Flexible photo retrieval (FlexPhoReS) : a prototype for multimodel personal digital photo retrieval

This item was submitted to Loughborough University's Institutional Repository by the/an author.

Additional Information:

- A Doctoral Thesis. Submitted in partial fulfilment of the requirements for the award of Doctor of Philosophy of Loughborough University

Metadata Record: https://dspace.lboro.ac.uk/2134/12924

Publisher: © Nor Azman Ismail

Please cite the published version.
This item was submitted to Loughborough University as a PhD thesis by the author and is made available in the Institutional Repository (https://dspace.lboro.ac.uk/) under the following Creative Commons Licence conditions.

[Creative Commons licence]

For the full text of this licence, please go to:
http://creativecommons.org/licenses/by-nc-nd/2.5/
Please note that fines are charged on ALL overdue items.

FOR REFERENCE ONLY
Flexible Photo Retrieval (FlexPhoReS): a prototype for multimodal personal digital photo retrieval

by

Nor Azman Ismail

A Doctoral thesis submitted in partial fulfilment of the requirements for the award of Doctor of Philosophy of Loughborough University

Research School of Informatics,
Department of Information Science,
Loughborough University,
Loughborough, Leicestershire
LE11 3TU United Kingdom

© Nor Azman Ismail 2007
Abstract

Digital photo technology is developing rapidly and is motivating more people to have large personal collections of digital photos. However, effective and fast retrieval of digital photos is not always easy, especially when the collections grow into thousands. World Wide Web (WWW) is one of the platforms that allows digital photo users to publish a collection of photos in a centralised and organised way. Users typically find their photos by searching or browsing using a keyboard and mouse. Also in development at the moment are alternative user interfaces such as graphical user interfaces with speech (S/GUI) and other multimodal user interfaces which offer more flexibility to users. The aim of this research was to design and evaluate a flexible user interface for a web based personal digital photo retrieval system. A model of a flexible photo retrieval system (FlexPhoReS) was developed based on a review of the literature and a small-scale user study. A prototype, based on the model, was built using MATLAB and WWW technology. FlexPhoReS is a web based personal digital photo retrieval prototype that enables digital photo users to accomplish photo retrieval tasks (browsing, keyword and visual example searching (CBIR)) using either mouse and keyboard input modalities or mouse and speech input modalities. An evaluation with 20 digital photo users was conducted using usability testing methods. The result showed that there was a significant difference in search performance between using mouse and keyboard input modalities and using mouse and speech input modalities. On average, the reduction in search performance time due to using mouse and speech input modalities was 37.31%. Participants were also significantly more satisfied with mouse and speech input modalities than with mouse and keyboard input modalities although they felt that both were complementary. This research demonstrated that the prototype was successful in providing a flexible model of the photo retrieval process by offering alternative input modalities through a multimodal user interface in the World Wide Web environment.
To my beloved parents who passed away,
Allahyarhamah Miah Ismail &
Allahyarham Ismail Othman

To my Wife,
Nor Azah Hussin

To my Children,
Syafiqah, Asyraf and Hazim
Acknowledgements

First of all, I would like to thank God for everything, including the ability to learn, the pleasure of sharing it and the wisdom of appreciating it.

I would also like to thank my supervisor, Dr. Ann O'Brien for her support, encouragement and enthusiastic discussions. I have been very fortunate to have the opportunity to work with her and I can never thank her adequately for her help on both academic and non-academic matters.

Thanks are also due to Dr. Anne Morris, my Director of Research, for making many valuable suggestions and to those academic related/support staff who provided invaluable help, from the early stages of my research until its completion.

I want to thank my beloved parents who have passed away: Allahyarhamah Hajjah Miah Ismail and Allahyarham Haji Ismail Othman, my wife Nor Azah Hussin, my children Nur Syafiqah, Muhammad Asyraf and Muhammad Hazim, my beloved father in-law who passed away Allahyarham Haji Hussin Haji Zakaria; my mother in-law Hajjah Ramlah Haji Awang and other family members for the love and support they have provided throughout the three years of my research, without which I would not have reached my current goals and aspirations.

A special vote of thanks is due to all of my friends, fellow research students and staff in the Research School of Informatics, Department of Information Science at Loughborough University. All my experiences and interactions with these people have shaped my work and helped make my graduate experience worthwhile and unforgettable. Finally, I would like to thank to the Ministry Of Science, Technology & The Environment, Malaysia (MOSTE) for sponsoring my research, and my home institution, University of Technology Malaysia, for granting me leave to carry out this research.
Table of Contents

Abstract
Acknowledgement
Table of contents
List of figures
List of tables
List of acronyms

CHAPTER 1: INTRODUCTION

1.0 Introduction 1
1.1 Motivation 1
1.2 Research focus 5
1.3 Research aim and objectives 7
1.4 Scope of the study 7
1.5 Significance of the study 9
1.6 Structure of the thesis 10

CHAPTER 2: LITERATURE REVIEW

2.0 Introduction 13
2.1 Image retrieval evolution 14
2.2 User studies in image collection 19
   2.2.1 Image users 19
   2.2.2 Studies of user queries in image collections 20
   2.2.3 Personal digital photos 26
2.3 User interface support for digital photo retrieval 32
   2.3.1 Photo annotation 33
   2.3.2 User visual access modes 36
   2.3.3 Retrieval strategies 37
   2.3.4 Studies related to digital photo applications 41
CHAPTER 3: METHODOLOGY

3.0 Introduction 72

3.1 Research methodology 72
  3.1.1 System development lifecycle models 73

3.2 Research design 77
  3.2.1 Stage 1: Pre-development needs evaluation exercise 78
  3.2.2 Stage 2: User Task description 80
  3.2.3 Low fidelity prototyping development and evaluation 80
  3.2.4 Stage 3: FlexPhoReS evaluation 81
    3.2.4.1 Pilot test 81
    3.2.4.2 Participants 82
    3.2.4.3 Experimental procedures 82
    3.2.4.4 The tasks 84

3.2.5 Data collection 86
  3.2.5.1 Computer screen and speech recordings 86
  3.2.5.2 Questionnaire for suitability, flexibility and user subjective satisfaction 86

3.2.6 Variables studied 87
  3.2.6.1 The independent variables 87
  3.2.6.2 The dependent variables 87
3.2.7 Data analysis techniques
  3.2.7.1 Paired-sample t-test
  3.2.7.2 Wilcoxon Signed Rank test
3.3 Assumptions in the research
3.4 Summary

CHAPTER 4: MODELLING FlexPhoReS

4.0 Introduction
4.1 Proposed model design
  4.1.1 The proposed model
    4.1.1.1 Small-scale user study outcomes
4.2 Task description
4.3 Process model
4.4 Summary

