - Word ‘hear’ phrase query list is one clear example.
Here is a typical ‘do you want to hear a joke’ input/output set of exchange strings:

Time: 12_05_31 Robot: Do you want to hear a joke
Time: 12_05_35 Participant: YES
Time: 12_05_39 Robot: Do you want to hear a joke
Time: 12_05_39 Participant: YES
Time: 12_05_39 Robot: Do you want to hear a joke
Time: 12_05_51 Participant: IS THAT IT?
Time: 12_05_51 Robot: You have got me, I am dead, give me 15 second to restart
Time: 12_06_07 Robot: ok, I am back, let's talk

For some reason the phrase would be repeated three times before the crash and restart. The engineers denied me access to the code, which was unfortunate for all sorts of reasons, one being that I could not diagnose the above problem. However, this was one of the reasons why the engineers instigated the auto-restart routines, because this crash loop happened quite often and I do not think they found the source of the problem either. The restart routine was successful most times but not every time, if the robot was not given 15 seconds to restart and further text input was presented, the crash restart routine was more likely to fail.

**Latency**

One participant reported experiencing latency with regard to the time that it took the robotic arm took to follow when they were moving around in the space in front of the exhibit. Other related reports under Theme 7: Engagement show that several participants felt the robot was not paying attention to them, because it was moving around and looking at other people whilst in conversation with them.

Latency is interesting because people interacting with the robot exhibited the most common aspect of delay in conversational flow, however it was not uncommon for participants or the general public to take a very long time to decide what question they would ask of the robot, or what response they might give to Chatbot output.
Noise
Ambient noise present in the surroundings of the Articulated Head in the Powerhouse Museum was a problem for a number of different reasons, but it was manageable until the major construction project gained momentum. After commencement of the construction project the ambient noise levels in the museum were raised considerably. The robots voice required more amplification so that the public could decipher what the Chatbot was saying.

Spatial audio also needed more amplification but this just added to auditory confusion. The ambient noise levels present in the laboratory area became unworkable. Ambient noise and the effects in relation to the spatial auditory system are discussed in more detail under Theme 12: Auditory Visual Interactive Environment.

Other operational aspects commented on
One participant said that it took a long time for the robot to wake up; this was related to their physical presence within the interactive space. The robot did wake up immediately upon User text input.

E-Mote utterances were output from the Chatbot text to speech engine from time to time. E-Mote references are directly linked to triggering facial expressions of the Articulated Head so they should really have been kept internal to the system and not sent to the text to speech engine. Furthermore, the range or E-Mote functions available were not fully implemented with the exhibit. The E-Mote functions presented an obvious route for utilisation of extended capabilities of the avatar, where the infrastructure and programming already existed.

Text Space
The kiosk screen present above the input keyboard displayed the two lines of text input by the user. This limited input string length was observed as being inhibiting for Users who wanted to input more than a short sentence. This restricted input string length is likely to be one of the contributory factors that
explains the imbalance of User input to Chatbot output discussed under section Observation 3 – The Input/output word ratio imbalance.

Facial and physical characteristics
There were some generally positive comments from participants regarding presentation of the Articulated Head’s face and movement related to voice, for example: Participant comments on the correlation between speech and facial animation was quite positive - “it does actually look like he is talking”

However, there was also some less positive commentary on the way in which the robots facial expression did not appear to follow the context of conversation taking place – smiling when happy for example. Indeed the robot confirms this with statements such as; “the body lacks any human emotions” – this statement makes it appear that the robot does not experience emotion in its body but the statement does not preclude it from experiencing emotion in its head or brain; Indeed the Articulated head did say things such as “I am very happy today” on a regular basis, indicating that it did experience humanlike emotions.

The female face/head displayed at one point during the Powerhouse exhibition, procured positive commentary on how real it appeared but also some negative commentary on how the Chatbot output did not alter with the change in physical presentation. The Chatbot still said things like ‘I am Stelarc, I am 62 of your human years’. The problem with this is that the female face looked more like 20 years old, not 62. The personality that is reflected in the Articulated Head’s performance as a result of Stelarc’s strong contribution to the programming of the Chatbot, clearly exhibits a male personality in the agency given.

One of the most interesting aspects related to the physical presentation that was reported by some research participant’s and also came up in conversations with members of the general public, was the perception amongst people interacting with the robot that it appeared somewhat like a Zoo animal, caged! There was a feeling of separation and distance experienced when interacting with the robot. This observation is and possible
realignment of the macro dynamic interactive environment to address this point is revisited in Physical Presentation and Operational Aspects.

**7.3.4 Theme 4: Mode of Interaction**

This theme is directly linked with sub section 9.2.4 Mode of Interaction.

It was mentioned in section 7.1.3.4 Observation 4 – Conversational diversity across the data set that the vocabulary of the Chatbot looks wider than the Users based on review of text input/output string data, but it was also noted in Theme 3: Physical Presentation and Operational Aspects above that user input strings length was restricted to what one could see on the kiosk screen. Hence the User input string length restriction is likely to be a contributory factor to the range of words used in conversational exchange.

There is a very large number of research participant and interpretive commentary references to problems associated with User input at the keyboard kiosk. These references relate to missing keystrokes, participant input mode references, and lack of punctuation characters available for User input from the kiosk keyboard. Participants commented that it was hard to type; keyboard input was slow, and required attention to typing. Problems such as typing mistakes and worn key labels affected fluent conversation. Some participants reported that they felt the tempo or speed of exchange was important to conversational flow, much the same way as it is in human ↔ human communication. These references span the entire participant group and are uniformly spread in reporting. Therefore this keyboard interaction holds a central position in the findings from this theme E-Appendix 31: Keyboard Evidence and also holds a central place in terms of major findings of this investigation.

To consider how this ‘text in-speech out’ modality of communication affects the interaction: Firstly, although this is an altered modality from that which is immediately natural to a human, in human ↔ human communication, it is less of a problem than it might have been fifty years ago. Many people who interacted with the Articulated Head were accustomed to working with computers and a keyboard. Texting and email are commonplace modes of
information exchange in modern society. However, this particular modality has the clear disadvantage in that one is normally looking at the keyboard when typing, not the entity one is attempting to communicate with, although one can look up before pressing return. It has already been noted that research participant interaction videos show an inordinate proportion of the participants time during interactions is being spent looking down at the kiosk keyboard, and then another significant but smaller proportion of interaction time looking directly at the face of the Articulated Head, waiting for a response. The time and attention-space left for the human to take in any other environmental information is greatly constrained by this single fact.

Having established text string input functionality to the Event Manager; it was then necessary to set up a specific speech prompt that triggered the Articulated Head to say ‘press enter’. I did this in order to combat audience disengagement with the exhibit and to save time of having to leave the lab to explain to patrons. This functionality was only used at times when I was present in the laboratory adjacent to the exhibit and could see a member of the public struggling to get a response from the robot, it was only used once with a single research participant to see what would happen if I introduced the apparent ability for the robot to see clothing.

Interestingly, one research participant commented in a Video Cued Recall interview that the text in-speech out altered modality when conversing with the Articulated Head would be less of a barrier to engagement in interaction, if the robot replied in the same modality. The participant said that she was very computer literate but the pre-conditioning that using computers, email, mobile communication tools and chat rooms in particular had given her, made the cycle of ‘text in-text out’ far more natural than the ‘text in-speech out’ modality that the Articulated Head installation displays. It was noted that; as it is with ‘text in-text out’, so to would it be with ‘speech in-speech out’. It is interesting also to note that some anecdotal and some concrete evidence collected from audience interactions with the Articulated Head at the museum supports this concept of matching modalities in information exchange.
exchange. Many people wanted, hoped and expected to be able to speak to the robot E-Appendix 32: Example 9.

Research participants recorded interacting with the Articulated Head did, at specific points in interaction, instinctively revert back to a speech modality of information exchange with the robot. This speech modality phenomenon was also observed with several members of the general public. One specific and very compelling instance is described in section 7.3.6 where a young man gets angry with the Articulated Head. The speech modality phenomenon appears to happen when the robot says something personal "to or about" the person interacting with it. Participants gave the impression in video cued recall interviews that they had forgotten momentarily that the Articulated Head was a robot. It is clear from some research participant reporting in the video cued recall interviews, and from other observations, that the human-machine barrier was temporarily suspended, that the human attributed human capabilities to the robot, that the robot had in some respects temporarily passed the Turing test during the interaction that was taking place. I decided to test this speech modality phenomenon further with one of the last research participants I interviewed and interestingly, from just that one short test, one appears to be able to elicit this response from a human in interaction by getting the robot to say something specific about the person interacting. I did this by using the aforementioned text string input functionality to the Event Manager.

The specific research participant test I mention above shows in the Video Cued Recall interview that; the robot said “I like your pink shirt” to which the participant typed “it is not pink” the robot replied “I am colour blind” the participant looked at the Articulated Head and said “Yeah, you are!” The two strings: “I like your pink shirt” and “I am colour blind” were introduced into the conversation manually via the text string input functionality to the Event Manager. This was the only time that text strings were introduced into a participant interview. The participant had no idea that I was able to make the robot speak independently from its preprogrammed response mechanisms, and it appears that because the robot had something specific
to say about the personal attributes of the User interacting with it, even if the
detail regarding the attribute was inaccurate, the user, momentarily at least,
attributes a sense of consciousness to the robot and drops the percept of
human ↔ robot information exchange in favour of a human ↔ human
exchange - the modality of communicating is altered in that instant from text
input to speech. This human speech response phenomenon (titled The
Freakish Peak) is a hypothesis emergent from this investigation and is
graphically represented by the diagram introduced earlier (see figure 2-1
page 18, section 2).

Whilst the test detailed above did introduce a real human consciousness to
the robot’s systems during a single short interactional conversation with the
Articulated Head, and this could be construed as cheating, concealing
truths from research participants in research experiments to maintain the
integrity of tests results is not unusual. The intention - and fact was that this
test was a simulation of a stimulus that will improve human engagement in
human machine interaction. The significant point is; if the robot could
establish personal attributes of the person interacting with it, via sense
apparatus afforded to the system (as is recommended in conjunction with
the design refinements detailed in the text and diagrams presented in
section 9) - and could then speak to this person about the attributes
identified, then the elicitation of the human speech response phenomenon
as detailed above, would not be simulated with the introduction of a human
consciousness as it was in this test. It would be ‘real’ in the respect that its
functionality was endogenous to the robots systems, and would therefore
not be cheating, but autonomous.

Furthermore, since this investigation was focused on finding ways in which to
improve human machine interaction - and research data prior to
conducting this test very strongly indicated the phenomenon’s existence, this
test and result constitute a valid attempt to triangulate findings from the
collected research data with extra sampling, to affirm the existence of this
phenomenon in a simulated test, and to ascertain the phenomenon’s
sustainability as a repeatable stimulus in the interactive environment. The
results of the test in conjunction with the other data and observations, confirms that the stimulus is repeatable. Variation of the identified attributes of the individual interacting with the robot would increase the sustainability of this phenomenon as a repeatable stimulus.

Development of the robots capability to identify attributes of a person interacting with it, in order to stimulate the Freakish Peak speech modality phenomenon and to increase engagement in human machine interaction is considered a key outcome of this research project. Some of the design refinements detailed in the text and diagrams presented in section 9 are specifically recommended to address this research outcome.

The attribution of perceived consciousness of the robot by the human interacting with it, is I believe, one of a set of keys to enhancement of audience engagement. Extension of multisensory capabilities to gather specific details pertaining to the interacting audience is essential to be able to invoke these kind of audience responses and furthermore, one would need to cater for the altered modality of the invoked instant in order to prolong the affect. I.e., the robot needs to be able to respond to the speech of the person interacting otherwise the invoked response fails to sustain engagement, hence the need for a voice recognition system.

The positive impacts upon human ↔ machine interaction that automatic speech recognition might bring to this and similar interactive environments is significant and discussed in more detail in relation to Theme 12: Auditory Visual Interactive Environment and in section 9.2.4: Mode of Interaction.

7.3.5 Theme 5: Movement, Tracking and Control
This theme is directly linked with sub section 9.2.5 Movement Tracking and Control.

The audio localizer functionality and its useful contribution to the Articulated Head’s tracking functions have already been discussed in theme 4 above. The stereo camera tracking system worked well for the vast majority of all
exhibition time with the Articulated Head, and although it obtained some negative responses in relation to robotic arm movement latency, and the not paying attention me phenomenon reported by research participants, it obtained some very positive reactions from participants and the public alike. Children were often observed trying to play hide and seek with the robotic arm, which was wonderful to see. In fact children’s engagement with the physical side of interaction was much stronger than that of their adult counterparts, possibly because typing has little appeal at a young age, and in the case of the really young ones, they may not have been tall enough to see the keyboard anyway.

One of the main complaints of participants reporting on movement and tracking of the Articulated Head was that it appeared distracted and was not paying attention to them. This attention phenomenon appears important because people seem to be put out when attention is not focused on them during conversation, not paying attention is perceived as rudeness.

There is another phenomenon that is clearly evident in participant interactions, and has also regularly been a feature observed in public interactions, that is; the wish of Users to exert some control over the robot and its responses.

The control word anomaly identified in the text input/output analysis presented under Table 7-7 - Word ‘say’ phrase query list identifies the words ‘say’ ‘sing’ and ‘dance’ as control words; that is words that were either the trigger or a central word in a phrase that triggered a specific response from the Articulated Head. One could argue that all input triggered a specific response, based on the selected 28000 or so preprogrammed responses available in the Articulated Heads Chatbot’s programming. Where these trigger related words are different, is that the User was expecting and trying to elicit control over a particular response.

Attempts at User exerted control over the robot manifest themselves in a variety of differing ways in interaction data, from the attempted control of
mimicry in conjunction with the face tracking software, when it was working, to the control of the movement of the robotic arm by mimicry and movement in the interactive space, to the use of the word “say” with a following phrase to control Chatbot output. Other examples include attempts at conversational control, by trying to get the Chatbot to give specific responses to various text input strings, to specific command strings intended to evoke physical or facial expressions such as; ‘smile’, ‘laugh’, ‘cry’, ‘sing’, ‘dance’.

All of these and similar instances of User exerted control of the robot are interesting in that they seem to nurture engagement, especially when the exerted control is perceived by the participant as working. It is also interesting that in this transaction the User is noticeably giving agency to the Articulated Head, and is receiving feedback that confirms that their efforts have been successful. In this respect the human user has contributed, albeit momentarily, to the artwork and exhibit through the act of interaction, they have left a mark. This aspect of interaction related to User exerted control is thought to be an important human instinct and is considered to provide an avenue for improvement of engagement in this human ↔ machine interaction as discussed in sub section 9.2.5 Movement Tracking and Control.

7.3.6 Theme 6: Dialogue
This theme is directly linked with section 9.2.6 Dialogue

7.1 presented the Text String Input/Output Analysis, which treated user input and Chatbot output largely as separate entities. The following looks at the dialogue and the conversational interaction largely from a combined perspective.