CHAPTER 5: FlexPhoReS PROTOTYPE

5.0 Introduction
5.1 Enabling technology
  5.1.1 The World Wide Web
  5.1.2 Speech technologies and SALT
  5.1.3 MATLAB
  5.1.4 Active Server Pages (ASP)
  5.1.5 JavaScript
  5.1.6 Enabling technology summary
5.2 Designing the prototype
5.3 FlexPhoReS architecture
5.4 FlexPhoReS components
  5.4.1 User Profiles
  5.4.2 Repository and access control
  5.4.3 Multimodal interaction
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.4.4</td>
<td>Retrieval strategies</td>
<td>130</td>
</tr>
<tr>
<td>5.4.5</td>
<td>Photo features</td>
<td>134</td>
</tr>
<tr>
<td>5.4.6</td>
<td>Help</td>
<td>137</td>
</tr>
<tr>
<td>5.5</td>
<td>Limitations of the prototype</td>
<td>138</td>
</tr>
<tr>
<td>5.6</td>
<td>Summary</td>
<td>138</td>
</tr>
</tbody>
</table>

**CHAPTER 6: USER EVALUATION OF THE PROTOTYPE**

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.0</td>
<td>Introduction</td>
<td>139</td>
</tr>
<tr>
<td>6.1</td>
<td>User background, gender and experience</td>
<td>140</td>
</tr>
<tr>
<td>6.2</td>
<td>Data analysis</td>
<td>141</td>
</tr>
<tr>
<td>6.3</td>
<td>Test hypotheses</td>
<td>142</td>
</tr>
<tr>
<td>6.4</td>
<td>Results of the user evaluation</td>
<td>143</td>
</tr>
<tr>
<td>6.4.1</td>
<td>Task completion time</td>
<td>143</td>
</tr>
<tr>
<td>6.4.2</td>
<td>Subjective satisfaction with the prototype</td>
<td>150</td>
</tr>
<tr>
<td>6.4.3</td>
<td>Acceptability of the prototype</td>
<td>155</td>
</tr>
<tr>
<td>6.5</td>
<td>Tests for statistical significance</td>
<td>157</td>
</tr>
<tr>
<td>6.5.1</td>
<td>Comparison between using mouse and keyboard input modalities and using mouse and speech input modalities in search performance</td>
<td>158</td>
</tr>
<tr>
<td>6.5.2</td>
<td>Comparison between using mouse and keyboard input modalities and using mouse and speech input modalities for subjective satisfaction</td>
<td>159</td>
</tr>
<tr>
<td>6.5.3</td>
<td>Comparison between using mouse and keyboard input modalities and using mouse and speech input modalities in photo browsing search performance</td>
<td>159</td>
</tr>
<tr>
<td>6.5.4</td>
<td>Comparison between using mouse and keyboard input modalities and using mouse and speech input modalities in keyword searching search performance</td>
<td>160</td>
</tr>
<tr>
<td>6.5.5</td>
<td>Comparison between using mouse and keyboard input modalities and using mouse and speech input modalities in visual example searching</td>
<td>161</td>
</tr>
</tbody>
</table>
6.6 Participants with significant experience in using online retrieval system
6.7 Summary

CHAPTER 7: RESEARCH FINDINGS AND DISCUSSION

7.0 Introduction and research findings
7.1 Phase One
7.2 Phase Two
7.3 Phase Three
7.3.1 Comparison between different input modalities in search performance and subjective satisfaction among all participants
7.3.2 Acceptability of FlexPhoReS
7.3.3 Discussion of Phase Three findings
7.4 Summary

CHAPTER 8: CONCLUSIONS AND FURTHER WORK

8.0 Introduction
8.1 Research summary and achievements
8.2 Research conclusions
8.2.1 Literature background
8.2.2 The aim
8.2.3 FlexPhoReS model
8.2.4 FlexPhoReS prototype
8.2.5 Research findings and evaluations
8.3 Research implications
8.3.1 Implications for knowledge
8.3.2 Implications for system developers
8.3.3 Implications for digital photo searchers
8.4 Limitations and suggestions for further work
8.5 Concluding remarks
APPENDICES

Appendix 1: Pre-development small-scale user study structured interview questionnaires
Appendix 2: Interview questionnaire of low fidelity informal evaluation
Appendix 3: Prototype sketch on A1 paper
Appendix 4: Low fidelity prototype (Microsoft PowerPoint version)
Appendix 5: Recruitment questionnaire
Appendix 6: Experiment procedure
Appendix 7: Questionnaire for user subjective satisfaction and acceptability
Appendix 8: Participants’ demographic characteristics
Appendix 9i to 9iv: Statistical analysis and tests for normality
Appendix 10: FlexPhoReS source codes files (CDROM attached) and services
# List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Projected number of digital photos printed</td>
<td>2</td>
</tr>
<tr>
<td>1.2</td>
<td>Research operational framework</td>
<td>9</td>
</tr>
<tr>
<td>2.1</td>
<td>Early generation of image retrieval systems</td>
<td>15</td>
</tr>
<tr>
<td>2.2</td>
<td>New generation image retrieval systems</td>
<td>18</td>
</tr>
<tr>
<td>2.3</td>
<td>Illustration of 2-D and 3-D display</td>
<td>40</td>
</tr>
<tr>
<td>2.4</td>
<td>The W3C multimodal interface framework</td>
<td>60</td>
</tr>
<tr>
<td>3.1</td>
<td>Lifecycle model for interaction design</td>
<td>76</td>
</tr>
<tr>
<td>4.1</td>
<td>‘Use case’ diagram for the photo retrieval system</td>
<td>97</td>
</tr>
<tr>
<td>4.2</td>
<td>Process model of FlexPhoReS</td>
<td>98</td>
</tr>
<tr>
<td>5.1</td>
<td>User interacting with Web</td>
<td>102</td>
</tr>
<tr>
<td>5.2</td>
<td>Speech recognition process flow</td>
<td>104</td>
</tr>
<tr>
<td>5.3</td>
<td>Speech synthesis or text-to-speech process flow</td>
<td>104</td>
</tr>
<tr>
<td>5.4</td>
<td>SALT Architecture</td>
<td>105</td>
</tr>
<tr>
<td>5.5</td>
<td>Matlab web server configuration</td>
<td>106</td>
</tr>
<tr>
<td>5.6</td>
<td>ASP page process and displayed</td>
<td>107</td>
</tr>
<tr>
<td>5.7</td>
<td>FlexPhoReS interface flow diagram</td>
<td>110</td>
</tr>
<tr>
<td>5.8</td>
<td>Series of screens shot (in PowerPoint) of the low fidelity prototype</td>
<td>112</td>
</tr>
<tr>
<td>5.9</td>
<td>Architecture of FlexPhoReS</td>
<td>114</td>
</tr>
<tr>
<td>5.10</td>
<td>Client-Server Interaction in FlexPhoReS</td>
<td>115</td>
</tr>
<tr>
<td>5.11</td>
<td>SALT pages elements</td>
<td>116</td>
</tr>
<tr>
<td>5.12</td>
<td>Speech recognition event</td>
<td>116</td>
</tr>
<tr>
<td>5.13</td>
<td>Prompts queuing</td>
<td>117</td>
</tr>
<tr>
<td>5.14</td>
<td>Adobe Photoshop Album</td>
<td>118</td>
</tr>
<tr>
<td>5.15</td>
<td>FlexPhoReS user interface</td>
<td>120</td>
</tr>
<tr>
<td>5.16</td>
<td>FlexPhoReS user profiles</td>
<td>125</td>
</tr>
<tr>
<td>5.17</td>
<td>FlexPhoReS access control architecture</td>
<td>126</td>
</tr>
<tr>
<td>5.18</td>
<td>FlexPhoReS login page</td>
<td>126</td>
</tr>
<tr>
<td>5.19</td>
<td>FlexPhoReS starting page</td>
<td>127</td>
</tr>
<tr>
<td>5.20</td>
<td>FlexPhoReS warning message and speech prompt</td>
<td>127</td>
</tr>
<tr>
<td>5.21</td>
<td>‘Tap’ and ‘talk’ in the FlexPhoReS</td>
<td>128</td>
</tr>
<tr>
<td>5.22</td>
<td>Visible photo name on top of each photo</td>
<td>129</td>
</tr>
</tbody>
</table>
5.23 FlexPhoReS retrieval strategies flow
5.24 Browsing and navigation result and categories
5.25 Searching by keywords
5.26 Photo names
5.27 Visual example searching results with photo example.
5.28 Sample photo and its corresponding histogram
5.29 Retrieval query options and process
6.1 Frequency use of online image retrieval features
6.2 Participants’ retrieval strategies experience
6.3 Average time taken to complete each task
6.4 Average time taken to complete each tasks with input modalities
6.5 Average of subjective satisfaction when using different input modalities (all participants)
7.1 Average time taken to complete search tasks with different input modalities
7.2 Means (average scores) of subjective satisfaction for FlexPhoReS across different input modalities
7.3 Acceptability mean of all participants
7.4 Suitability of mouse and keyboard input modalities of FlexPhoReS for photo retrieval tasks
7.5 Suitability of mouse and speech input modalities of FlexPhoReS photo retrieval tasks
7.6 Suitability of both input modalities for photo retrieval tasks
7.7 Both input modalities are complementary for photo retrieval tasks
7.8 FlexPhoReS flexibility
7.9 FlexPhoReS flexibility over the World Wide Web
7.10 Graph of Percent reduction vs. Retrieval strategies
7.11 Participants’ choices of FlexPhoReS input modalities for photo retrieval tasks
<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Annotation techniques</td>
<td>35</td>
</tr>
<tr>
<td>2.2</td>
<td>When is speech input and output appropriate?</td>
<td>53</td>
</tr>
<tr>
<td>2.3</td>
<td>Relative effectiveness of low-fidelity prototype vs. high-fidelity prototypes</td>
<td>63</td>
</tr>
<tr>
<td>2.4</td>
<td>Characteristics of evaluation paradigms</td>
<td>67</td>
</tr>
<tr>
<td>3.1</td>
<td>Mouse and keyboard input modalities tasks</td>
<td>84</td>
</tr>
<tr>
<td>3.2</td>
<td>Mouse and speech input modalities tasks</td>
<td>85</td>
</tr>
<tr>
<td>5.1</td>
<td>Problems and recommendation</td>
<td>113</td>
</tr>
<tr>
<td>5.2</td>
<td>Words and their meanings for speech interaction</td>
<td>121</td>
</tr>
<tr>
<td>6.1</td>
<td>Participants task completion times</td>
<td>144</td>
</tr>
<tr>
<td>6.2</td>
<td>Means and standard deviations of task completion time for all participants (n=20)</td>
<td>145</td>
</tr>
<tr>
<td>6.3</td>
<td>Participants search tasks completion</td>
<td>148</td>
</tr>
<tr>
<td>6.4</td>
<td>Means and standard deviations of time taken to complete each search task for all participants (n=20)</td>
<td>149</td>
</tr>
<tr>
<td>6.5</td>
<td>Participants (1-10) Subjective satisfaction scores</td>
<td>152</td>
</tr>
<tr>
<td>6.6</td>
<td>Participants (11-20) Subjective satisfaction scores</td>
<td>153</td>
</tr>
<tr>
<td>6.7</td>
<td>Means of subjective satisfaction when using different input modalities</td>
<td>154</td>
</tr>
<tr>
<td>6.8</td>
<td>Acceptability of the prototype for participants (1-10)</td>
<td>156</td>
</tr>
<tr>
<td>6.9</td>
<td>Acceptability of the prototype for participants (11-20)</td>
<td>157</td>
</tr>
<tr>
<td>6.10</td>
<td>The Pair sample t-test for search performance using different input modalities</td>
<td>158</td>
</tr>
<tr>
<td>6.11</td>
<td>Wilcoxon Signed Ranks test for subjective satisfaction when using different input modalities</td>
<td>159</td>
</tr>
<tr>
<td>6.12</td>
<td>Wilcoxon Signed Ranks test for photo browsing search performance using different input modalities</td>
<td>160</td>
</tr>
<tr>
<td>6.13</td>
<td>Wilcoxon Signed Ranks test for keyword searching search performance using different input modalities</td>
<td>161</td>
</tr>
<tr>
<td>6.14</td>
<td>Pair sample t-test for visual example searching search performance using different input modalities</td>
<td>162</td>
</tr>
<tr>
<td>6.15</td>
<td>List of null hypotheses explored</td>
<td>163</td>
</tr>
</tbody>
</table>
7.1 Average time taken to complete search tasks and percentage reduction with different input modalities