During conversational interaction between humans and the Articulated Head, many recurring audience questions put to the robot resulted in recurring robot responses, this is already evident from review of the Text String Input/Output Analysis tables.
A key aspect of conversational interaction that is emergent from participant reports of the experience, and has been raised in several other places already in this thesis, is that participants appeared to be drawn into a sort of interactive vacuum, where they were focused almost exclusively on the keyboard and face, at the expense of almost everything else. This stands in stark comparative contrast to normal human ↔ human conversation in a number of differing and very important ways. Whilst it is normal to make eye contact at various intervals during human ↔ human conversation, it is also normal for the gaze to be free to take in the visual environment. The eyes and mind build a panoramic vision of the immediate surroundings, and it is not uncommon for aspects of these panoramas to become subjects of interest in the conversational foci-taking place. Table 7-16 - Word ‘Harry’ phrase query list - clearly helps to illustrate that the environment within which the exhibit and human User are situated, can influence conversational foci and be reflected in conversational references.

The same is true of spatial audio in this respect; sounds and voices may become the subject of discussion. The brain uses head related transfer functions, which play a pivotal role in helping humans locate sound sources in environmental space. “Head movements are known to help listeners to perceive the location of a sound source” (Mason, Kim, & Brookes, 2009). Therefore it follows that constriction of gaze and head movement will inevitably inhibit a human’s auditory visual perception. A human’s focus on the conversational foci taking place is desirable, but only if the conversation taking place makes sense and has flow over a protracted period of time, or number of exchanges. Focus on conversational foci to try and make some sense of it, at the expense of everything else, is not desirable within the context of the interactions under investigation herein.

There were all sorts of anomalies identified in conversational interaction between humans and the Articulated Head. The following is an example of an anomaly reported:
“The robot contradicts himself, by saying that he is both self-programmable and unable to self-program.”

The robot stated in one conversation that he (it) was a libertarian. The participant comments that the robot being a libertarian was rather odd, as he was attached to a post in a museum so the ideology must be rather difficult to implement.

There were many reports with mixed messages regarding conversational interaction; some parts of conversation were reported to flow well, whereas other sections were reported to go badly. Sometimes participants reported banal responses, whereas at other times they reported the robot making good sense. Generally participants experienced a little of both good sense and nonsense in conversational interaction.

Vocal characteristics of the robot affected intelligibility of its voice. Participants who were having trouble understanding what the robot was saying, cited pronunciation and inflection, or lack thereof as the main cause of misunderstandings. Some participants complained that the robot was speaking too fast, whereas others said it was speaking too slow. Generally participants felt it was difficult to understand the robot due to the speed of talking and the lack of inflection in the voice.

There was a range of very coherent responses and engaging aspects of language picked up on in recorded interactions and interviews. One participant reported that they felt the robot was being nice to them and therefore, they felt like being nice back. This participant believed that this process of reciprocal niceness improved the conversation, and sense of engagement as a result.

Some participants reported it being difficult to converse with the robot. There appeared to be a number of different reasons for this, but the difficulty of keeping a strand of conversation flowing, tethered with the a waning will to keep trying, appeared to be the main underlying reason for participants
experiencing this difficulty. Some latency was reported in sections of the interaction where the participant had to wait for the robot to respond to an input.

Questionable links between exchanges, banal replies to User input from the robot, repeated sections of sentences or repeated responses to User input were all recorded. There were many cases where the robot’s response did not match the question asked. Many sections of conversation suggest that the robot did not understand the user input. Some participants reported the conversation being shallow and said that conversation could only be maintained with short superficial input (small talk).

Several participants reported their pleasure that the robot showed signs of being interested in them. This was a pattern that has come up in many places in the data. It appears that it is important that the robot pay attention to the human interacting with them.

There are a multitude of other micro dynamic conversational exchanges that could be examined in more detail, for instance; some people perceived that the robot was being rude at times, some said it was sleazy, some reported concordance and agreement with the robot whereas others reported misunderstandings and disagreement.

However, the macro dynamic messages that emerge from review of the dialogue are that, conversational flow rarely extends beyond a couple of exchanges and that humans find driving the conversation hard work, that humans want the robot to be interested in them and generally report more positive feedback when the robot shows an interest, or agrees with them.

Fundamentally, as is typically the case with human communication, engagement appears to be directly linked to the robot showing an interest in the audience and finding concord with them on subjects of the human’s interest. The result of this concord, when found, is that the human reports liking the robot. To address this macro dynamic finding in conjunction with
Observation 1 – *The anthropomorphic stance,* a few simple recommendations for improvement of conversational interactive engagement are put forward in section 9.2.6 *Dialogue.*

The opposite of the above, where participants have experienced rudeness or disinterest emanating from the robot, invariably seemed to nurture some degree of human disengagement from the interaction. Disengagement was often shown by changes of subject in conversation, but has been observed to put members of the public off altogether. *Table 7-31 - Word ‘talk’ phrase query list* – displays the phrase ‘I don’t want to talk about that now’ emanating from the robot, and this phrase has been observed to engender frowns and other displays of upset from humans interacting with the robot.

One very interesting general public interaction I observed entailed a young man typing rude words into the kiosk. One of the engineers had set up a Chatbot response that called the User a loser when they used this particular four-letter word. Upon being called a loser by the Chatbot, this young man instinctively went into a fit of rage and started shouting expletives, pointing his finger and shaking his fists at the robot. I found this immensely amusing but also very interesting because it showed me that the instinctive nature of humans to default to the modality of speech is very strong, and lingers close to the surface of human machine interaction even when subdued.

Another interesting aspect of dialogue exchange is that Stelarc’s personality shows through in the Chatbot programming from time to time. *Table 7-33 - Word ‘western’ phrase query list* – showed Stelarc’s strong interest in the Western Bulldogs and country and western music in particular. This observation raises the idea that concordance with a human User in conversational interaction could fairly easily be simulated in dialogue by copying, storing and then repeating preferences of the User in subsequent conversation.

*Table 7-35 - Word ‘old’ phrase query list* - identifies the two most consistently common phrases to appear in both User input and Chatbot output, ‘what’s
your name’ and how old are you’. User answers to these questions clearly present opportunities for enhanced engagement in conjunction with memory and representation as discussed in section 9.2.

Opportunities, which allow the robot to show more interest in, and concordance with the User, and to initiate conversational topics rather than being driven by user input, are expected to enhance conversational engagement in human ↔ machine interaction.

7.3.7 Theme 7: Engagement
This theme is directly linked with section 9.2.7 Engagement.

Table 7-29 - Word ‘time’ phrase query list identifies that the ‘spare time’ phrase is normally prefixed with ‘what do you do in’ and that this phrase does encourage User input and also indicates that the robot was interested in the User. The crosspollination between themes is slowly but surely becoming much more obvious. Various aspects pertaining to engagement have already been raised in the themes preceding this one.

The robotic arm movements are linked to participant reports saying that the robot was not paying attention to them, and have also been cited as having a possible link to erroneous stereo microphone input. Participants report that the robot appeared distracted. Latency in human input is linked to people having to think hard to work out what to say and then needing time to pay attention to the typing, to get the words input correctly. Participants report becoming bored or exhausted with the interaction after a while, because it was difficult to keep a conversation going, they felt they were not getting anything back. Participants wanted the robot to be interested in them and found it annoying when the robot appeared to ignore them and pay attention to other people in interactions. Participants became excited when the robot showed signs of being interested in them or remembered something individual about them. When the robot was nice or friendly, participant’s tended towards being nice and friendly back.
The same things seem to crop up over and over again in the data, but the most prominent message of all with regard to engagement is that; it would be more engaging if you could speak to it. It is interesting to note that although some participants instinctively approached the interaction with speech, some participants who were already in the throes of text input still tipped over into a speech modality, this invariably happened after a preceding exchange which strongly suggested the robots possession of consciousness. It appears that the trigger can be accidental or deliberate and still procure the same outcome. However the triggers do appear to relate to the robot being conscious of a particular aspect or feature of the human, that it would be unlikely to know without senses or consciousness, hence triggering the attribution of consciousness in the human mind leading directly to the subsequent speech action.

It is not clear whether systematic deliberate placement of triggers will consistently stimulate the same response but this observation clearly provides opportunities for development and further testing as discussed in section 9.2.7 Engagement.

7.3.8 Theme 8: Emotions
This theme is directly linked with section 9.2.8 Emotions.

The emotions theme shows a strong correlation with previous themes. The anthropomorphic stance taken by the interacting audience manifests in the checking of emotions just in the same way as it transpires in the checking of preferences. This checking of emotions appears rather like a human prodding another creature that appears to be dead, just in case there is some life left in the creature. Strangely though, if one had a reaction from another creature after prodding, one would almost immediately assume that there was some life left in the creature and the prodding would stop. However, this prodding by the participant to gain a response appears to continue even after confirmatory responses from the robot were received. It is as if the human somehow can’t believe it was true that the robot possessed the capacity for emotion. This further highlights the conflict
between what one wants to believe and what one ultimately must know only exists within the biological condition.

Amazingly though, there are instances in the research data where a research participant has attributed the capacity for feeling emotions to the robot. For instance there is an exchange where the participant asks the robot if he was embarrassed, the robot replied saying that he doesn’t have feelings to which the participant replies, “I don’t believe that!”. In the Video Cued Recall interview the participant confirmed that she didn’t believe that. This suggests that although it is difficult to elicit attributions of conscious existence from the human towards the robot, it appears equally difficult to reverse the attribution once given. This directly links with the hopes, wants and beliefs of the human interacting with the robot. The term Scotoma describes the condition where the mind sees what it chooses to see rather than what is actually there. For example, the mind can fill in gaps when reading text so that the sentence you are reading makes sense even when words are missing. It is as if the minds eye has a blind spot, but the mind itself allows sense to be made from the otherwise normal data, and clearly if one knows what the mind wants to see, then one can give the mind a helping hand in seeing it, through simulated performance that points in the desired direction! This human predilection for Scotoma represents opportunities for leverage in improvement of this human machine interaction as discussed in section 9.2.8 Emotions.

It seems to me that the systematic checking of preferences and emotions by the human interacting with the Articulated Head was really all about sharing existential experience. Humans appeared to want to be in charge of the interaction but simultaneously wanted the robot to be able to share common objects of consciousness and sedimented layers of experience as perceived through any and all of the senses, and to be able to discuss and have opinions related to those elements raised during the interaction.

Emotions expressed by the robot and by the humans in conversational interaction, needed to be acknowledged by the Articulated Head’s
performance in various ways to show some concordance with the human’s existential experience. This needed doing, not just in words but also in gestures, facial expressions and tone of voice. Empathy, sorrow, fear, anger, loathing, happiness, humour and so on needed some simulation of human-type responses present in the Articulated Head’s performance. If not human-type responses, then at the very least, some robotic type responses that the human could easily draw parallels with. This expansion of nuances in the virtual performer’s performance was needed in order to combat the conflict between hopes, wants and reality that appears to exist in the human mind in relation to interactions. The aim would be to promote human favour of the hopes and wants that nurture belief in, and attribution of existential experience to the robot.

7.3.9 Theme 9: Senses – Related

This theme is directly linked with section 9.2.9 Senses – related.

The senses have been raised and their crosspollinations are evident in several other themes, and have been a subject of interest throughout this thesis. Table 7-30 - Word ‘hear’ phrase query list references the Chatbot’s use of the word ‘hear’. The use of the word inferred to the User that the robot could hear. Several pieces of evidence already presented point to the fact that people interacting with the Articulated Head hoped, wanted and expected the robot to be able to see and hear them, this message is loud and clear in evidential data.

It has been noted that there was much less reference to the other senses, smell, taste and touch, in research data. Possibly this was because of the void and distance created between the Articulated Head and its audience by the enclosure. Perhaps the aforementioned caged zoo effect reduced the relevance of these other senses in the interactive environment to some extent, or perhaps the human instinctively thought the robot less likely to possess these senses anyway. Whatever the reason, further discussion of possible extensions to the robot’s performance in relation to these senses is conducted in section 9.2.9 Senses – related.
7.3.10 Theme 10: Memory Related

This theme is directly linked with section 9.2.10 Memory related.

In textual evidence for this theme, there is an engaging example of where a participant has shared a secret with the Articulated Head; examples of where the robot shows self-awareness and displays a possible capacity for learning, and teaching in one case are also present. There are further examples of participant wants and hopes regarding conversational flow. Review of the Table 7-30 - Word ‘hear’ phrase query list shows the phrases ‘question I don’t hear everyday’ and ‘I only hear that type of thing’: both phrases infer memory related to previous conversations.

The robot stated that its prime directive was to collect new knowledge, yet there is scant evidence of the robots ability to do this in conversational interaction. There is more than one example in the empirical data where the robot has told a participant that he was searching for something in memory or on the Internet, and then he failed to follow through with the results of that search. In one case the participant expressly asked for the results and was not given them. The participant commented in Video Cued Recall that they had a fear that they maybe harbouring unrealistic expectations of the robot. However, if the robot had followed through with some results, the participants fear or holding unrealistic expectations would have been allayed, then repeated re-affirmation of their expectation being realistic would very likely to lead to embedded enculturation that would have a positive, and possibly permanent affect upon future interaction, leaving the participant with the belief that the robot does possess the skills to search its own memory and present results.

Sometimes the Articulated Head did succeed in showing that it had a memory of conversational themes that had just passed. However, this memory seemed to be limited to the recall of short strings, and often the section of string chosen did not fit well into the context of the conversation currently taking-place, and sometimes the pasted string did not fit well into the constructed sentence either. Sometimes the Chatbot remembered a
person’s name during interaction and other times it did not. Sometimes the Chatbot carried over a name that was stored from a previous interaction into the current interaction; this was also true of subjects that it seemed to recall.

The Articulated Head, and more specifically its Chatbot appeared to display a distinct lack of memory during interactions. Participant reports have shown that this can lead to frustration and annoyance. Some active display of memory is believed to be a very important factor in tying together all the findings from the themes that have preceded this one. How expansion of memory capabilities might have improved the Articulated Heads performance in human ↔ machine interaction is discussed in more detail in section 9.2.10 Memory related.

The concept of thinking is normally shown in conversation by recalling strands of previously discussed or digested information, and then representing new reason based on consideration of these strands and synthesis of their contributions to meaning in the context of ‘now’. This process sounds complicated and indeed it is. Fortunately the biological brain is uniquely equipped for performing such processes. However, simulation of thinking in the Articulated Head’s performance could be, and was to some extent achieved through the construction of pre-programmed responses. The problem is that to simulate thinking convincingly, synthesis of prior information and its contribution to meaning in the context of ‘now’, must be included in the simulations ingredients. Pre-programmed strings can simulate synthesis of prior knowledge, but cannot be equipped with the facility to place synthesis of it in the context of ‘now’ without the inclusion of some current variable. Current knowledge pertinent to the immediate situation and surroundings could, and should have been pre-programmed into the Chatbot’s memory repertoire for inclusion in conversation, but it is the inclusion of current variables in simulation that makes the performance really convincing. It is for this reason that the Articulated Head needed to be able to establish at least some current variable conditions in order to place
conversational foci in the context of ‘now’. Ways in which this could be achieved are discussed in section 9.2.1 Memory related.