7.2 Means (average scores) of subjective satisfaction for FlexPhoReS across both input modalities

7.3 Statistical significance tests for different input modalities in search performance and subjective satisfaction

7.4 Average acceptability rating and standard deviations of FlexPhoReS among all participants
<table>
<thead>
<tr>
<th>Acronym</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANOVA</td>
<td>Analysis of variance</td>
</tr>
<tr>
<td>ARIA</td>
<td>An Agent for Annotating and Retrieving Images</td>
</tr>
<tr>
<td>ASP</td>
<td>Active Server Pages</td>
</tr>
<tr>
<td>AT&amp;T</td>
<td>American Telephone and Telegraph</td>
</tr>
<tr>
<td>CASE</td>
<td>Computer-aided software engineering</td>
</tr>
<tr>
<td>CBIR</td>
<td>Content-based image retrieval</td>
</tr>
<tr>
<td>CDROM</td>
<td>Compact Disc Read Only Memory</td>
</tr>
<tr>
<td>CERN</td>
<td>European Organization for Nuclear Research</td>
</tr>
<tr>
<td>CUA</td>
<td>Common User Access</td>
</tr>
<tr>
<td>DOM</td>
<td>Document Object Model</td>
</tr>
<tr>
<td>DTMF</td>
<td>Dual tone multi frequency</td>
</tr>
<tr>
<td>EMMA</td>
<td>Extensible MultiModal Annotation Markup Language</td>
</tr>
<tr>
<td>GUI</td>
<td>Graphical user interface</td>
</tr>
<tr>
<td>HTML</td>
<td>Hypertext Mark-up Language</td>
</tr>
<tr>
<td>HTTP</td>
<td>HyperText Transfer Protocol</td>
</tr>
<tr>
<td>IBM</td>
<td>International Business Machine</td>
</tr>
<tr>
<td>IIS</td>
<td>Internet Information Services</td>
</tr>
<tr>
<td>JAD</td>
<td>Joint Application Development</td>
</tr>
<tr>
<td>LUCID</td>
<td>Logical User-Centred Interaction Design</td>
</tr>
<tr>
<td>MARS</td>
<td>Multimedia Analysis and Retrieval Systems</td>
</tr>
<tr>
<td>MATLAB</td>
<td>Matrix Laboratory</td>
</tr>
<tr>
<td>MERL</td>
<td>Mitsubishi Electric Research Laboratory</td>
</tr>
<tr>
<td>MIT</td>
<td>Massachusetts Institute of Technology</td>
</tr>
<tr>
<td>MMI</td>
<td>Multimodal Interaction</td>
</tr>
<tr>
<td>NLM</td>
<td>National Library of Medicine's</td>
</tr>
<tr>
<td>PDA</td>
<td>Personal Digital Assistants</td>
</tr>
<tr>
<td>PDH</td>
<td>Personal Digital Historian</td>
</tr>
<tr>
<td>PhotoTOC</td>
<td>Photo Table of Contents</td>
</tr>
<tr>
<td>QBE</td>
<td>Query by Example</td>
</tr>
<tr>
<td>QBIC</td>
<td>Query by Image Content</td>
</tr>
<tr>
<td>QUIS</td>
<td>Questionnaire for User Interface Satisfaction</td>
</tr>
<tr>
<td>Acronym</td>
<td>Description</td>
</tr>
<tr>
<td>---------</td>
<td>-------------</td>
</tr>
<tr>
<td>RAD</td>
<td>Rapid Application Development</td>
</tr>
<tr>
<td>SALT</td>
<td>Speech Application Language Tags</td>
</tr>
<tr>
<td>SPSS</td>
<td>Statistical Package for the Social Sciences</td>
</tr>
<tr>
<td>SQL</td>
<td>Structured Query Language</td>
</tr>
<tr>
<td>SR</td>
<td>Speech Recognition</td>
</tr>
<tr>
<td>TTS</td>
<td>Text to Speech Synthesis</td>
</tr>
<tr>
<td>UEL</td>
<td>Usability Engineering Lifecycle</td>
</tr>
<tr>
<td>URI</td>
<td>Universal Resource Identifier</td>
</tr>
<tr>
<td>W3C</td>
<td>World Wide Web Consortium</td>
</tr>
<tr>
<td>WI-FI</td>
<td>Wireless Fidelity</td>
</tr>
<tr>
<td>WWW</td>
<td>World Wide Web</td>
</tr>
<tr>
<td>XHTML</td>
<td>eXtensible HyperText Markup Language</td>
</tr>
<tr>
<td>XML</td>
<td>Extensible Mark-up Language</td>
</tr>
</tbody>
</table>
CHAPTER 1

INTRODUCTION

1.0 Introduction

This chapter presents an introduction to the research.

• Section 1.1 presents the motivation for the research.
• Section 1.2 outlines the research focus
• Section 1.3 presents the research aim and objectives
• Section 1.4 gives the scope of the study
• Section 1.5 gives the significance of the study
• Section 1.6 illustrates the structure of the thesis.

1.1 Motivation

Each of us sees this world in his or her own way and has his or her own personal reactions to it. Photography is the imaging technique that records information similar to that which we receive using our eyes (Efford 2000, p.2). Photographs are made in order to convey a certain vision or idea and most people take photographs to capture and express their feelings about people, nature and the world around them (Taylor et al. 1982, p.4). They can also create a story that was never planned to be told (Taylor et al. 1982, p.68). The word “Photography” comes from the Greek words photo, for “light,” and graph, for “drawing.” “Drawing with light” is a way of describing photography (Britannica 2005).
The process of making photographs involves recording light patterns, as reflected from objects, onto a sensitive medium through a timed exposure. The process is done through mechanical, chemical or digital devices usually known as cameras (Wikipedia 2005). The first photograph was produced in 1826 by the French inventor Nicéphore Niépce on a polished pewter plate covered with a petroleum derivative called bitumen of Judea (London and Upton 1994, p.365).

Technology advances all the time and the world of photography changes with it. In January 2004, Kodak announced that they would no longer produce traditional film cameras in North America and Europe (BBC 2004). This was interpreted as a sign of the possible end of film photography and the advance of digital photography technology (that uses an electronic sensor to record the image as binary data). In everyday life, people already have large collections of printed personal photos and the new digital photo technology helps collections grow further. An industry report from Lyra Research Inc. forecasts that the number of digital photos printed worldwide is expected to exceed 40 billion units by 2008 (Figure 1.1) (Lyra Research 2005).


**Figure 1.1:** Projected number of digital photos printed (Lyra Research 2005)
Chapter 1

Introduction

Digital photo technology has resulted in more people having large personal collections of digital photos and sharing them with others, especially with their family and friends. Web based photo galleries are one of the methods of sharing that allow photo users to publish a collection of digital photos online in a centralised and organised way (Van House et al. 2004, p.5). According to InfoTrends in 2002, the most popular use of digital photos was sharing online, either via e-mail or web sites. "We see online sharing emerging to be a category as popular as e-mail, but even stronger," says analyst Lia Shubert, and estimates 15 million people had already shared their photos online (Graham 2002).

Undoubtedly, the internet platform will play the central role in transporting personal digital photos. A recent survey conducted by InfoTrends/CAP Ventures, showed that 26% of Internet users had photos posted to an online photo service in 2004, a rise of 19% from 2003. In 2004 there were over 825 million photos stored at online photo services, and sharing and printing activities continue to increase (O'Keefe 2004). More and more collections of digital photos are now connected to the internet and this has encouraged the development of digital photo systems to provide user friendly and flexible user interfaces to retrieve the growing libraries of large personal photo collections (Shneiderman and Kang 2000, p.88).

Effective and fast retrieval of both digital photos and traditional printed photos has not always been easy, especially when the collections grow into thousands (Ismail and O'Brien 2004, p.1045). Many research projects have proposed approaches to retrieve digital images from large image collections (Flickner et al. 1995; Hiroike et al. 1999; Laaksonen et al. 1999; Chen et al. 2000; Cox et al. 2000; Nakazato et al. 2003). These approaches can be divided into two types of interactions: browsing and searching.

Personal digital photo collections are different from organisational digital image collections. The users of personal digital photos are normally the owners and are
familiar with the contents of their photos, as they are generally related to their life and memories. Some people may have developed a personal classification and social uses for their digital photos (Rodden 1999; Van House et al. 2005). When browsing, the users often look through the entire set of photos, whereas when searching, the focus is on free text (keyword) searching. There are currently a number of digital photo systems using these different retrieval approaches: commercial products, experimental systems and freeware packages. Among them are Apple iPhoto5 (Apple), Ulead iMira (Ulead 2005), Adobe Photoshop Album (Adobe), Personal Digital Historian (Shen et al. 2003), FotoFile (Kuchinsky et al. 1999), AT&T Shoebox (Mills et al. 2000), PhotoTOC (Platt et al. 2003), PhotoFinder (Kang and Shneiderman 2000) and various packages bundled with digital cameras. Some of these systems provide browsing, free text searching and even a range of limited visual content based retrieval.

In web applications, graphical user interfaces (GUIs) have become the user interface of choice. For many years, they have provided the user with a common look and feel, visual representations of data and direct control using mouse and keyboard input modalities as standard input devices. However GUIs only become possible to implement when computer hardware can produce accurate bitmap displays and can interactively manipulate accurate screen presentations to the users (Roope 1999, pp.1-2). Moving beyond mouse and keyboard, multimodal interfaces are expected to be more transparent, flexible, efficient and powerful for human-computer interaction (Oviatt 2003, p.286). Multimodal interface technologies have been applied with some success to problems in certain domains such as personal information management (e.g. Personal Digital Assistants (PDA) applications) (Deng et al. 2002). Numerous theoretical and empirical studies have investigated the potential of multimodal interfaces. The array of multimodal systems currently ranges from simulation and training applications to verification system security that will increasingly affect our lives (Oviatt et al. 2000, p.263).
Some recent studies have involved designing systems that combine either speech and pen input (Benoît and Goff 1998) or speech and lip movements (Zhai et al. 1999). One study has developed a multimodal interface for digital image retrieval. Käster and colleagues proposed a multimodal system that combines mouse, keyboard, speech and touch screen interface for standalone content based image retrieval (Käster et al. 2003). The usability experiments of the study showed that users well appreciate this multimodal interface for image retrieval. It is not clear, however, that multimodal interface approaches and technologies are fully suited for web based personal digital photo retrieval applications. Personal digital photo users in general have individual differences in performance, needs and abilities in using different modes of interaction.

The prototype in this study was developed to offer users flexibility in performing photo retrieval tasks. It allows users to use either mouse and keyboard input modalities or mouse and speech input modalities to control and retrieve digital photos through the World Wide Web environment.

1.2 Research focus

A useful photo retrieval interface study should address the problem of the user's interface requirements, together with providing a formal evaluation of the user interface itself. The proposed model, Flexible Photo Retrieval System (FlexPhoReS) provides a range of retrieval interface facilities which allows users to utilise mouse and keyboard input modalities and mouse and speech input modalities to retrieve digital photos through the World Wide Web environment. Research questions of the study are specified as follows:

1. What are the user interface properties and components suitable for a personal digital photo retrieval system?
2. What are the requirements of personal digital photo users for photo retrieval tasks?

3. How to design and demonstrate a flexible user interface (for a web based personal digital photo retrieval system) which allows users to utilise mouse and keyboard input modalities and mouse and speech input modalities to retrieve digital photos through a World Wide Web environment?

4. Will the use of mouse and speech input modalities improve user’s search performance (in time taken to complete search tasks) and receive the approval of digital photo users compared to mouse and keyboard input modalities?

This research attempts to answer the above questions. The first question required a review of image retrieval evolution, user studies of image collections and user interface support for a digital photo retrieval system. For the second research question, a data collection exercise on how people retrieve and organise their digital photos was carried out to discover user-oriented requirements. The result of the exercise was to provide supportive guidelines in developing a flexible interface for a web based personal digital photo retrieval system.

For the third and fourth research questions, an investigation was carried out on multimodal interface implementation and how mouse and speech input modalities and mouse and keyboard input modalities could be employed in a single web based personal digital photo retrieval system. A prototype was developed and evaluated to demonstrate the prototype’s ability, the user’s search performance, subjective satisfaction and acceptance of a flexible user interface for a web based personal digital photo retrieval system.
1.3 Research aim and objectives

The aim of this research was to design and evaluate a flexible user interface for a web based personal digital photo retrieval system. The aim is expressed in a set of specific objectives that systematically set the direction of the research and are listed as follows:

1. To design a model of a flexible user interface for a web based personal digital photo retrieval system.

2. To develop a prototype of a flexible user interface for a web based personal digital photo retrieval system.

3. To evaluate the prototype in order to measure users' search performance, subjective satisfaction and acceptability of the system.

1.4 Scope of the study

In line with the aim and objectives of the research, the scope of this research is as follows:

1. The development of a flexible user interface for a web based personal digital photo retrieval system (FlexPhoReS) as an outcome of a literature review and small scale user study.

2. The implementation of the prototype based on the system model of FlexPhoReS.

3. Evaluation of the prototype in order to measure users' search performance, subjective satisfaction and acceptability of the system model and prototype.
A flexible user interface is defined in this research as an interaction system that offers:

- the ability to allow digital photo users to utilise mouse and keyboard input modalities and also mouse and speech input modalities to retrieve digital photos through a World Wide Web environment.

- the ability to allow digital photo users to use photo browsing, keyword searching and visual example searching features to retrieve personal digital photos.

- the ability to use the prototype across the World Wide Web environment allowing digital photo users to retrieve digital photos at any time and any place through a web browser.

A personal digital photo collection is defined in this research as follows:

- one that allows users to retrieve their own digital photos from an authorised web-based photo repository.

- implies that users are normally familiar with the photo contents as they relate to their life and memories.
Chapter 1

Introduction

Figure 1.2: Research operational framework

Figure 1.2 illustrates the research operational framework. A system model of FlexPhoReS was constructed based on a literature review. Interview data from a user study also was used to assist the construction of the model. In the development of the prototype, the World Wide Web, Extensible Mark-up Language (XML), JavaScript, Active Server Pages (ASP), Hypertext Mark-up Language (HTML) and Matrix Laboratory (MATLAB) were used as the enabling technologies to achieve the aim of this research. To measure users' search performance, subjective satisfaction and acceptability, prototype usability testing in a controlled experiment was conducted. Further information was gathered through a questionnaire and session recordings software.

1.5 Significance of the study

Integrating multimodal interaction styles into a photo retrieval system is necessary for a user interface design to cope with a variety of users having different needs. This study attempts to fill the gap between the field of multimodal interface and the design of user interfaces for a web based personal digital photo retrieval system.
The contributions of this study to multimodal interface and digital photo retrieval research include:

1. Investigation of image retrieval evolution, user studies in image collections, user interface support for digital photo retrieval and multimodal interface through a literature review and small-scale user study to provide instructive understanding of the user needs for the digital photo retrieval process and how to implement a multimodal interface.

2. Model development of the flexible user interface that employed a multimodal interface for web based personal digital photo retrieval process.

3. Development of a prototype flexible user interface for web based personal digital photo retrieval prototype system.

4. An evaluation of the prototype system with digital photo searchers gives an understanding of the users' search performance, subjective satisfaction and acceptability of the web based personal digital photo retrieval system with flexible user interface.

5. The details of the FlexPhoReS prototype, research procedures and tasks could be use as a guideline for designing and developing systems in related area of research.

1.6 Structure of the thesis

The thesis is organised as follows:

Chapter 2 reviews the literature of previous research relevant to this work. This includes (a) image retrieval evolution; (b) user studies in image collections; (c) user interface support for digital photo retrieval; and (d) multimodal interfaces. The
review of image retrieval evolution presents image retrieval evolution from the early to later development of image retrieval systems. The review of user studies on image collections presents the image users, previous research on user queries in image collections and personal digital photo user studies. Studies on user interface support for digital photo retrieval are discussed including digital photo annotation, visual access modes, retrieval strategies and studies related to personal digital photo system development. The multimodal interface section describes multimodal user interfaces and applications; user interface engineering that presents multimodal user interface guidelines, framework for multimodal interface, prototyping and evaluation paradigms. The considerations from the literature review provide a foundation for the FlexPhoReS design and development.