7.3.11 Theme 11: Outliers

This theme is directly linked with section 9.2.1 Outliers.

Having reviewed the data in the nodes sorted under this theme, I found that virtually all of them could be satisfactorily sorted under Theme 1: Anthropomorphism or Theme 6: Dialogue. This left me with just four very short snippets of nodal data that just did not fit anywhere else.

Two of the snippets very briefly relay comments related to science fiction references. One participant comments:

"I think the display or the head could become more engaging, for me personally, if it kind of bought up some of these issues with artificial life dominating human life".

There were a few other references to science fiction identified in the text input/output analysis and participant interaction data, such as references to Star Trek, R2D2, Star Wars and the phrase 'life the universe and everything'. Given the context of the robot being presented as a Thinking Head, perhaps some more Chatbot programming related to these subjects would be interesting.

The other snippet of information that could not be placed comfortably under another theme relates to one participant indicating that she felt that there was a third party involved in the interaction, she states:

"it was paying attention to me, which was nice and that sense of being watched was, I guess you would say, more natural than the sense of something else watching us - It would be like me having this conversation with you and a security camera taking us, that sort of, someone else sitting behind that security camera watching what we were doing, where as, I attributed the camera 'it' to - it then interpreting what I was doing" -
Interviewer interjects - "so effectively like you meeting a robot that was being watched by the aliens that created it" participant confirms with a "yeah, for sure and I think the other thing is too, you can sort of immerse yourself in the conversation and interaction but as soon as you notice there is something else up there it sort of - snaps out the magic of it I guess - so as soon as you notice there is cameras and there is, it sort of makes you go oh - someone might be controlling that ...it destroys the illusion".

The participant was alluding to the fact that hiding aspects of the exhibit is important to creating the illusion, which sustains engagement. The participant was aware of the fact that she was being recorded before she entered the interaction at the time, but nevertheless, the magic of the illusion is both reported and destroyed within her statement, and presumably the same was true within her interaction as well.

The participants comments strongly suggest that immersion in the magic of the illusion in this interaction can be reasonably easily reached, but it is important not to present any features in the exhibit design that would destroy the illusion quickly, because the illusion is just as easily (if not more easily) deconstructed.

7.3.12 Theme 12: Auditory Visual Interactive Environment

This theme is directly linked with section 9.2.1 The Auditory Visual Interactive Environment.

This theme reflects upon, and evaluates the auditory visual additions tested in parallel with research participant interaction with the Articulated Head in the Powerhouse Museum. Within the context of the embodiment of the human participant’s brain during this interaction, the critical importance that dimensional layout and display have upon the effectiveness of audio visual aids and the strength of spatio-temporal contextualizing cues in relatively unconstrained interactive public exhibition spaces is considered.

Conclusions presented in section 9.2.12 The Auditory Visual Interactive Environment contribute a refined experimental project and exhibit design,
aimed at expediting more encouraging participant reportage of the enhancement of engagement in Human Computer Interaction with this, and similar types of interactive installation exhibits.

**Video Cued Recall Outcomes**

The Video Cued Recall interview data identifies a very broad range of phenomena experienced by participants during interaction with the Articulated Head. The following text predominantly imparts only features that directly link participant experience to the auditory visual additions.

Extracts of Video Cued Recall interview data, where research participants were making direct reference to the auditory visual environment, are presented in as Appendix 3 The Auditory Visual References Table. These extracts have been specifically selected as they clearly impart the overarching impact of the auditory visual interventions. They are letter coded according to; Column ‘S’

The following is one example of a short extract:

‘Participant explains that although he was aware that the projections were there, he did not really pay any attention to them as he was focused on the interaction, on the keyboard and face.’

The overarching message extracted from all the Video Cued Recall data related to the auditory visual environment, including but not limited to the comments listed in the table in Appendix 3, was that most participants did not notice the auditory visual additions and even when they did, the contribution to the experience of interaction with the Articulated Head was generally perceived to be insignificant, peripheral, subconscious or even divorced from the mainstay of their experience of interaction with the Articulated Head.
This apparent failure of the auditory visual additions to enhance engagement in participant interaction with the Articulated Head installation was upon first impressions very disappointing.

However, a positive message of partial success of the auditory visual additions was also present within the Video Cued Recall data, and although this message is carried by a much quieter voice, it cannot be ignored. Perception of particular phenomena presented by the auditory visual additions, which were touched upon by some participants, are in fact a very close - or even exact fit to phenomena specifically targeted as desirable outcomes of the auditory visual design:

a) The concept of the ‘audio bubble’ creating a more intimate environment as mentioned in extract F.
b) The fact that extracts B & I identify that the words in the projections related in some way to the conversation that was taking place. (The processing of thought)
c) The fact that one extract J, appeared to have picked up on the contextual intentions of the auditory composition.

These points raise serious questions because one cannot consider the auditory visual additions to be either a failure or a success in light of the analysis; If the auditory visual additions were conveying to participants some aspects of phenomena intended by design but their contribution was perceived to be insignificant, very small, peripheral, subconscious or even divorced from the mainstay of the participants experience of interaction with the Articulated Head, rather than a centrally related, embedded, augmenting and enhancing aspect of participant engagement in this human computer interaction, then the following questions become pertinent;

a) What features of the designs were ostensively obscuring or diluting the perceived phenomena’s impact upon participants within the environment?
b) How can one effectively amplify the impact of the noticed phenomena and convey the presence of those intended phenomena, which have not yet been recognised by the participants.

c) What other mediating considerations and constraints beside the above should be taken into account in order that one might expedite more encouraging participant reportage of the enhancement of engagement in human computer interaction with the Articulated Head or similar interactive installations, whilst retaining integrity of the exhibit within the wider museum environment?

**Mediating considerations and constraints**

Appendix 1 is a complete plan-view diagram of the exhibit layout and surroundings. Please refer to it for clarification of any features of the exhibit that are referenced by enclosing in [ ] brackets in the following text, including dimensions and position of any physical aspect of the enclosure and identification of speaker mounting positions.

**The Spatial Audio System**

Loudspeakers [L1, L2 & L3] projected their dispersion patterns out toward the opposite [Wall A]. They were mounted to the left hand side of the enclosure [apex] just above ground level, spaced evenly and inset into the enclosure wall for the purpose of maintaining a flush face to the exhibit enclosure walls. The direct sound dispersion angle of each speaker was too low to allow any direct sound to reach the ears of a person standing at the kiosk. Two loudspeakers [L4 & L5] mounted similarly to [L1, L2 & L3] but to the right hand side of the enclosure [apex] when facing the robot, projected their direct sound dispersion pattern away from a person standing at the [kiosk] - out towards the [large doorway] into the Engineering Excellence Awards Exhibition (EEAE) and [Wall B]. This meant that only low-level reflected sound from hard surfaces, which were considerably further away from the speakers than was the case on the left hand side of the [apex] was reaching the ears of a person at the [kiosk]. No direct sound from those speakers was reaching the ears of the person standing at the kiosk at all.
Bob Katz, in his book, Mastering Audio comments “did you know that wearing a hat with a brim puts a notch in your hearing at around 2Khz (Katz, 2007, p. 46)”. The reflective surfaces that feature as part of the enclosure were obscuring direct sound from reaching the ears of a person standing at the kiosk and were also affecting the frequency characteristics of source sounds in a similar way.

Frequency cancellation, the colouring effects of comb filtering and the altered frequency content of reflected sound, caused by speaker mounting positions and their subsequent dispersion patterns, were all ubiquitous within the spatial auditory environment. These confounding aspects of the built environment compromised aspects of the experimental design, having a pivotal impact upon the ability to balance the Distance Based Amplitude Panning (DBAP) (Lossius, Baltazar, & de la Hogue, 2009) system effectively. Balancing this system was critical to the effectiveness of virtual sound source placement within the spatial audio environment. Distance Based Amplitude Panning is one of a range of auditory spatialisation techniques, chosen in this case, as it is a useful technique for auditory spatialisation where placement of loudspeakers units in irregular or undesirable positions becomes necessary. Although Distance Based Amplitude Panning allows for irregular speaker placement, the system still has limits in this regard.

An Ono Sokki LA-1210 (Type 2) (“ONO SOKKI-Products Information & Service,” 2012) sound pressure level (SPL) meter, which conforms to [JIS C1502 Type 2, IEC 60651 Type2, Draft IEC 61672-Aug 1998 Class2, ANSI S 1.4 Type2] standards, was used for taking sound pressure level readings. The unit’s omnidirectional microphone was placed at the centre of the expected listening position of a person standing at the kiosk - 1.65m. Full bandwidth 20-20Khz pink-noise was sent from Max environment out to each speaker in the SPAT array in turn, and then a decibel reading was taken. A purpose built iPhone Max app, created using the c74 (“C74,” 2012) object provided the ability to control the volume of each channel of the SPAT Max patch remotely, whilst taking readings at the kiosk. The spatial audio system was
balanced by obtaining the same sound pressure level reading at the listening point from each speaker in the array.

Ambient noise in the museum was measured on a normal day before the aforementioned construction project started at the normal kiosk position. The reading produced a typical sound pressure level of around 83db. The maximum sound pressure level reading obtainable from the most obscured speakers in the SPAT array (L4 & L5), using pink noise at the maximum system volume obtainable without clipping was 86db. Although human hearing perception has high resolution and a change of 1db in sound pressure level of a source present within the sound field maybe perceptible to the experienced listener, the average listener does not readily notice it (Katz, 2007). Because all speakers in the SPAT array must be balanced to the same sound pressure level reading as (L4 & L5) to render a balanced DBAP system, they were all subject to the same governing constraints. Therefore, the effective system headroom for raising source levels above the ambient noise present within the spatial auditory environment was equal to or less than 3db, rendering sound pressure level variation of virtual sound sources barely perceptible to an audience at the kiosk position.

Furthermore, this balancing act left an unacceptable situation created by the exhibit, whereby the volume of audio output from speakers (L4 & L5), although supplying the correct sound pressure level reading for the meter at the kiosk listening position, rendered the audio volume in the museum space to the right of the enclosure [apex], far too loud to retain integrity of the exhibit within the wider museum environment. This problem manifested itself with several speakers in the array to a greater or lesser degree, depending on obscurations and/or distance from the meter at the kiosk position.

Other confounding and compromising aspects of the spatial auditory system design included: speakers [L8, L6, U5 & U7] mounted on [Wall A & Wall B] of the Engineering Excellence Awards Exhibition, which were probably too far away from a participant standing at the kiosk to fully contribute to the spatialisation. Speakers [U2, U3 & U4] were directed towards the participant
with the top half of their diffusion pattern intersecting directly with the participant’s ears. Unfortunately a significant amount of low absorption coefficient surface material was affecting the diffusion field. Glass rose 1.83m from [Pillar A] out to the [apex] of the enclosure and subsequently protruded into the speaker diffusion field, obscuring approximately 40% of the signal from the audience. The low absorption material reflected a significant amount of high frequency content emanating from those speakers, effectively muffling the clarity of audio material and the intelligibility of the robots voice projected from those units.

Mounting speaker [U8] was impossible due to the building design and restrictions on fixtures and fittings allowed in the Powerhouse Museum. Omission of this speaker further impacted the effective dimension of the spatial auditory field.

The ceiling height in the Powerhouse Museum was approximately 8m and the building in which the exhibit sat was similar in size to a large aircraft hanger, which acted as a giant reverberation hall, colouring the acoustic environment. The prevailing mixture of less than ideal speaker positioning, high ambient noise and the creation of an unhelpful acoustic environment by the enclosure and surrounding surface materials, rendered a drastically blurred and ostensively ineffective spatial auditory sound field.

**The Thought Cloud Projections**

The projection screen consisted of a sticky back plastic GlassVu Mk3 high-grade projection film mounted on the right hand side of an area of Glass between [pillar A and pillar B] behind the Articulated Head, when viewing from the kiosk position. To the left hand side of the [projection screen] a large glass door provided internal access to the enclosure from the Video Cued Recall interview lab. The glass between [pillar A and pillar B] rose approximately 3m from the ground and the projection film rose from a height of 1m to the top of the glass screen, covering an area approximately 2m². The distance from the Articulated Head screen, mounted on the end effector of the robotic arm, to the projection screen, varied according to the
movement of the robotic arm but was normally in the region of 2.5m to 3m from the centre of the screen at a head height of 1.65m. A significant proportion of projection screen was obscured from the audience by the robotic arm and the screen mounted on the end effector. The degree of obscuration varied a little but because the robotic arm tended to follow audience movement, obscuration remained fairly constant.

The Video Cued Recall interview data includes extracts D & E that comment that the participant did not really notice anything about the projections and that they were focused on finding out what the robot could do, or were focused on the head itself. Extracts G & I also comment on the fact that although they were aware of the projections they too, were generally more focused on other things. We do know from the wider body of Video Cued Recall interview data that people interacting with the robot were predisposed to focus on the head and keyboard. It is clearly evident from the video data that an inordinate amount of participant gaze and time was directed at the keyboard during interaction with the Articulated Head.

To put forward a plan for the unrestricted design refinement and presentation of auditory visual stimuli accompanying this, or similar exhibits, that takes into account the tangible, embedded, and embodied interaction reports of research participants gathered from Video Cued Recall interviews and the subsequent findings from the themed interpretive phenomenological analysis presented herein, conjunct with the aforementioned aspects of auditory visual presentation and perception thereof, a fairly radical yet clearly palpable set of improvements, which address all the confounding and compromising aspects of the current exhibit design has to be contemplated.

Acknowledgment of the impact that dimensional layout and display have upon the effectiveness of audio visual aids, and the strength of spatio-temporal contextualizing cues presented in this relatively unconstrained interactive public exhibition space, was critical.
Within the context of audience embodiment and in order to secure a powerful, immersive and dynamic auditory visual experience for them, auditory visual presentation of stimuli must remain a central critical concern to all design decisions made from concept through to the completion of any such installation. The compromises made with dimensional layout and display when installing the auditory visual additions in the Powerhouse, constituted the reasons that jeopardised full realisation of the design intention (Paine, 2010). One cannot compromise on the technical details of presentation of auditory visual stimuli with each obstacle that presents itself within the design and construction sequence of an exhibit, and then expect the outcome to retain veracity with conceptual intentions. In fact, with each compromise made, one experiences a dilution and depletion of the impact of auditory visual stimuli upon the interacting audience.