Chapter 3 describes the research methodology including research design, task description, prototyping, the evaluation experiment and assumptions employed in this research.

Chapter 4 describes the proposed model of FlexPhoReS based on the literature review and data collected from a small-scale user study. It also discusses the task description and presents the process model of the proposed system.

Chapter 5 discusses the development of FlexPhoReS prototype. This includes a discussion of the enabling technology used in the prototype design and development, prototype design based on the task description and in conjunction with the process model, presenting the architecture and components of FlexPhoReS system. This chapter also illustrates and explains the capabilities of FlexPhoReS prototype.

Chapter 6 presents the results of the user evaluation with the FlexPhoReS prototype. This includes a discussion of participants' background, presents the participants' data and statistical significance tests.
Chapter 7 presents an analysis of the user evaluation. This includes discussion of the test hypotheses and statistical significance. This chapter also presents the results of the study based on usability testing of users' search performance, subjective satisfaction and acceptability of the FlexPhoReS prototype.

Chapter 8 concludes the thesis by discussing the overall research achievements and implication of the research. In addition, it discusses the limitations of the research and points to future research directions.

There are ten appendices included in this thesis to supply more detailed information about certain subjects discussed. Appendix 1 provides the details of the small-scale user study structured interview questionnaires. Appendix 2 provides the structured interview questionnaire of low fidelity informal evaluation. Appendix 3 presents the FlexPhoReS prototype on A1 paper. Appendix 4 presents FlexPhoReS low fidelity prototype (Microsoft PowerPoint version). Appendix 5 provides the recruitment questionnaire. Appendix 6 presents experiment procedure. Appendix 7 presents the questionnaire for user subjective satisfaction and acceptability (suitability and flexibility) of the FlexPhoReS system. Appendix 8 presents participants' demographic characteristics, Appendix 9i-9iv presents the statistical analysis and normality tests of the evaluation and Appendix 10 presents the FlexPhoReS source codes files (CDROM attached) and services.
CHAPTER 2

LITERATURE REVIEW

2.0 Introduction

This review of the literature is divided into four areas of research relevant to the study:

- Section 2.1 reports topics related to the evolution of image retrieval from the perspective of early to later developments.

- Section 2.2 is a review of user studies of image collections. It includes the image user, previous research on user queries in image collections and personal digital photos.

- Section 2.3 reports topics related to user interface support for digital photo retrieval. The review includes photo annotation, user visual access modes, retrieval strategies and other related systems.

- Section 2.4 presents the subject of multimodal user interfaces. Speech interfaces and information retrieval; multimodal user interface and applications; and user interface engineering related to multimodal interface are covered.
2.1 Image retrieval evolution

The purpose of an image database is to store and retrieve an image or image sequences that are relevant to a query. There are a variety of domains such as information retrieval, computer graphics, database management and user behaviour which have evolved separately but are interrelated and provide a valuable contribution to this research subject. As more and more visual information is available in digital archives, the need for effective image retrieval has become clear (Del Bimbo 1999, pp.1-2).

Most traditional image retrieval systems have relied on manually annotation of images by text which can be time-consuming, inconsistent, laborious and expensive. In these, human cataloguers manually produced metadata for images and the text-based image retrieval systems could only retrieve images by matching words from a user query to words that were manually annotated to the images (Del Bimbo 1999, pp.2-3). Figure 2.1 shows the early generation of image retrieval systems.

![Figure 2.1: Early generation of image retrieval systems (Del Bimbo 1999, p.3)](image-url)
Although advances in technology have led to a large amount of research on automatic image annotation to support text-based image retrieval which automatically assigns metadata in the form of captioning or keywords to a digital image. This has not been fully successful and it is still typical to have manually annotated text-based image retrieval as normal way of describing the attributes of images when adding metadata such as captioning, keywords or descriptions to the images before the image retrieval process can be initiated. Advantages of text-based image retrieval include:

- Text based indexing is high-level indexing because it can be used to describe almost any aspect of an image at varying degrees of complexity. Thus a text based search does not have to be highly specific (Eakins and Graham 1999, p.21).

- The searcher can use natural language, Boolean expression and a control vocabulary in the search query (TASI 2004).

There are several disadvantages commonly associated with text searching (Del Bimbo 1999, p.4; eVision 2001, p.4):

- Search by text is language-specific and context-specific. Users have to choose a language that specifies the search.

- Search by text is error-prone. Typographical errors can give incorrect results or an empty result set.

- Search by text can be cumbersome. The users must know the appropriate keywords or descriptions used in the image database. In fact, keywords are inherently subjective and not unique.
• Users might feel frustration if they can picture exactly what images they want, but are unable to translate the images into image retrieval search terms.

• Creating keywords for a large number of images is time consuming.

The type of data used to retrieve images has a direct impact on the internal organisation of the retrieval system. Although keywords and free text have the ability to give a rich and detailed description of image content, textual information for every image in the database requires humans to personally describe them.

Based on some of these disadvantages, automatic indexing and retrieval based on image content or content based image retrieval (CBIR) has become more attractive for some image retrieval applications (Rui et al. 1999, p.40). Although CBIR solves some of the problem of keyword-based image retrieval, it too suffers from several disadvantages (Eidenberger 2000, p.513):

• Bad retrieval results because of the semantic gap and subjectivity of human perception. CBIR uses visual characteristics (low level) of image features which refer to the specific characteristics of the pixels that form the digital image, rather than words that have been attached to the image. As with textual annotation, different persons or the same person in different situations may judge visual content differently.

• Bad image querying performance. CBIR technique involves complex mathematical computational distance functions for the comparison of image features. This can lead to bad and unacceptable processing and response times.
• Complex interfaces: CBIR interfaces tend to be complex and difficult to use. Users do not always find it easy to select appropriate features for a specific querying process.

Text based image retrieval and content based image retrieval techniques complement each other to some extent. Text-based image retrieval can capture high level abstraction and concepts but cannot recognise visual queries. In contrast, CBIR is able to recognise visual queries but unable to capture high level concepts.

Nowadays it is possible to create, process, broadcast and store digital images with increasing efficiency (Kauniskangas 1999, p.15). The new generation of image retrieval systems support retrieval by image content not only performed at the image's primitive level (i.e. colour, shape and texture) but also at the image's logical level using keywords in the textual domain. Image processing, pattern recognition and computer vision are increasingly an integral part of the system's architecture and operation (Del Bimbo 1999, p.4). Automatic indexing and retrieval based on image content becomes an attractive technique for developing image retrieval applications. Several systems incorporate relevance feedback which allows a form of user interaction. These systems automatically create new queries based on sample images identified as relevant by the user in previous queries (Wood et al. 1998, p.14). Figure 2.2 shows the newer generation of image retrieval system.
These image retrieval systems are normally composed of three databases (Rui et al. 1999, pp.54-55).

i) The image collection database that contains the raw images for visual display purpose.

ii) The image feature database that stores the image features extracted from the images using feature extraction techniques. The information needed to support content-based image retrieval.

iii) The text annotation database that contains the keywords and textual information of the images.

This generation of image retrieval systems has inspired further research directions in order to expand the capability of image retrieval, including new user interfaces (Del Bimbo 1999, p.13). However in order to develop an image retrieval system, it
is important to explore not only the technical aspects of image retrieval and databases but also user aspects of image retrieval.

2.2 User studies in image collection

2.2.1 Image users

In a large variety of applications such as map-making, weather forecasting, shopping, works of art and personal photographs, images are increasingly used to express information. In other words, most individuals use images and interact with images in different ways at different times. Graphic designers, for example, use images in their daily job. Other groups of individuals such as librarians are required to find images on behalf of others (Eakins and Graham 1999, p.11). The following examples are offered of some activities which depend on the use of images, summarised from Eakins and Graham (Eakins and Graham 1999, pp.11-13):

- Crime prevention
  Photographic information becomes very important because the police might be using the information to identify people for proof of evidence in crime.

- Medicine
  Health professionals use and store image information in the form of X-rays or other scanned images for diagnosis and monitoring purposes.

- Fashion and graphic design
  Photographs and graphics from the real world might provide inspiration to visualise the final product. Two dimensional sketches and three dimensional geometric models are used to present ideas.
• Publishing and advertising
In the publishing and advertising industries, photographs are used extensively to illustrate books and articles in newspapers and magazines. Most newspaper publishers maintain their own libraries of photographs and will reuse them if necessary.

• Architecture and engineering design
In architecture, photographs are used to record interior and exterior shots of buildings including particular features of the design.

• Historical research
Historians from a variety of disciplines such as art, sociology and medicine use visual information sources to support their research activities. Archaeologists rely heavily on images as in some cases the information is the only evidence available.

To understand the needs of image users, studies of user queries in image collections were examined.

2.2.2 Studies of user queries in image collections

Most research in image retrieval user studies has been concerned with problems involved in retrieval, for example, how people search for images? What users need? What attributes of an image are indexed? and how images are used?

The study by Enser and McGregor (1992) on the Hulton Deutsch Picture collection was believed to be among the first major studies of user demand for image information. With approximately ten million items (mostly black and white negatives and prints), it was the largest such archive in Europe at that time. They analysed some 2722 queries after the queries had been answered and classified them into two categories: unique (equivalent to the specific level, e.g. an image of
King George V) and non-unique (which appears to encompass both the generic and abstract levels, e.g., Kings of England). They also noted that within these basic categories, the requirement could be subject to refinement (by time, location, action, event, or technical specification). In this study, only requests of the type 'unrefined unique requests' were found to be easily satisfied by the classification scheme in use at Hulton Deutsch. 'Unique' requirements were the most common overall; those without refinements made up 42% of the total sample, and those with refinements made up 27%. These were followed by non-unique requirements with refinements (25%) and those without (6%). They also noted that non-unique subjects needed a more detailed indexing system (Enser 1993).

Armitage and Enser (1997) extended this work further by analysing and categorising additional user queries for still and moving images across seven libraries with more specialised content and classified a sample of 1749 written queries. They found four major types of queries as follows: i) image content ('find me some images of...'); ii) identification/attribution/provenance checking; iii) accessibility of image/artist of work (e.g., ownership/viewing availability); and iv) miscellaneous (e.g., requests for administrative procedures only, or unusable queries). According to Goodrum (2000, p.65), the finding formed the foundation framework for queries with 4 main categories (who?, what?, when?, where?) and 3 levels of abstraction (specific, generic, abstract).

Keister (1994) examined the characteristics of 239 user queries at the National Library of Medicine's (NLM) Prints and Photographs Collection. The user group which was involved in this study included picture professionals, researchers, TV, film or media personnel, clothing designers, health professionals and the museum and academic communities. She found that the museum and academic community often had precise citations to the preferred images. Health professionals asked for images in keeping with the NLM's orientation, i.e., images can be accessed by appropriate topics, such as a particular disease. Picture
professionals, which included groups of still picture researchers, TV, film or media personnel, think visually and use art and graphic jargon to describe specific preferred images. Consequently this study confirmed that most image requests can be easily searched without images, but some subjective needs are only satisfied by the user's interaction with a substitute image.

In both a manual and computerised environment, Hastings (1994) cited in Jorgensen (1999) conducted a qualitative analysis of the search behaviour of eight art historians. The results show that searches become more complex in the computerised environment compared to a manual environment. She also identified three types of search styles:

- **Browse searcher**
  
  Browse searchers formed their own image categories and used more complex images.

- **Subject searcher**
  
  Subject searchers imposed a preconceived image classification scheme and used textual information to assist the classification of object and activities.

- **Text searcher**
  
  Text searchers worked mainly from textual information.

The above studies emphasise that image collections indexed for specific groups of users may be searched by a broader group than the original target audience. This reflected one of Del Bimbo's disadvantages of textual searching where the index language used may not be appropriate for all of the users who are searching the images (Del Bimbo 1999, p.4).
An exploratory study by Jorgensen (1998) was concerned with how participants describe, view and retrieve images. The study indicates that the most typical reported attributes in these tasks were object, person, social status, colour, body part, location (specific) and activity. Consequently, in order to create a template for image descriptions, three classes of image attributes were suggested:

- **'Perceptual' classes**: related to physical content of the image.
- **'Interpretive' classes**: stimulated perceptually but require additional internal interpretive and intellectual processes in order to name the attribute.
- **'Reactive' class**: which related to responses such as conjecture or emotion.

In addition image queries can be modified by a number of other more detailed considerations, for instance date, location, activity, event, image orientation, perspective or type of representation.

The study by Ornager (1997) presents types of journalist user queries conducted in newspaper image archives. A user typology was proposed based on the journalists' queries:

- **'The specific inquirer'** who asks very specific questions because they have a specific pictorial query in mind;
- **'The general inquirer'** who asks very broad questions because they want to make their own choice;
- **'The story teller'** inquirer who tells the story and is open to suggestions from the archive staff;
• 'The story giver inquirer' who passes the story over to the staff wanting them to choose the photograph(s); and

• 'The fill in space inquirer' who only think about the size of the photograph in order to fill an empty space on the newspaper page (Ornager 1997).

She also suggested the minimum elements to be included in subject analysis and photo indexing were as follows:

- Name of person (who?)
- Name of phenomenon (what?)
- Geographical information (where?)
- Time information (when?)
- Specific events (what?)
- Mood and emotions (shown or expresses)
- Associative meaning (what?)
- Size of photo

Markkula and Sormumen (1998, 2000) conducted a study that addresses photo needs and searching behaviour of journalists. In the context of photo needs, they found that journalists' requests fell into four categories (Markkula and Sormunen 1998):

- Concrete objects (i.e. people, buildings or places);
- Themes or abstractions interpretable from the photo;
- Photo's background information (such as specific news events and films and television programmed); and
- Known photographs (searched for or requested by time of publishing, photo taken date, place or the photographer)
With half of all requests for persons, the first category of requests dominated the use of the photograph archives. In the context of searching behaviour, they found that journalists tended to make single-word or single-phrase queries and were likely to find photos through the names of persons, places or events. They also found that browsing frequently required less effort and time than formulating a query. In order to make the query stage more effective to journalists, Markkula and Sormunen (2000, p.283) recommend a ‘hybrid’ system which supports traditional concept-based indexing and classification methods and develop user interfaces to support photo browsing. Automatic visual methods could then be applied to the browsing stage.