So, it is not a case of form over function or vice versa, it is very clearly a case of function within form. This means that function and form are of paramount importance together. Synergy rather than compromise with critical concern for the details of both visual and technical design decisions, centered around the unchangeable aspects of audience experience, presented by virtue of their embodiment, must be adhered to from the outset of any such project, if it is to be wholly cohesive and successful in meeting the conceptual intentions and outcome criteria set by the body of interested parties contributing to its implementation.

To this end section 9.2.12 The Auditory Visual Interactive Environment contributes an experimental projects and exhibit design refinement, aimed at giving an audience a more immersive and enveloping auditory visual experience and expediting more encouraging participant reportage of the enhancement of engagement in Human Computer Interaction with this, and similar types of interactive installation exhibits.
8. Pre Analysis: Key Concepts

This section presents key concepts that help to explain what has been found in the themed data analysis, how it impacts human ↔ machine interaction and how the concepts help inform design refinements presented in section 9.

8.1 A Situated Robot with Physical Separation

Solving the problems presented by the various operational obstacles presently hindering the quasi-autonomous prolonged and successful operation of robots in human environments is not that simple. The successful navigation of autonomous robots in human environments is uniquely complicated, because of the vast number of variables presented by obstacles and moving objects in three-dimensional spaces. This is one of the reasons why we have what I term robotic devices as oppose to actual robots, the difference between the two is a grey area but it is not necessary to hold that debate herein, suffice to say that we are calling the Articulated Head a robotic device and a robot, because it was not free to move in space as a human can but it did claim and present humanoid features and capabilities to some extent. It is by virtue of the fact that robotic devices are limited or restricted in one dimension or another that they function successfully within the environment in which they are designed to operate, by simplifying the number of confounding considerations that must be taken into account during their design.

The Articulated Head was one such example of a robotic device that was restricted in one specific dimension; it was a situated robot that was fixed to the ground, and was therefore unable to move from its base. This significant fact simplified its operation greatly, because the list of navigational requirements for its successful operation was dramatically reduced. Furthermore the Articulated Head was separated from its audience by an enclosure. The Articulated Head’s robotic arm was capable of navigation in three-dimensional space within a limited half spear, with a radius of between two and three metres from its base. The enclosure was situated just outside this navigational boundary, this meant that it had no obstacles to navigate and was able to move freely within its immediate space. Therefore
Articulated Head did not actually have to navigate physically in a human environment, because its space was its own, and humans were effectively excluded from that space. This physical separation helps to explain the lack of references to the sense of touch in the empirical research data set, and also explains the caged animal zoo exhibit reports by research participants in Theme 3: Physical Presentation and Operational Aspects, but the Articulated Head and its environmental conditions were very much connected and operational in the human environment on almost every other level. The Articulated Head certainly navigated the human environment on auditory visual and mental planes.

8.1.1 Key Concept 1: Intentionality

Intentionality is especially relevant in relation to analysis of interactions that were taking place between humans and the Articulated Head, because it relates to the hopes, wants and beliefs of the interacting human.

Dennett’s intentional stance (D.C. Dennett, 1989) describes an innate ability that we as human beings are endowed with. That is not to say that other biological beings are not endowed with this ability to a greater or lesser extent too - but for the purposes of this discussion only, we shall limit our frame of reference to human endowments. The innate ability described is that we are able to predict on a regular basis, and with a significant degree of accuracy, outcomes from a set of indicative circumstances, even if these predictions sometimes turn out to be incorrect. This predictive ability of the intentional stance (D.C. Dennett, 1989) has at its heart intentionality. In what follows I explain what the term refers to and why it is important to this investigation of human ↔ machine interaction.

“Intentionality is aboutness. Some things are about other things: a belief can be about icebergs, but an iceberg is not about anything; an idea can be about the number 7, but the number 7 is not about anything; a book or a film can be about Paris, but Paris is not about anything. Philosophers have long been concerned with the analysis of the phenomenon of intentionality, which has seemed to many to be a fundamental feature of mental states and events” (D. C. Dennett & Haugeland, 2012, p. 1).
One might argue that an iceberg is about something, it is about an accumulation of frozen water, or that the number 7 is about the quantity or measure of something. However leaving this semantic argument aside, Dennett and Haugeland’s quote, most importantly, raises the analysis of the phenomenon of intentionality in this thesis, and more specifically the identification and analysis of phenomena, which appear to extend from it.

“Phenomena with intentionality point outside themselves, in effect, to something else: whatever they are of or about. The term intentionality was revived by Franz Brentano, one of the most important predecessors of the school of phenomenology” (D. C. Dennett & Haugeland, 2012, p. 1)

In essence intentionality is said to be the aboutness of mental states and phenomena such as wants, hopes and beliefs: latent or static charges that a belief or hope may carry, and what those charges appear to want to gravitate towards, that being the thing that it (the belief or hope itself) is about. Notably when this description of intentionality is subjected to closer examination it becomes clear that intentionality comes with the concept of intentional relations, which carry with them some interesting characteristics and attributions that do not follow the rules normally consistent in ordinary relations: Dennett and Haugeland point out that, a belief can be about both real and non-existent entities. The possibility of the inexistence of the object of intentionality, the object that a thought, hope, want or belief is pointing to, is especially relevant in relation to analysis of interactions that took place between research participants and the Articulated Head and how one can improve this human ↔ machine interaction.

“Brentano called this the intentional inexistence of the intentional objects of mental states and events, and it has many manifestations. I cannot want without wanting something but what I want need not exist for me to want it” (D. C. Dennett & Haugeland, 2012, p. 2)

Why this quote is important to this investigation is that the intentional inexistence of intentional objects appears to be a plausible explanation for some of the participant behaviours exhibited and observed during
interaction with the Articulated Head. The research data shows that participants certainly hoped, wanted and believed in capabilities of the Articulated Head that did not exist.

### 8.1.2 Key Concept 2: Combinatorial Explosion

Daniel C Dennett speaks of “combinatorial explosion” ([D. C Dennett, 1997, p. 77](#)) in relation to a systems design. To illustrate this, Dennett employs a “thought experiment” (often used by philosophers). Dennett asks the reader to imagine a hypothetical competition between a human and a seemingly hyper-intelligent Martian. The human and the Martian pit their predictive skills and methods, being what Dennett calls the “intentional stance” and strategy of the human as opposed to what he calls the “Laplacean deterministic physical stance” utilised by the Martian, against each other respectively, to see which of them can make an accurate prediction first.

Based on the details of an observed telephone call the human and the Martian use their predictive skills to determine what they think will happen as a result of the content of this phone call. The phone call proceeds as follows:

*The telephone rings in Mrs Gardners kitchen, she answers and this is what she says: “Oh hello dear. You’re coming home early? Within an hour? And bringing the boss to dinner? Pick up a bottle of wine on the way home then, and drive carefully”. ([D. C Dennett, 1997, p. 68](#))*

The human predicts the arrival of a car at Mrs. Gardener’s house with two humans in it, one carrying a bottle of wine. The human makes the same prediction as the Martian - but arrives in an entirely different way (and much faster) leaving the Martian amazed at the apparent intellectual dexterity of the human, who the Martian had previously perceived to be the lesser intelligence. The apparent magical predictive ability of the human amazes the Martian because the Martian is bereft of any knowledge and skills that the intentional stance and strategy purvey. The Martian assumes that the human must have arrived at the same conclusion by calculating all the possibilities and variables that the Martian had calculated - and had done so much faster; therefore the human must possess previously unrecognized
processing powers. Work smart, not hard, is the underlying moral of this tale. However, in relation to systems design and “combinatorial explosion”:

“Increasing some parameter by, say, ten percent - ten percent more inputs or degrees of freedom in the behavior to be controlled or more words to be recognised or whatever tends to increase the internal complexity of the system being designed by orders of magnitude. Things get out of hand very fast....” (D. C Dennett, 1997, p. 79).

8.1.3 Key Concept 3: Combinatorial Reduction

Therefore, it makes sense when designing and redesigning systems to consider the opposite of “combinatorial explosion” (D. C Dennett, 1997, p. 79), here the term combinatorial reduction springs to mind - so this term will be used henceforth. Combinatorial reduction is thus defined as the employ of methods, which achieve design refinements, whilst avoiding or reducing the impacts that combinatorial explosion might have upon their implementation.

Indeed combinatorial reduction is at work in the design of many devices, where restriction of particular parameters is desirable in order to render the devices operation successful in the physical and practical worlds.

Combinatorial reduction as a rule of thumb

Combinatorial reduction as a rule of thumb is a necessary and frequently desirable aspect of both project management and systems design, given the ramifications of combinatorial explosion (D. C Dennett, 1997, p. 79). The Articulated Head was no exception with regard to this rule of thumb, and reference to retrospective combinatorial reductions made to the original operational design perspectives as presented in the grant application E-Appendix 2: From talking heads to thinking heads and in the diagram in Figure 2-4 (Herath, 2012) were implemented, in the regard that it became too difficult to maintain and secure the successful functional status of the Data Logger, Sonar Proximity Client, Audio localizer and the Face Tracker Software, or to implement some of the original plans, due to the size and commitments of the project team involved. Further reference to
combinatorial reduction and its employ is made in the design refinements put forward in section 9.

8.1.4 Key Concept 4: Embodiment

The book, How The Body Shapes the Mind (Gallagher, 2005, p. Introduction) references the seemingly deterministic and certainly influential aspects of embodiment and its role in conditioning us as ‘soon to be’ humans. Once separated from our mothers’ wombs, this conditioning, so powerful has already been, that we can see reflection of our own form in the face of others as soon as we open our eyes, and are almost instantly capable of facial imitation, the smile and so forth. Gallagher goes on to suggest that embodiment is an inescapable fact and that we, as a human brain and mind are first and foremost embodied, and that this embodiment has inexorable and as yet unfathomed consequential influences upon our very nature and existence, of our sense of self! This influence of embodiment is said by Gallagher to go beyond consciousness into the unconscious mind and even into what he terms as the “prenoetic” or “before you know it” (Gallagher, 2005, p. 5). Embodiment certainly shapes and possibly even facilitates the existence of our sense of self; it influences the structuring of consciousness and therefore influences our perception of everything phenomenal and intentional experience included of course. The theoretical viewpoint projected by Gallagher is, as a human being, not hard to swallow, even the prenoetic element, that which we are not yet conscious of. Indeed a few minutes of quiet contemplation and reflection upon our consciousness and sense of self, renders this theoretical perspective as not only likely but seemingly a sure thing.

The associations that embodiment imposes upon a human brain has a very powerful influential, if not deterministic affect upon a humans perception of their lifeworld and the phenomena experienced through their senses. It then follows that this deterministic affect is present and at work in the human ↔ machine interaction under investigation in this thesis.
Embodiment and its ramifications are a critical aspect considered in relation to the evaluation of The Creative New Media Additions detailed in section 6 and evaluated under Theme 12: Auditory Visual Interactive Environment.

Embodiment of the avatar that was the face of the Articulated Head is also important in terms of a human’s perception of it as an intentional agent. Human perception of the performance of robotic devices when observed in action, generally procures the attribution of a largely unintelligent status. If the Articulated Head did not have a screen displaying a face on the end of the industrial robotic arm, and the arm moved randomly or in predefined patterns rather than actually tracking you as the human observer of it, then a similar perception and the attribution of a largely unintelligent status would be likely. However, because the industrial robotic arm was capable of displaying more complex actions and movements in three dimensions, which more realistically resembled aspects of human movement, it was more likely to procure human observers attribution as an intentional agent. The moment the arm begins to track your movement and position as the observer, or performs a specific task of one form or another that requires some complexity, it immediately, from the human observational perspective, became an intentional agent, it appeared to have a mind of its own and could therefore, possibly represent a threat.

8.1.5 Key concept 5: Identification of agency:
Self-preservation is a primal instinct of all living creatures and the will to defend ones existence in the space-time continuum is an overwhelmingly compulsive intrinsic mechanism of all living creatures. To decide upon defensive actions, creatures generally identify moving, and possibly intentional agents, and subsequently attribute the possibility of intelligence and the ability to commit to intentional acts to them, therefore raising their status as possible threats until it has been established otherwise.

Charles Abramson of Oklahoma State University, when discussing the biological criteria for fine-tuning of intentional agency in regard to worms, comments that:
“Only organisms with central nervous systems are capable of fine-tuning their bodily movements for the performance of intentional acts” “Internally generated flexible behaviour appears to be confined to organisms with central nervous systems” (Abramson, 2012)

Whilst Abramson does not concede that internally generated flexible behavior can appear to be exhibited by a pre-programmed agent, with either random or sensor controlled threshold routines, it should be noted that the above extracts were made in the context of reference to biological entities only and it is normal for humans to only attribute the ability of internally generated flexible behavior to other creatures. Where intentional acts of an object of attention other than another creature are perceived by a human observer, the human very quickly concludes that an external, and very possibly human agent (because of the intelligence required) must be involved, that that agent is trying to exact control over the object in question, and that the external agent may also have extensional intentional acts motivating the control of the object of attention in question. The human suspicion of a controlling agent is rather neatly demonstrated under Theme 2: Expectations and enculturation by the children in the Powerhouse Museum who thought I was controlling the Articulated Head.

Abramson notes that “mental states such as beliefs and desires are primarily identified through the performance of intentional acts, which presuppose the notions of trying and control” (Abramson, 2012)

Trying and control as words are an adjective and conjunct noun that are clearly identifiable as characteristics operational in intentional acts, both are synonymous with human behavior, and both can be exacted through given agency. For just one example of this, one can try to convince the human interacting with the Articulated Head, that the Head is interested in the human and is executing an intentional act - by controlling the Articulated Head’s arm and screen position to face the coordinates of the interacting human. This can be achieved by including a stereo tracking camera linked to the Articulated Heads industrial robotic arm motor commands through a
preprogrammed coded interface, as was the case with the Articulated Head’s implementation. However, let us not forget that exacting trying and control through given agency entails the ramifications of system design synonymous with “combinational explosion” (D. C Dennett, 1997, p. 79). To endow any non-human object with given agency that can process and perform the diversity of intentional acts with the apparent consciousness, dexterity, flexibility and speed of a human being is thus far beyond realization through application of science and technology, whilst we may be able to imagine it, we are not able to realize it within the confines of our knowledge of the physical world. The reasons for this are fairly simple; we do not have sufficiently sophisticated sensory apparatus and processing units that can perform at levels and resolutions comparable with the human brain. It is worth noting that the primal instinct for self-preservation appears to gravitate towards other moving and possibly intentional agents, because these are the attributes that a human possesses by virtue of embodiment and a central nervous system, and presumably because a human naturally recognises that these attributes proactively make them a threat to other beings. That is not to say that other threats to human safety and existence do not exist within the environment, clearly they do, it is just that in the hierarchy of active attentional immediacy in relation to this primal instinct, self-protection gravitates towards moving, thus possible intentional agents, first.

This means that movement can easily attract human attention because it activates a human primal instinct that initially attributes the status of ‘possible intentional agent’ to the moving object, but what happens after the immediacy of this attribution to retain human attention.

**8.1.5.1 Human abilities taken for granted**

Human beings have a tendency to take their own abilities for navigation, decision-making and maneuvering within a three dimensional environments for granted. They can effortlessly avoid physical contact with moving objects, some of which have the potential to harm them such as collisions with cars and other people. They can circumvent obstacles that stand in the way of their path to reaching a destination and can easily recognise elements that are present within the immediate environment. They can identify at a glance
the constitution of many elements that surround them, animal vegetable or mineral. They can also identify many liquids; solids and gases through comparative analysis of data collected by their sensory apparatus in conjunction with the cross-referencing of their memory and lived experiences at lightning speeds. They can conceive, design, build, pick up, move and rearrange three-dimensional objects and remember their proper places in their operational spaces with ease. Humans are incredibly versatile, complex and dexterous entities. Human are nothing short of completely amazing and thus it takes them only a very short period of time to separate attribution of given or predetermined agency from that of agency delivered by an active and present intentional agent in an interactive environment.

Extension of the time period required for human identification and attribution of these differing types of agency is important in relation to the interactions that were taking place between the Articulated Head and its interacting audience because it represents a clear target for one way in which interaction between humans and machines can be improved. User input string data analysis presented in section **Theme 1: Anthropomorphism** identifies a human proclivity towards anthropomorphism in this interactive environment, hence providing an opportunity for trying to exact some control over the period of time taken by the human to separate attribution of given or predetermined agency from that of agency delivered by an active and present intentional agent in this interactive environment by catering for this anthropomorphic proclivity through the use of suggestive, generative targeted intentional language acts. This suggestion is discussed in more detail in 9.2.6 **Dialogue**.

The following quote sums up the remarkable demonstrable capabilities of our brain, and poses the question how does it do what it does?

“Now somehow the brain has solved the problem of combinational explosion. It is a gigantic network of billions of cells, but still finite, compact, reliable and swift, capable of learning new behaviours, vocabularies, theories, almost without limit. Some elegant, generative, infinitely extensible principles of representation must be responsible. We have only one model of
such a representation system: A human language. So the argument for a language of thought comes down to this: what else could it be? We have so far been unable to imagine any plausible alternative in any detail” (D. C Dennett, 1997, p. 77).

The point of presenting the quote above is that it nicely articulates the humans language of thought as a key target for investigation, indoctrination and manipulation in relation to extension of the time period taken by the human to separate attribution of these types of agency. Furthermore, since the overtly problematic issues associated with combinatorial explosion in machine development cannot readily be solved by the current ‘state of the art’ in science and technology, a focus on developing the human’s perception of engagement and agency in interaction as opposed to intensive machine development, is clearly highlighted as one of the key ways in which interaction between humans and machines can be improved in such interactive environments.

8.1.5.2 Internally generated flexible behaviors: The source of agency
The human instinctive opposition to the idea that a non-biological object conducting intentional acts can actually be representing generative or internally generated flexible behaviors has enormous implications with regard to participant interaction with the Articulated Head, and indeed to human computer interaction in general because: If the object of attention is not a biological being, then any concept of relationship building in reflective, rather than reflexive interaction, transcends the object of attention by transference, such that the object of attention now becomes the agency and source of that agency rather than the object itself. This point of departure between the object conducting intentional acts and the source of its agency - that which is trying to control it, has profound implications when attempting to answer the big question being:
- How can the interaction between humans and machines be improved?

8.1.5.3 Interacting with what? The giver of agency
Asking another question can highlight the problem with this big question from a semantic and theoretical, if not practical and physical perspective: What
exactly is it that the human in this relationship is interacting with? The human on an intellectual level is, in fact, interacting with the giver or givers of agency to the machine, and not the machine itself – the human is interacting with a third party (another human or humans) through the machine, and the machine itself is the medium through which this interaction is taking place. Naturally the agency present during interaction is limited by the ramifications that “combinational explosion” (D. C Dennett, 1997, p. 79) brings to the system design, and consequently improvement of interaction between the parties involved holds a fairly direct linear and correlative relationship with the scope and limitations of the agency given to that machine.

8.1.5.4 Extension of agency given to the machine

The extension of the agency given to the machine by increasing the number of parameters to be controlled or providing “more inputs or degrees of freedom in the behavior to be controlled or more words to be recognised or whatever” (D. C Dennett, 1997, p. 79) will directly expand the scope of interaction taking place, and will therefore result in perceived improvement in interaction between the human and the machine, regardless of the semantic and theoretical point just made in the paragraph above.

8.1.5.5 Transparency of the medium

Put even more simply, on a very practical, physical, tangible level, without any consideration for the mental states and beliefs that may exist in the human minds that are a party to these interactions: If the machine becomes the medium through which communication between the real generative intentional agents is taking place during in these interactions, then the transparency of the ‘human to human’ communication taking place through this medium, is directly proportional to the transparency (or bandwidth) of the medium itself, therefore combinational reduction is deterministic with regard to the big question:

- How can interaction between humans and machines be improved?

Put in another very practical way: combinational reduction and its deterministic influence upon the scope of any agency given to the machine, defines the scope of any interactions that can take place. That is, it defines
the limitations imposed upon these interactions, thus directly attenuating any scope for improvement of interaction between the human and the machine. This point is true, simply because, the machine is not really the entity at the other end of the interaction to the human in the projected human ↔ machine relationship referenced in the big question.

Given the statements made in the paragraphs above, the Martian in Dennett’s hypothetical scenario, who is imbued with the Laplacean deterministic style physical stance and strategy (or perhaps I should say - limited by combinational reduction because he or she is bereft of the predictive abilities that intentionality affords the human) would no doubt conclude that the only way to improve interaction between humans and machines would be to expand the scope of agency given to the machine exponentially until such time as the machine, being the medium through which communication between the real generative intentional agents is taking place during in these interactions, becomes transparent. The Martian would likely conclude that regardless of the ramifications of combinational explosion, if one conducted enough calculations and adopted ever increasingly ingenious technological design features, then transparency of the medium would eventually be realized and the Martian would therefore set out on his/her quest to solve all the problems of combinational explosion to achieve transparency of the medium.

However, the human, perhaps through reading Dennett’s writings, though more likely because of intuitively recognising from their own experience of existence and practice in both artistic and technical creation, would note that the human brain has solved the riddle of combinational explosion in a way that is incomprehensibly complicated, beyond the scope of current human scientific endeavours and manufacturing capabilities to replicate in the space afforded by a football pitch, let alone a skull! The human would very likely conclude that the Martians quest was folly and choose a more simple option. After all, for the human, life is much too short to contemplate setting out on a quest to achieve the seemingly impossible, especially when there is clear evidence in front of the human that the Martian, who uses a
Laplacean deterministic style physical stance and strategy, performing billions of calculations to solve the problems of \textit{combinational explosion}, whilst simultaneously engineering the transparency of the interactive medium, is in fact, bereft of the ability to identify, let alone understand the consequences of intentionality, which are in fact consequences that are derivatives of the only example of an entity that has demonstrably already solved the problem of \textit{combinational explosion}, the biological brain.

Naturally the Martian does not believe that there is a simpler option, because the Martian does not understand the human capacity for the perception of, or projection of intentionality from thought and action, especially when it comes to the “\textit{intentional inexistence} of the intentional objects of mental states and events” (D. C. Dennett & Haugeland, 2012, p. 2).

\textbf{8.1.5.6 Transparency of medium and scope of interactions}

There are of course many situations where there is not a direct coloration between the transparency of the medium and the scope of any interactions that can take place through it. That argument would only hold true if each, and all of the entities involved in the interaction were machines, which they are not in this case. Although the ‘transparency of the medium/interaction scope’ correlation holds true in the physical and practical world, it does not hold true with regard to the mental states of the real generative intentional agents being the two (or more) human entities involved in these interactions.

To elucidate this point more clearly, one must first consider the fact that the term \textit{agency given} carries with it the implicit charge of its installation to the machine, and that the scope of any agency given, is only restricting the scope of interaction on a physical and practical level, not on a semantic or mentalist level, by virtue of the fact that perceived agency can be invoked and evoked rather than just installed, and can indeed, by virtue of the possibility of the “\textit{intentional inexistence} of the intentional objects of mental states and events” (D. C. Dennett & Haugeland, 2012, p. 2) be a non-entity from the perspective of the designers, or supposed givers of agency. It is quite possible for a person interacting with the Articulated Head to perceive agency that was not given by any third party to the machine.
Humans can and do perceive and believe in things that are based on very flimsy evidence with little or no concrete substantiation in the physical or practical world at all.

I make the above point explicitly here because the empirical evidence collected from Video Cued Recall Interviews of participant interactions with the Articulated Head, has clearly identified that some participants did indeed perceive, believe and act upon agency that was not intentionally given to the machine.

The constructed realities of the human mind do not have to be substantiated and scientifically accepted realities of the physical world to be real and true to the mind in which the constructed reality exists.

This point, or opinion, depending largely on your philosophical perspective of what constitutes the current state of play with regards to the explainable universe, relates to this investigation because the empirical evidence collected from participant interactions with the Articulated Head during this investigation has clearly identified that some participants did indeed construct their own realities from the perception of agency that was not intentionally given to the machine.

The degree to which each side of this interplay between levels of the physical and practical as oppose to the semantic and mentalist worlds impact upon the scope of interactions that have taken place in the human ↔ machine interaction under investigation here, is more apparent in the research data with examples including participants believing that Articulated Head could see and hear, think, feel emotions and conduct such activities as flirting.

In a scene towards the end of the Harry Potter film (Warner Bros., 2013b) Deathly Hallows part two, Harry asks Professor Dumbledore the question;
Professor, is this real? - or is this all happening in my head?

To which the professor replies:

"Of course it’s happening inside your head Harry, why should that mean it’s not real?"

This line rather nicely sums up a critical point surrounding performance and perception of it that I have been trying to put across in this thesis.

The key concepts presented above that impact the interactions that have been under investigation herein, hopefully now make it clear that both practical and physical restrictions, agency given, the transparency of the medium, intentionality and the various mental states of any, and all of the biological entities a party to interactions, including any actions, events and behaviours perceived or instigated, whether accidental or otherwise, real or imaginary, rational or irrational, reflexive, reflective or reaction-al and relational, that may result from installation, invocation and/or evocation, all have a role to play in the interactions - and therefore impact upon this investigation.

8.1.6 Key Concept 6: Acquisition of knowledge

The investigation of how human ↔ machine interaction can be improved is also influenced by acquisition of knowledge from the data in analysis, for example; there was that that one could establish or deduce from observation of interaction; the Articulated Head turns to face the participant or the participant appears to be losing interest and so forth - but this tells us nothing of what the participant is actually thinking or feeling. Then there is that that one could establish or deduce from what a participant said in the video cued recall interviews; the Articulated Head is getting angry or he is so rude, this is participant declarative knowledge about something perceived and does not require that something to exist in the real practical or physical world, for it to exist in the mind of the one that declares this knowledge. There is also that that can be established or surmised through actions and repetition of actions, including other aspects of interactivity such as the
speed of participant responses to familiar scenarios. These actions typically manifest during interactions where a participant displays repeated actions, or learned behaviours, as the result of enculturation to the specific interactive environment. This procedural knowledge may be implicit to the interaction but not explicit in participant declaration. There is also that deduced and/or induced knowledge, existential experience and constructed realities that the person analysing the research data brings to the analysis table and findings, further complicating the interpretation of data.

So, given all that has been said, though it is not exhaustive of all related theories and theoretical perspectives in existence, it should now be clear that this study had a very broad brief, with the central interactions under investigation defying a reductionist methodology due to the contribution and combinatorial explosion bought to the investigation by the inclusion of the human minds in the interaction equation. That is to say, the only clear example of an entity that has demonstrably already solved the problem of combinatorial explosion, the biological brain, brings unto the investigation the same problem that it alone has solved.

This point, with hindsight may appear obvious to you as the reader, if so, then good, because this point defined the chosen methodological approach (phenomenology) and vouchsafes the methods adopted during this investigation and also frames the validity of any findings firmly within the realms of interpretation of phenomena. Correlations between types of phenomena experienced during interactions, the frequency with which they occur, and interpretation of the conditions that appear to support manifestation of that particular phenomena, whether this be with just one participant or across a range of participants, these are the main features of focus for this investigation.

It comes as no surprise then that Franz Brentano, was “one of the most important predecessors of the school of phenomenology” (D. C Dennett & Haugeland, 2012, p. 1). The potential for “intentional inexistence of the
intentional objects of mental states and events” (D. C. Dennett & Haugeland, 2012, p. 2) is a uniquely located phenomena, known only to exist in the biological brain of creatures of consciousness. Therefore, since phenomenology, as a methodology, is focused on the phenomena of consciousness, its status as an appropriate candidate for employ in this investigation is thus ratified.

So, what exactly do we need to know in order to address the big question that we are trying to answer in this study?

1. **We need to identify any phenomena manifest during interactions.**
2. **We need to identify conditions that appear to support manifestation of these phenomena.**
3. **We need to establish whether the phenomena identified, appears to have a positive or negative influence upon human ↔ machine interaction**
4. **Then we need to consider the conditions that appear to support manifestation of any particular phenomena, and how those supporting conditions might be increased or decreased in frequency during interactions, according to whether the phenomena’s influence on interaction has a positive or negative effect upon engagement.**

“One doesn’t reduce Turing machine talk to some more fundamental idiom; one legitimizes Turing machine talk by providing it with rules of attribution and exhibiting its predictive powers. If we can similarly legitimize “mentalistic” talk, we will have no need of a reduction, and this is the point of the concept of an intentional system. Intentional systems are suppose to play a role in the legitimization of mentalistic predicates parallel to the role played by the abstract notion of a Turing machine in setting down rules for the interpretation of artifacts as computational automata” (D.C. Dennett, 1989, p. 67)

Legitimisation rather than reduction is part of the role of an intentional system. Since all the entities active in the interactions that took place between an audience and the Articulated Head were intentional systems of one form or
another, what we really need to know is what was actually taking place during these interactions, and what are the human perceptions of participants during these interactions - so that we can find empirical evidence that effectively legitimises the apparent conditions under which occurrences of phenomena experienced proliferate, and then adjust the intentional systems to proliferate positive influential phenomena of engagement, whilst obliterating negative influential phenomena of engagement, thereby improving human ↔ machine interaction.

8.1.7 Key Concept 7: The beaconing of succor
Terrel Miedanger in a short story called “The soul of the mark III beast” discusses how a human in the story does not want to kill a robot beetle. The human is given a hammer and asked to smash the small moving beetle. The human feels it is not fair to the little beetle to smash it. The human reluctantly hits the beetle once with the hammer and the beetle begins to wince and make a noise as if it were in pain, the human is now even more reticent to strike it again. The theory that a machine can beacon succor earning some empathy from the human assassin in the story is interesting, because it represents another possible lever for increasing audience engagement, it is not only that some empathy maybe earned, it is also that the threshold of this empathy giving is different for each human individual. Some individuals might find it easy to destroy the beetle whereas others may have a real problem.

This empathy felt by humans is another aspect of the anthropomorphic stance. Some would not buy the beaconing of succor and have no qualms whatsoever wheedling the hammer of destruction with vigor, possibly even reveling in the act, but many humans do appear predisposed to feel empathy. For one example; my wife, when proof reading the reduction presented in section 3 of this thesis, expressed feeling sorry for the robot because it was devoid of the senses that people hoped, wanted and expected it to have. One way or another the concept that empathy might easily be evoked in the human, raises consideration for it as a lever for increasing audience engagement as is discussed in section 9.2.7 Engagement.
9. The Blueprint of Emergent Recommendations

It should be noted here that any suggestions and recommendations put forward in this blueprint are very unlikely to be implemented with the Articulated Head, because it has now been decommissioned and its hardware has been allocated to another purpose. Nevertheless these recommendations were established from the research conducted in conjunction with the Articulated Head, and thus they are presented as if they could be implemented with the Articulated Head. Many of the recommendations will also apply to similar interactive exhibits.

The Emergent Themes from this investigation identify several small crimes committed against the flow of human ↔ machine interaction in the environment under investigation. Many of these small crimes were simple and practical in nature. Some of the smaller crimes identified had a larger impact on human ↔ machine interaction than others. Some of the small crimes were not negative influences on the interaction as such; they were more like opportunities for enhancement of audience engagement with the interactive exhibit that were not taken advantage of.

The major crime identified by analysis of empirical human centered research data in this investigation was simply that the macro dynamic interactive design of the Articulated Head, including many features of the exhibit design enclosure and auditory visual environment, failed to put the human (the customer) first in this interactive environment. The form in the design of many features of the project and exhibit took precedence over function, and simplicity of implementation in functionality took precedence over implementation of interactive functionality that would have nurtured greater flow in this human ↔ machine interaction. Whilst combinatorial explosion and technological barriers and constraints are real and important concerns, they should not mask or preclude the stimulus for trying, and most certainly not stall the quest for progress; moreover they should inform the passage for advancement. Once again, these statements are not to suggest that the Articulated Head did not engage its audience, it is just to say that in hindsight, and armed with the findings from this human centered research
investigation, it is easy to say that much more could have been done. It is hoped that the blueprint of recommendations that follow, will help others avoid the pitfalls that ensue rejection of functionality on the basis of assumed technological complexity, or constraint to advancements, and will also help anyone intending to embark upon the design, construction and implementation of any such similar interactive exhibit, by highlighting the critical implications that the embodiment of any human brain present in an interactive environment, brings to their perception of it, and hence to the design table for that environment.

What follows is effectively a macro dynamic realignment of the interactive exhibit design that addresses the barriers to human ↔ machine interaction that have been identified through this investigation in The Emergent Themes, based on concepts that help explain the theme findings and inform the new design.

9.1 The ecosystem
An ecosystem is defined as a biological community of interacting organisms and their physical environment (Jewell, Abate, & McKean, 2005), quite what one calls an environment that incorporates these factors and a lot of technology such as was the case with the Articulated Head exhibit, I am not quite sure? However the term and concept of an ecosystem is helpful in terms of setting out the main concept for the interactive design refinement that follows. The key idea is that the interactive design needs to create a separate microcosmic-ecosystem for the exhibit that is partially separated from the wider ecosystem in which it is situated (the museum or other larger exhibition space). This simple concept allows most of the practical technical and technological barriers to interactive flow identified through this investigation to be addressed.

For real exponential improvement of human ↔ machine interaction, the machine must ergonomically meet the human on their own playing field on as many different levels as is conceivably possible, physically, practically, mentally and psychologically. This is an enormously tall order, which ultimately leads to the point that now seems simple and obvious, that
humans in this interaction seem to want the robot to behave and act much like a human would, and in as many different ways as possible. Humans want the robot to come across as an existential being that can share their experience, preferences and sensory perception of the immediate surrounding environment with them. For this interaction to be truly fluent over protracted periods of time, the robot's performance must at least attempt to match that of a human. However, this somewhat impossible task is made significantly easier by two key factors in relation to the context of Articulated Head ↔ human interaction being:

1) Most interactions that take place are short with external factors to the interaction often dictating the tendency for the audiences' departure.

2) The audience hope, want and expect the robot to display predictable human preferences and capabilities.

So what exactly do we actually know about the human experience of the interactive environment under investigation? – and how exactly does this knowledge inform the design refinement?

9.2 The blueprint themes

9.2.1 Anthropomorphism

This sub section is directly linked with Theme 1: Anthropomorphism.

We know that humans are predisposed to the anthropomorphic stance as previously discussed and we know that humans can be convinced of an illusion momentarily, especially when they are not given the time for deep critical reflection. Humans can be tricked even when they know, or at the very least must suspect that an illusion is just that, an illusion, and it cannot actually be true, especially if the hope and want to believe that it is true. Examples of this included, the rabbit in the hat trick and a whole range of other pieces in a magician’s repertoire. Indeed without human partiality for believing the impossible magicians would be redundant.
This point is simple; if humans want to believe that the robot possesses consciousness, human preferences and capabilities such as senses, which clearly they do according to the research data, then convincing them that this is the case should not be so difficult – surely all one needs to do is present a cascade of evidence to the human that says that the robot does possess these attributes over the short duration of the interaction. If this cascade of evidence is not criticized by the robots subsequent actions during the interaction then, is it not possible that they might depart the interaction remaining fairly convinced?

9.2.2 Expectations and enculturation
This sub section is directly linked with Theme 2: Expectations and enculturation.

We know that humans hope, want and expect the robot to be able to see and hear, if not smell, taste, touch and feel. Therefore we must try to convince the human that the robot can see and hear. Smell, taste, touch and feel are probably more difficult illusions to sustain though not necessarily impossible. Smell is probably the easiest to implement by using some form of air constituent analysis sensor apparatus, and because it would not require physical contact, it therefore avoids the ensuing health and safety issues. Implementation of taste, touch and feel are more problematic. Nevertheless, the illusion of sight and hearing can be achieved.

The robot already shows signs of being able to see because it is an intentional agent and displays the intention of looking at the person it is interacting with. We know that because it is not always focused on its audience they have reported it as being distracted and/or consider that it is not paying attention to them. If the robot were endowed with the ability to establish features of current variables related to the immediate interactive environment and/or pertaining to the person interacting with it, then that person is likely to be convinced that it can see. This can be achieved in a number of different ways but two obvious examples spring to mind;
1) Present variable features in the spatial environment that the human would not immediately attribute as parameters that are likely to be under the control of the robot's systems and then endow the robot with the ability to raise these features in the course of conversational foci. This also gives the robot food for initiating conversation, which partially addresses the finding that it was hard to keep a conversation going. This arrangement includes one example of a way in which the avatar can be silently mutli-present, and extension of this silent multi-presence idea can further extend multisensory capabilities as described in point 2 below.

2) Provide avenues for the robot to be able to establish features of the individual interacting with it that, again, the robot can raise in conversational foci. For example, if one of the features in the spatial environment that the robot had control of, was a large white wall behind, and some distance away from the human interacting with the robot, then the human would be unlikely to attribute the wall as being a parameter that the robot might control, especially if they have their back to it anyway. If this wall could be made to change colour from white to blue or green, which can be achieved through lighting or by using materials that are electro-colour sensitive such as types of St Gobain Glass ("St Gobain Glass," 2013) then a high resolution camera could take a picture of the human and apply a pixel colour analysis to the picture after chroma-keying out the background of the image. The pixel analysis could then be used to establish various features of the human such as height and the predominant colours present on the lower and upper half of the body. It maybe possible to establish hair colour with reasonable accuracy as well. This would allow the robot to say something about the height or garments that the human is wearing, hence lending significant weight to the idea that it can see. Furthermore this would dispel any human feeling that the robot was not paying attention to them.
9.2.3 Physical Presentation and Operational Aspects

This sub-section is directly linked with Theme 3: Physical Presentation and Operational Aspects.

It is interesting that physical presentation of the speaking face was reported to be convincing, even though articulation of a phoneme was the same no matter the context because there was not a co-articulation model. The fact that the face did not incur detrimental reporting from research participants, suggests that the head speech, lip and jaw movement worked fairly well – maybe it would not be completely convincing if you were asked to concentrate on it, and I am sure anyone lip-reading would have had trouble, but in general it did not appear to be a significant impediment to intelligibility or satisfactory interaction?

There are many aspects of the physical enclosure that had affects upon the interaction; these are picked up in section 9.2.12 The Auditory Visual Interactive Environment.

With regard to the operational aspects of the exhibit, the data logger, sonar proximity client, audio localizer and face-tracking software functionality are now seen as being critical functionalities for addressing various aspects of participant reports in relation to improving this human-machine interaction. The most useful function of the face tracking software other than mimicry, is that it is possible for it to recognise and differentiate between faces with reasonable accuracy, yes, it may make some mistakes but we should remember that, just as it is with colour recognition, so to with face recognition. The goal in this interaction is to legitimize the role of the intentional agent and if the face tracking software in cooperation with the data logger, could store the name and face pattern of an individual, then clearly it would allow recognition of the face in future interactions to trigger robot initiation of the use of the humans name. This functionality addresses commentary from several themes because it clearly suggests that the robot can see, has memory and is interested in the human. It also suggests to the human the possibility of relationship building.
To address the audio localizer and sonar proximity client problems, the new design refinement presented shortly suggests amalgamating these functions into one microphone system placed very close to the human in the new microcosmic-ecosystem environment created for the exhibit. The new arrangement would reduce the erratic movements of the Articulated Head’s robotic arm, as well as addressing more fundamental problems listed in the following headings.

### 9.2.4 Mode of Interaction

This sub section is directly linked with Theme 4: Mode of Interaction.

We know that the keyboard input system was reported to be difficult for participants, and many said it was an unnatural mode for communication. This negative impact upon the human ↔ machine interaction was reported universally across the research participant set, and observations of the public use of the kiosk keyboard also confirm this. It is a major finding of this investigation that the kiosk keyboard input system was the perpetrator of the most major crime committed against human ↔ machine interaction in this interactive environment, because it failed to put the human first, it failed to meet the human on their own playing field in terms of modality of communication and furthermore, it sucked the life out of any other attempts to legitimize the intentional agents performance by drawing the attentional status of the human brain into concentrating on a typing activity with a less than conducive interface at the expense almost everything else related to the exhibit for a very significant percentage of the interaction time. We know that speaking to the robot is something that humans would prefer as their mode of interaction.

We know that there are a very large number of mistakes in input strings. We know that speech recognition technology is not perfect but it is getting better every day and there are models of speech recognition technology that no longer require training before operation. Therefore it is a recommendation, or should say instruction that installation of speech recognition technology is an imperative addition to improving the flow of human machine interaction with any such similar exhibit. The argument that it makes too many mistakes does not wash because textual data analysis
shows that keyboard input is unlikely to be better and could possibly be much worse, only time and experimentation will confirm this. There is the argument that speech recognition technology would not work in a public exhibition environment because of acoustics and ambient sound, but the new microcosmic-ecosystem environment presented in The Auditory Visual Interactive Environment addresses this concern directly.

9.2.5 Movement Tracking and Control

The control word anomaly identified in Theme 5: Movement, Tracking and Control should be explored in the programming the Chatbot and in the pre-programmed actions of the Articulated Head. For example, if the human asked the robot to dance or sing the robot could/should have a small repertoire of responses to the control words such as; conducting a short dance or song line or saying no, why should I dance or sing for you. Providing a range of control words and responses to them would enhance engagement.

What we do know is that people like to be able to make the robot do something. The humans use of the control word ‘say’ shown in Table 7.7 - Word ‘say’ phrase query list confirms this. It appears that the opportunity to contribute something to the artwork, to leave ones mark, even if it is only momentarily is compulsively engaging for humans. This phenomenon seems synonymous with the human impulses associated with graffiti and carving ‘I woz here’ into the wood of a park bench.
The projections project detailed in 6.8.3: The Thought Cloud Textual Projections in conjunction with the spatial environmental design refinement presented in 9.2.12 The Auditory Visual Interactive Environment addresses the weak impact of the projections reported by humans in interaction with the exhibit, and represents excellent opportunities for humans to contribute temporary markings to the interactive artwork, hence improving engagement in human ↔ machine interaction.

9.2.6 Dialogue
This sub section is directly linked with Theme 6: Dialogue.

A key finding of this theme is that programming dialogue to cater for the anthropomorphic stance of the human in conversational foci is likely to enhance engagement in the interaction, and this is probably also the most sensible approach to expansion of the Chatbot’s conversational repertoire.

During the course of this investigation I discussed expansion of the Chatbot repertoire with the programming engineers and one of the key concerns was related to the fact that, creating what they called a ‘say anything’ Chatbot would entail the ramifications of combinatorial explosion, requiring a team of programmers and a protracted period of time in order to achieve this. I agree with the programming engineers, and so recommend the employ of combinatorial reduction of the task by limiting the development of the Chatbot’s repertoire to the subjects that we know from text string data analysis are the main subject of interest to the human.

We know that the human appears to test and probe the robot with their anthropomorphic stance, checking its preferences against their own existential experience of the lifeworld. We know that a biological “systems desires are those it ought to have, given its biological needs and most practical means of satisfying them. Thus intentional systems desire survival and procreation, and hence desire food, security, health, sex, wealth, power, influence and so forth”(D.C Dennett, 1989, p. 49)
Maslow’s hierarchy of needs (Maslow, 1954) focuses on human needs and the innate curiosity of humans. The hierarchy of needs is represented in the diagram below:

My recommendation is that expansion of the Chatbot’s conversational repertoire should be based on the E-Appendix 13: Top 100 User Input String Words. The list should be examined to extrapolate the words that related to the hierarchy of needs as shown in Figure 9-1 above, by working from the base of the triangle to the tip and starting with the words ‘food’ or ‘eat’. Conversational strings should be expanded by providing a range of Chatbot responses to common User input probes thereby enhancing engagement in conversational interaction by variation. Furthermore the preferences of the User input to the system should be stored in the data logger (which serves as the robots memory) with the humans name and face recognition pattern so that the Chatbot in subsequent interaction can use these preferences for conversational initiation. Extrapolation of the humans name and other details can of course be achieved by targeting form field type questions such as ‘what is your name?’ at the human. This approach caters for the most common intrigue of the human anthropomorphic stance and probing, whilst combating the ramifications of combinatorial explosion in Chatbot
repertoire expansion, simultaneously to facilitating expansion of the memory and knowledge base that the robot can draw upon - leading to enhancement of engagement in future interactions.

Chatbot accord with User preferences should be displayed in conversational responses regularly because reciprocal niceness was reported to improve the conversation, and sense of engagement as a result. Humans like people who share agreement with them, and thus it is likely that this concordance will find favour in human ↔ machine interaction too.

A database/memory library related to music, film, TV and game preferences are another way in which the Chatbot’s conversational repertoire can be expanded; again User preferences could be stored. The library would provide opportunities from Chatbot initiation of conversation and could easily show concordance with User preferences by the Chatbot picking music and films from similar genre categories as the User’s stored preference for discussion.

The robot should be given that capability to search the web and other data sources so that it can present results on specific subjects such as the weather, temperature, and sport results. Some facilitation for presenting results both visually and/or through the Chatbot should be implemented.

9.2.7 Engagement
This sub section is directly linked with Theme 7: Engagement.

You, the reader should now be able to feel the re-platting of the untangled strands of the aforementioned knot beginning to come together. Engagement is the theme that all the other themes effectively contribute to.

We have already discussed several ideas that relate to face recognition, human feature recognition, memory, data/knowledge base access and speech recognition. The functionality and ideas already discussed contribute very significant ways in which the avatar can be given agency and can be multisensory and multi-present with regard to its interactive environment and
knowledge acquisition. The ways in which the functionalities and ideas put forward contribute to the characteristics of the virtual performers performance, and subsequently condition affect and engagement of the audience are clear.

9.2.8 Emotions
This sub section is directly linked with Theme 8: Emotions.

Human predilection for Scotoma has been discussed before (see 7.3.8), if one knows what the mind wants to see anyway, then one can give the mind a helping hand in seeing it, through simulated performance that points in the desired direction!

Utilisation of the aforementioned E-mote command repertoire for facial expression linked directly to User input regarding emotions is one obvious way of getting the Articulated Head to appear to have emotions. This would address in part, the need for concordance to be shown, not just in words but also in facial expressions. Gestures and tone of voice could also be linked here; sad could indicate slower talking and lowering of the industrial robotic arm whereas excitement and happiness could be simulated by the opposite. These simulations would indicate to the human that the robot is able to share common objects of consciousness and sedimented layers of experience. There would be some technical issues to overcome with implementation of the above functionalities, especially with regard to control of voice inflection, but the majority of the framework was already in place with the Articulated Head.

9.2.9 Senses – related
This sub section is directly linked with Theme 9: Senses – Related.

Much has already been said regarding simulation of the senses, suffice to say that convincing the audience that the robot can see and hear is paramount to enhancing legitimization of the intentional agents performance and the human perception of its presence as an existential being active in the interaction.
9.2.10 Memory related
This sub section is directly linked with Theme 10: Memory Related.

Crosspollination of memory in the themes above has already constituted its integration in the re-plating process with reference to the acquisition and storage of Users preferences, and in relation to access to knowledge bases for display and/or inclusion in conversational foci.

9.2.11 Outliers
This sub section is directly linked with Theme 11: Outliers.

The outlier theme only left two subjects of interest in analysis. The Sci-Fi theme is already integrated in the re-plating process with the inclusion of a database library related to music, film, TV and game preferences. The other subject of interest related to a research participant feeling that a third party was watching them during their interaction. Indeed a third party was watching them. To address this issue one simply conceals any cameras and other sensory apparatus within the exhibit structure so that they are essentially invisible and their functionality does not distract the human in interaction.

Then, after making sure that the exhibits functionalities perform successfully with autonomy, one removes the third party altogether and lets the exhibit speak for itself!

9.2.12 The Auditory Visual Interactive Environment

The Design Refinement
The experimental projects and exhibit design refinement detailed below integrates recommendations from the themes analysis under the headings in section 9.2 above, and stems directly from the explanation of research participant reporting imparted in Theme 12: Auditory Visual Interactive Environment. The design refinement is a plan for the creation of a microcosmic-ecosystem and interactive environment for the exhibit, aimed at giving an audience a more immersive and enveloping auditory visual
experience and expediting more encouraging participant reportage of the enhancement of engagement in human ↔ machine Interaction with this, and similar types of interactive installations. The description of the design that follows assumes implementation in the Powerhouse Museum but could easily be applied to many exhibition environments.

Some people might ask, what does all this have to do with research? Research is by definition the systematic investigation into, and study of materials and sources in order to establish facts and reach new conclusions: this investigation systematically studied and observed human interactions with the Articulated Head and has identified that humans like it when the robot pays attention to them. Humans have indicated in the research data that they want to use speech modality in communication. When the robot demonstrates a level of agency that is pre-conceptually dismissed by the human as being beyond its capabilities, the human momentarily attributes the robot with real human capabilities and consciousness, and in that instance has been shown to instinctively switch to a speech modality. (see 7.3.4 ‘pink shirt’ & 7.3.6 ‘young man’). This phenomenon’s relationship is illustrated on the following page.
Whilst this phenomenon does provide a way to demonstrate a clear peak in human engagement with the robot, which meets with the big question; How can interaction between humans and machines be improved? Apprehension of this new speech modality of communication needs initiating at the speech recognition apprehension point indicated in figure 9-2 above. Sustainability of the enhanced engagement of the audience in this interaction would require that the robot could demonstrate the capability of speech recognition. Beyond this, the sustainability of the interaction and conversation would rely heavily on the diversity and coherency of the robots.
‘database of conversational possibilities’ and the number of attributes about its audience and the immediate environment that it could bring into conversation.

With the antecedent of the Freakish Peak Phenomenon hypothesis, and in an attempt to address many of the drawbacks of the Articulated Heads exhibit design and performance identified through the data analysis in sections 7 & 9 of this document, a design refinement was established. Figure 9-3 presents a rudimentary schematic depiction of a proposed environment that would assist in promoting human-machine interaction. The schematic in Figure 9-3 forms a useful indicator of the proposed shape and scale of the redesign that is then extended by the subsequent Figures 9.4 & 9.5, both of which provide further insight into key aspects of the design.

**A refined experimental project design**

![Diagram](attachment:grapher_file.png)

*Figure 9-3 Interactive environment design refinement & layout.*

**Grapher File (Mac Only)**

i. The round black circles indicate loudspeaker positions
ii. The red objects in the middle indicate the humans position in the installation
iii. The yellow object indicate the Articulated Head's position
iv. The concave shape indicates a large surrounding projection screen which would be acoustically porous.

\[
\begin{bmatrix}
x \\
y \\
z
\end{bmatrix} = \begin{bmatrix}
sin t \\
sin u \\
\cos t
\end{bmatrix}, t = .9\ldots 0.8\pi, u = 1\ldots 2\pi, b = 4
\]
The \(x, y, z\) formula above was entered into Apple’s proprietary Grapheer software application. The formula generates the concave grid shape in Figure 9-3 above. The other objects present in Figure 9-3 above are created and positioned with other formulae not shown here. The colours and rendering of the concave screen are a function of the Grapheer software application. By manipulating the value of \(t\) in the formula one can close down the size of the circular cut out in the base of the concave screen, similarly, if one changes the value of 0.8 figure just before \(\pi\) in the formula one can close down the size of the circular cut out in the ceiling of the shape. You can also manipulate the width of the entrance gap in the shape by manipulating the value 2 preceding the second \(\pi\) character in the formula. The formula below Figure 9-3 above, which creates the concave screen shape, can be unpacked and shown as follows:

\[
x = 0 + 0.7(\sin [t] \cdot \sin [u]) \text{ where } t=0.9, 0.8; u=\pi, 2.\pi; b=4
\]
\[
y = 0 + 0.7(\sin [t] \cdot \cos [u]) \text{ where } t=0.9, 0.8; u=\pi, 2.\pi; b=4
\]
\[
z = 0 + 0.7(\cos [t] \cdot \cos [b]) \text{ where } t=0.9, 0.8; u=\pi, 2.\pi; b=4
\]

The black dots in Figure 9-3 above indicate prospective speaker mounting positions. The yellow objects in the middle indicate the robot and the red indicate an interacting audience. No kiosk is indicated in the design because the Video Cued Recall interview and Interpretive Phenomenological Analysis data clearly indicate participant problems with typing, and that the modality of information exchange appeared to be channeling the interacting audience’s focus between the keyboard and the head. The majority of the interacting audience’s attention time was predictably spent looking at the kiosk keyboard. User input data clearly showed a very high rate of spelling mistakes and incomplete sentences present in typed kiosk keyboard input. Therefore voice recognition software and a purpose-designed microphone array system utilizing well known standard audio compression techniques including ducking, gating and/or switching would be expected to produce a comparable, if not better standard of system text input whilst liberating the participant’s gaze and opportunities for visual meanderings - hence immediately rendering any
projections more tangible. The diagram in Figure 9-3 is not made specifically to scale, it is just for illustrative purposes to help elucidate the idea of a redesigned, more intimate and immersive interactive enclosure for the exhibit. The exact shape of the concave screen, speaker-mounting positions and placement of the robot and audience would depend on a number of factors, such as projection options and dispersion field angle specifications of the speakers.

There is a multitude of ways in which these ideas could be incorporated into any similar large-scale installation or exhibit, with variations to the scale, dimensions and layout entirely dependent on the scenario in question, therefore the details in figure 9-3, 9-4 & 9-5 are simply meant to convey the basic concepts, which they do so adequately.

Whilst it could be argued that Figure 9-3 could have been presented with more sophistication by the use of higher-end CAD software or 3DS Max, Maya etc..., and that high-end software depictions would look good for any hypothetical prospective client, which indeed they would, - without the specific needs of any prospective client and the exhibition/installation scenario in question, any such depictions would be somewhat superfluous making little difference to elucidation of the ideas already conveyed.

Consideration of health and safety would also affect the design. Previously a 1.83-3m glass barrier separated the audience from the robot. In this new design, it is envisaged that both the audience and the Articulated Head would be on raised platforms. The robot would be separated from the audience by a moat with a raised outer lip to deter any further human movement towards the robot. The moat gully and some of the wall rising out towards the human would be connected with pressure sensitive trip switches for disabling the robot should any foolish individual choose to ignore the obvious warning signs sent to them by the raised moat lip and labeling saying “do not cross this line”. The moat would completely surround the Articulated Head and the trip switches would disable the robot before any person could enter its navigational space. To be doubly sure of health and
safety, a laser and photosensitive resistor trip switch arrangement would also protect the navigational space as a failsafe contingency plan.

![Figure 9-4 Enclosure Cross Section View](image)

Design refinement materials and layout need further consideration to take account of expense and other issues such as disabled access and durability for example. However, the diagram 9-5 below shows a rough (not to scale) plan view of the area shown in Figure 9-4 above.
This arrangement addresses another issue present in the wider Video Cued Recall interview and Interpretive Phenomenological Analysis data; that of the audience reporting that the glass barrier reduced intimacy of the engagement and made the Articulated Head more like a caged animal.

Further key aspects of this design that address concerns identified through themes analyses are as follows:

**The Spatial Auditory System**

All eight loudspeakers in the array are mounted in a relatively uncompromising position so a direct uncoloured sound image is communicated to the audiences’ ears. The problem of the audience receiving a sonic image comprised of substantially reflected sound from surface materials, is dramatically reduced as 1) no acoustically reflective material is present between the participant and the speakers and 2) the speakers are mounted so that their projection field propagates a significant proportion of their output, out from under or over the concave screen surrounding the audience. The upper four-speaker array prorogation will be reflected off the floor and out into the museum.
The same is true of the lower four-speaker array except that prorogated waves will take longer to reach the museum ceiling. The screen arrangement also means that less reflected sound would enter back into the audience arena as the screen effectively acts as an acoustic shield surrounding the audience. Furthermore the concave screen reduces the effects of ambient noise from entering the audience’s arena, therefore making a clear and tangible contribution to raising the effective headroom available within the audio system. This in turn allows for much clearer balanced spatial audio definition within the immediate space surrounding the exhibit. Accurate psychoacoustic placement of virtual sound sources would be achievable.

The speaker arrangement allows for a controlled and consistent exhibit sound pressure level contribution to the wider museum environment, whilst retaining the positive aspects of the robots singing voice with other sounds from the exhibit attracting a wider audience, a feature that was indicated as being desirable in extract L in Appendix 3; *The Auditory Visual References Table*. The speaker array also allows for experimentation with a range of spatialisation techniques including Vector Based Amplitude Panning (VBAP), Vector Distance Panning (VDP), DBAP and Ambisonics, Virtual Microphone Technique (ViMiC), Wave Field Synthesis (WFS) and Ambisonics Equivalent Panning (AEP). [Lossius, personal communication]. The projection screen also provides partial isolation from ambient noise, which would otherwise hinder accurate voice recognition. Spatial audio presented from within the partially isolated acoustic environment could easily be cancelled from voice recognition signals by phase reversal and addition of the spatial auditory source signal to the voice recognition microphone signal hence dramatically increasing system intelligibility whilst still retaining the benefits of spatial auditory cues. "High levels of envelopment and room impression created by surround sound have been shown, through auditory research, to be the descriptive features most closely related to positive emotional responses" (Rumsey, 2001, p. 46), hence more positive experience should be reported by patrons of the exhibit simply because of the increased spatial auditory envelopment created by this design refinement, regardless of any other features of the design that have a positive effect upon engagement.
The Thought Cloud Projections

The concept for the Thought Clouds draws from an analogy with the famous rabbit in a hat trick - you know that the rabbit cannot really have been in the hat as you checked it out for yourself at the beginning of the trick. However, you believe you just saw the rabbit come out of the hat - and seeing is believing. Enhancement of the illusion of the Articulated Head thinking, brought about in the minds of the interacting audience by the Thought Cloud projections, relies on allowing the audiences visual experience to meander freely with their thoughts and imagination. Extract G in Appendix 2: The Auditory Visual References Table suggests that although relations between the text in the projections and some connection with the user input can be made, it was not clear enough and probably needed to be made more explicit. Stelarc showed a stronger interest in the more simple Director projections, which do have more explicit connections to user input and ECA output as the exact words are extracted for display, rather than related words. Perhaps some experimentation with a mixture of both was called for so extension of the projection capabilities was explored with the use of Vizzie objects (Cycling, 2010) and Jamoma (“Jamoma,” 2012) within the Max/MSP (Cycling 74, 2012) programming environment.

We know that participant processing of auditory visual as oppose to auditory only information, increases cognitive load and this has been shown to reflect in increased participant reaction times (Stevens, Gibert, Leung, & Zhang, 2011). It is clear that the audience needs to be able to make a tangible visual connection with the projections for a sufficiently long duration of time, in order to view, read, digest and identify connections between the displayed words, input text, Embodied Conversational Agent output and their own thought processes. To evoke attribution of the ability of conscious thought to the Articulated Head, the audience’s imagination must dream up links between the projections and the thread of conversation taking place. Fortunately the cognitive processing time required for dreaming up these links is still thought to be relatively short in comparison to the time it took to type a sentence on the kiosk keyboard, implementation of voice recognition in our design refinement, delivered along with emancipation of gaze,
dramatically increases the available time for auditory visual processing during interaction. This trick of the imagination can be very powerful and convincing but it does require an immersive interactive environment with a good degree of visual freedom afforded to the audience, in order for the trick to work. With the above points in mind, the concave visual screens design refinement surrounding the audience and the recommendation for removal of a kiosk and keyboard, in favour of voice recognition for input was conceived. Removal of the kiosk screen, which displays typed input from the keyboard and THAMBS is also desirable, because it is another visual distraction for the audience and since the audience do not need to look at what they have typed anymore with voice recognition, it would no longer be a necessity.

Options for high absorption coefficient materials to be used in the construction of the display screen and options for the display of the Thought Clouds from a concave screen needs more investigation. Possibilities exist for projection onto the screen or display emanating from the screen. However, ("Surround Vision - Projections on any type of surface," 2007) provides a video with examples of options for projecting onto curved screens and ("Command Australia - GlassVu Carbon Projection Film," 2012) provides one example of projection screen material. Another consideration, and possible visual enhancement to the exhibit, which would also help to attract an audience from within the wider museum environment, is the idea that the display screen could be visually active on both sides, allowing the exhibits display to be visible on the inside and outside of the exhibits screen.

The blueprint of recommendations above is effectively the re-platting of the separated strands of the aforementioned knot into a new considered and more conducive interactive environment designed to promote and sustain the magic of the interaction for the human user for long enough to present a convincing, powerful, immersive and enveloping interactive experience without encumbering the project team and producers with unmanageable complications represented by issues such as combinatorial explosion.
It would be very interesting to conduct this research study again with implementation of the new design refinement, and then do a direct comparison between the two sets of data to see how much of an improvement the design actually delivers, sadly this is unlikely to transpire.

Section 10 now concludes this thesis with a short discussion of the future of robotics performance within the context of the fusion of science and technology within the arts.
10. Future Robotics Performance Horizons

In early December 2011 the Thinking Systems Frontiers Conference was held in the Powerhouse Museum. This conference marked the end of the Thinking Systems five-year initiative. The conference content was billed as Intelligent Machines, Robots, Human Computer Interaction and the Science – Arts Nexus. This conference also marked the end of the Australian Research Council grant application related to the Thinking Head; E-Appendix 2: from talking heads to thinking heads. This thesis marks one of the last outcomes from that original grant.

10.1 The Fusion of Science and Technology within the Arts

Science and technology have a symbiotic relationship and their influence can be seen in numerous performance situations within the arts, whether pre-programmed and interactive, or live in nature. Examples are extant across all sectors and genres of the entertainment industries, from the projections and spatial audio used in live musical and dance performance contexts to interactive audiovisual installations present in museums and 3D animations used in film, games design and television.

Advancements in electronics in particular in recent years has driven the evolution from analogue to digital technology, providing much faster, instantly transportable new ways to store, retrieve and manipulate performance data. These technological developments allow far greater cross-fertilization and homogenisation of cultural influences in modern artwork than was previously possible in our analogue world. Indeed, the speed of data transfer has enabled the crossing of traditional barriers such as geographical distance, transforming the generation of many forms of artwork via collaboration over the Internet.

However, for all the fantastic advancements in science and digital technologies that have transpired in recent years, it is interesting to note that the influence on the field of robotics appears, on first impressions, to have been a somewhat blunt tool in terms of enhancing the conditioning affect
and engagement of audiences in the performance of these robots. In attendance at a robotics symposium held at the University of Sydney in December 2010, Professor Simon Penny presented a talk titled ‘60 Years of Situated Machines – Robotic Art as a site for technical and aesthetic innovation, activism and intervention’ (Penny, 2010) in which he showed a video of a wireframe robot controlled entirely by analogue components, motors and sensors, exhibited in the 1960’s. The robot responded to the movement and sounds generated by its audience to find a focus of attention. Although this robot could not speak, to all intent and purposes, it reacted to its audience in a remarkably similar way to the Articulated Head. The audience of this analogue robot interacted with it in much the same way as the Articulated Head’s audience did when they were not standing at the kiosk typing.

Whilst the digital domain allowed for the programming of an attention model and the incorporation of a conversational agent for the Articulated Head project, which could reply to communications with some success, the significant progress made in the development of digital technology and electronics in recent year has not necessarily led to parallel progress being made in the improvement of performance of robots in terms of conditioning affect and engagement of their audiences.

Nevertheless, huge advancement in robotics has taken place over the last two decades and whilst mankind is still quite a long way from the August 2011 National Geographic Magazine articles’ opening optimistic statement; “Robots are being created that can think, act and relate to humans. Are we ready?” (NGM, 2013, p. 66), mankind is making very significant inroads in the performance of robotics, especially in the area of prosthetics and bionics. A January 2010 National Geographic Magazine (NGM, 2013, p. 35) article titled “Bionics” details recent headway in prosthetics and bionics. Prosthetic limbs, which can function and feel like a real limb and are electronically connected to the brain for control are a reality. It is now possible to give people who are blind new vision by electrode arrays placed on the retina and connected to the brain.
“Can robots be constructed that possess a conscience, arguably the most human attribute?” (NGM, 2013, p. 84).

With regard to this question, the answer is probably unlikely to be answered in the near or perhaps even distant future; a robot can appear to possess consciousness, but only by performance and illusion. To possess conscience is to have consciousness and self-awareness, to be able to critically evaluate one’s own positive and negative influences upon the existence of all that surrounds you in the spatio-temporal environment, including the consciousness and conscience of other living entities. Conscience is learned by reference to one’s own layers of experience and may also be an innate feature of the design of the biological brain. Some humans claim to have no conscience at all, but this observation usually turns out to be untrue. To endow a robot with a conscience would currently require the transplant of a biological brain.

There are many robotic projects taking place globally. For example British scientists and a consortium of European universities are working on the Emote project (Castellano, 2012), a robot that can teach and respond to children’s moods. Cambridge University UK is working on an emotional robot called Zoe (“Face of the Future Rears its Head,” 2013).

Robotics has a huge amount to contribute to mankind and whilst real walking, talking, thinking, learning, emotional robots with whom humans can relate to, and build a relationships with, are still the realms of imagination and artistic portrayal in films – it is probably only a matter of time before science and technology catch up with these portrayals.

**Conclusion**

The number one best selling box office feature film hit of all time, released in 2009 was titled Avatar (Cameron, 2009). The film raised a gross income of nearly 3 billion dollars. The film story includes a disabled soldier who is given new limbs and a new life in the body of an avatar. The depiction is interesting because the avatar in the film is a complete three-dimensional
being and the soldier can enter and control this body in a real science fictional world.

Avatars do represent opportunities for entering virtual worlds but the progress with prosthetics and bionics will probably come to meet the portrayal of the avatar in the film, whereby human brains can essentially be situated in, and control robotic bodies where the robotic body supplies the brain with the nourishment for existence. Whilst this is just a speculative view of future developments rather than research facts, this scenario would of course represent emancipation from the constraints of serious physical disabilities for some otherwise active human brains, therefore potentially making a significant contribution to their quality of life.

There is no sign of a slowdown in the research and development of electronics and digital technologies. Emergent technologies spring up at regular and high frequency intervals. Companies such as Apple are investing large sums of money into speech recognition technology, and there is no sign or shortage of interest, or money being directed towards robotics research and development projects. Avatars, agency and performance have a central role to play in the pursuit of these new robotic horizons; therefore I think we can all look forward to some very interesting developments over the next few decades.
11. Bibliography

11.1 References.


12. Appendices (Hard Copies)

12.1 Document appendices
12.1.1 Appendix 1 – Plan View Diagram of the Exhibit
12.1.2 Appendix 2 – The Auditory Visual References Table
12.1.3 Appendix 3 The Expenses Table

12.2 Electronic Appendices
12.2.1 Appendix 1 - An Introduction to Head 0+
12.2.2 Appendix 2 - From Talking Heads to Thinking Heads
12.2.3 Appendix 3 - Articulated Head – System Quick Start Guide
12.2.4 Appendix 4 - THAMBS documentation
12.2.5 Appendix 5 – Participant Information Pack
12.2.6 Appendix 6 – Participant Consent Form
12.2.7 Appendix 7 – Participant Release Form
12.2.8 Appendix 8 – Participant Questionnaire
12.2.9 Appendix 9 – Ethics Amendment Approval
12.2.10 Appendix 10 - Road Traffic Accident Investigative Scenario
12.2.11 Appendix 11 - Creative New Media Additions Documentation
12.2.12 Appendix 12 – Slideshow of Powerhouse Exhibit Construction
12.2.13 Appendix 13 – Top 100 User Input String Words
12.2.14 Appendix 14 – Top 100 Chatbot Output String Words
12.2.15 Appendix 15 - Favourite - April2012-Results Preview
12.2.16 Appendix 16 - User Input String Phrase Table
12.2.17 Appendix 17 - Chatbot Output String Phrase Table
12.2.18 Appendix 18 – Example 1
12.2.19 Appendix 19 – Example 2
12.2.20 Appendix 20 – Example 3
12.2.21 Appendix 21 – Example 4
12.2.22 Appendix 22 - Chatbot’s top 1000 words
12.2.23 Appendix 23 – Example 5
12.2.24 Appendix 24 - Top 1000 User Input String Words
12.2.25 Appendix 25 - Top 10,000 User Input String Words
12.2.26 Appendix 26 - Top 10,000 Chatbot Output String Words
12.2.27 Appendix 27 - Example 6
12.2.28 Appendix 28 - See
12.2.29 Appendix 29 – Example 7
12.2.30 Appendix 30 – Example 8
12.2.31 Appendix 31 - Keyboard Evidence
12.2.32 Appendix 32 – Example 9
Appendix 2

Auditory Visual References.

In the table below; Column ‘S’ provides an alphabetical letter for each section - so that that letter elsewhere in this paper can reference the extract. Column ‘AV’ indicates the active AV conditions present in the interaction that each extract refers to. Combinations of the AV conditions are diverse so a key for clarification is provided below the table. Extract dialogue is imparted in the right column of the table and is row-separated to indicate different sections of experiential reflection.

<table>
<thead>
<tr>
<th>S</th>
<th>AV</th>
<th>VCR Interview and IPA Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>3, 5, 6, 10</td>
<td>There were no references to the auditory or visual environmental additions brought up by the participant as a result of the researchers targeted questioning.</td>
</tr>
<tr>
<td>B</td>
<td>2, 3, 5, 6, 7, 10, 11</td>
<td>Participant did notice the projections and comments that he could see that the text was related to conversation but they were only noted later in the interaction. Participant says that they did not notice when the projections were switched during the interaction.</td>
</tr>
<tr>
<td>C</td>
<td>4, 6, 10</td>
<td>The interviewer asks specific questions about the spatial audio and the projections. Participant comments that they did not really notice anything about either the audio or the projections.</td>
</tr>
<tr>
<td>D</td>
<td>1, 4, 6, 7, 10</td>
<td>Interviewer questions the participant on what she noticed about the physical environment. The participant appears not to have noticed anything about the physical environment and says that she was focused on finding out what the robot could do.</td>
</tr>
<tr>
<td>E</td>
<td>1, 4, 5, 6, 7, 10</td>
<td>Interviewer questions the participant about the influence of the projections during the interaction and the participant confirms that he did not really notice them; he was focused on the head itself.</td>
</tr>
<tr>
<td>F</td>
<td>1, 4, 5, 6, 7, 10</td>
<td>Spatial audio is noticed as creating a more intimate environment for the participant to converse in, the participant refers to it as an “audio bubble” noted as being subconscious some of the time.</td>
</tr>
<tr>
<td>G</td>
<td>1, 4, 5, 6, 7, 10</td>
<td>Participant explains that although he was aware that the projections were there, he did not really pay any attention to them as he was focused on the interaction, on the keyboard and face.</td>
</tr>
<tr>
<td>H</td>
<td>1, 4, 5, 6, 10</td>
<td>She was a bit disappointed that she didn’t get it clearly (the meaning of the projections) but she liked it on the screen.</td>
</tr>
<tr>
<td>I</td>
<td>1, 3, 5, 6, 10</td>
<td>The participant comments on the projections “so I didn’t really notice it at first until I was actually trying to get the robot to follow me, that’s when I noticed and you’ll see on the video, I kind of looked at it for a while, I noticed that when I typed, it kind of matched some of the words or ideas from the words perhaps, and then it felt like it would throw out really strange words like ‘parasite’ - or something like that is the word that it felt like if threw out, and I didn’t pay much attention to that afterwards as I wasn’t sure what it was doing - so I felt like it had something to do with what it was processing or something like that but I wasn’t really sure so I just sort of left it”.</td>
</tr>
<tr>
<td>J</td>
<td>1, 3, 5, 6, 10</td>
<td>The interviewee was asked, “did you notice anything about the spatial auditory environment?” to which the participant replied “no not really” but then elaborated in the next coded section... Participant elaborates on the spatial auditory environments “so, when you first mentioned to me the sound environment, I didn't...”</td>
</tr>
</tbody>
</table>
### Appendix 3

<table>
<thead>
<tr>
<th>Description</th>
<th>Category</th>
<th>Purpose</th>
<th>Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 * 8/8 MOTU 896mrk3 Firewire Audio Interfaces interface</td>
<td>Spatial Audio</td>
<td>Needs good ventilation</td>
<td>$2,800.00</td>
</tr>
<tr>
<td>8 * Genlec 8020B – 649$AUD each</td>
<td>Spatial Audio</td>
<td>Needs good ventilation</td>
<td>$5,192.00</td>
</tr>
<tr>
<td>Firewire + Video Graphics card for PC</td>
<td>Projections &amp; Spatial audio</td>
<td>PCIe</td>
<td>$603.00</td>
</tr>
<tr>
<td>PC (DH)</td>
<td>Head</td>
<td>Non project utilisation</td>
<td>$2,500.00</td>
</tr>
<tr>
<td>Cables &amp; Misc</td>
<td>Projections &amp; Spatial audio + Head</td>
<td>Various for installation</td>
<td>$500.00</td>
</tr>
<tr>
<td>1 VM Fusion License</td>
<td>Projections &amp; Spatial audio</td>
<td>Simultaneous Windows + Mac OS</td>
<td>$60.00</td>
</tr>
<tr>
<td>Max/MSP license for Windows</td>
<td>Projections &amp; Spatial audio</td>
<td>Programming of projects</td>
<td>$250.00</td>
</tr>
<tr>
<td>KVM switch</td>
<td>Projections</td>
<td>Screen space expander</td>
<td>$200.00</td>
</tr>
<tr>
<td>2 * Monitor screens</td>
<td>Projections &amp; Spatial audio</td>
<td>1 for each computer</td>
<td>$200.00</td>
</tr>
<tr>
<td>WIFI Relay Substation Nodes</td>
<td>Projections &amp; Spatial audio</td>
<td>For remote control</td>
<td>$1,300.00</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td>$13,605.00</td>
</tr>
</tbody>
</table>

*Table 12-1 – Cost Implications*