Choi and Rasmussen (2002) conducted a study on users’ queries employed through the information seeking process in American history. Data were collected from unstructured interviews and questionnaires. The study involved 38 faculty and graduate students of American history in 1999 in a local setting using the Library of Congress American History Photo Archives. The study indicates that across various types of users, there was a similarity in image query formulation. Users tended to search for relevant images on the basis of title, date, subject descriptors and notes provided.

Eakins et al. (2004) conducted a study that concentrated on user requirements for image databases. The study involved 125 experienced image searchers. The findings indicate that the majority of participants were more interested in concept based retrieval than content based retrieval. These findings strengthen the present dominance of concept based image retrieval (Enser 2000). The study also suggested that an image retrieval interface should have the following components:

- free text entry
- specification of overall focus and sharpness of image
- selection of the type of original image required (e.g. painting, photograph)
• selection of image file criteria (size and type)
• emotional impact of image
• selection of technical terms (where appropriate)

The authors also concluded that even experienced image searchers are conservative in their expectations and that system designers must be aware of this even when designing more innovative systems.

More recently, Othman (2005) conducted a study that addressed the image retrieval behaviour of university academic staff in the area of creative multimedia which covers a wide range of subjects: art, computer designs and graphics, history, culture and heritage, broadcasting, and engineering. The study involved 35 participants. Web based Google Image Search category was used as an image retrieval search tool. The findings indicate that most participants requested images with image captions. The most important criteria for relevance judgment were technical attributes, topicality, and completeness of the image. The study also indicates that all of the participants were sympathetic to the idea that visual search mode would be relevant and they would like to find similar images before they can choose the best, although they did not expect it as an important feature in an image retrieval system.

2.2.3 Personal digital photos

Photography is a universal medium that can be used to express a certain vision or idea (Taylor et al. 1982, p.73). It can be an effective medium for storage and communication of information among individuals (Efford 2000, p.1). Most photo snapshots record a particular scene or event to help people to remember. For example, a news photograph implies what people would have seen if they had been there themselves (London and Upton 1994, p.356) while a personal photo collection related to family life could be very valuable and a welcome gift when a child reaches adulthood (Shriver 1980, p.76). "Digital photography is part of a
complete process, from capturing a picture to the social use of these photos" (Vroegindeweij 2003, p.5). Digital photos are personal digital photos when they are owned by the user and are under their direct control. Some user studies related to personal digital photos collection have been conducted.

Rodden (1999) and Rodden and Wood (2003) investigated users' practices in organising personal photo collections, initially non-digital and subsequently digital. While institutional image collections tend to be organised objectively in an attempt to suit the needs of a possible wide range of users, personal photographs, according to Rodden (1999, p.1), are part of the wider group of personal documents. For that reason, organising personal photos can be related to organising personal documents. Based on a study of personal document management (Malone 1983), Rodden (1999) applies these observations to personal photographs. In his study of non-digital personal photographs, he gained some insight into how computer-based systems might be designed to help people organise photos. Twelve people ranging in age from 24 to 62, with an average age of 36 were interviewed about their behaviour in organising their physical photograph collections and were asked for their opinions on a number of possible features of a computer-based system. The study was investigative in nature; the subjects were asked open-ended questions on how they might imagine organising their photos with computer support. The findings indicated that people would like to have their photographs categorised and ordered mainly in order to make browsing easier. On searching categories, people usually search for a particular photograph they have remembered, for instance, they might have some idea of the photo's location where the photo was taken. Interviewees were more interested in expressing searching requirements textually with the aid of annotations instead of constructing image queries. Content-based image retrieval techniques would be used in the image browsing stage.
In a second study, Rodden and Wood (2003) extended the investigation to digital photographs. Thirteen participants were studied over a period of six months on their use of a prototype digital photo management software (Shoebox). The prototype software allowed browsing folders by event and time as well as some advanced features such as content based image retrieval and speech annotation. Results showed that people used the simple features effectively and reported they found their digital photographs easier to organise than had been the case with non digital photographs. Unlike institutional image collections, the participants did not engage in making detailed textual annotations their photographs. The findings indicated that because people are familiar with their own photographs they usually find what they need by browsing.

A number of studies have concentrated on users' practices in photo sharing. Frohlich et al. (2002) identified photo sharing practice by combining an ethnographic study and observations, interviews and self-recording techniques. A total of eleven home personal computer owning families that were already involved in some form of digital photography technology were interviewed about their use of conventional and digital photos. They also completed photo diaries and recorded conversations about photo use that occurred spontaneously over a three month period after the interviews. The results illustrated the strengths and weaknesses of past and present technology for photo sharing and found that digital photography had not replaced conventional photography among the families. Very few families organised their digital photo collections on the personal computer. The computer monitor was found an inappropriate tool for synchronously reviewing and sharing photos with others (Frohlich et al. 2002).

In a more recent study, Crabtree et al. (2004) employed ethnographic studies of paper-based photographs to consider requirements for distributed collaboration around the digital photographs. The study concentrated on the practical achievement of sharing 'at a distance' to support the distributed approach to photo sharing. The findings raise a number of challenges for future development of
digital photo distributed systems such as combining collaborative interfaces to support the movement and manipulation of photographs and to permit the flexibility of real-time sharing.

Van House et al. (2004) identified a set of social uses of personal photos from interview sessions. The discussion of the social uses was divided into three issues (summarised from Van House et al. (2004, pp.6-7)):

(1) Memory; the memory of a photo has informal and emotional components. Favoured photos were usually spoken of in terms of the memories and emotions evoked rather than quality of the photos;

(2) Relationship; photos can be used, not just to remember people and events, but to maintain existing relationships and even create new ones;

(3) Self expression; photos can be used as self-expression and self-presentation. Self-expression is about giving expression to the “authentic” self while self-presentation is about influencing others' views of one-self.

The study also provides a catalogue of what non-digital and digital photo users said and demonstrated about their personal photography practices from five different perspectives (summarised from Van House et al. (2004, p.4)):

(1) Camera use
Participants learned to use their digital cameras more and reserved analogue cameras for photo expeditions. Most participants believed that analogue cameras created better quality images and so captured more “important” images on the analogue cameras;
(2) Photo taking patterns;
The popular patterns of taking photos are based on family and friends, vacations, special events, and pets. Two distinctive types of photos were identified (i) "art" (taken for aesthetic reasons) and (ii) "fun" (funny in and of themselves, or in the context in which they were to be used). Most commonly, between 10% and 25% of photos taken were put on display in photo albums, frames, bulletin boards, refrigerators, and the like. The review of photos online showed that people use online sites for the same purposes (friends and family, vacations, events);

(3) Storage and retrieval
Printed photos are more appropriate than digital media for archiving photos. Many users are "too lazy" to annotate their own photo collections. However, most participants did minimal annotation, most commonly a scribble on the outside of an envelope of prints noting date, location, or event or people. Time was found as a main organising principle and most digital camera users had no more than one layer of folders, given a descriptive name about place, event, or person. On the issue of audio annotation, the reactions were mixed. In essence, people did not want to do the recording. However the preference was for face-to-face storytelling which balanced any perceived benefits of audio annotation;

(4) Photo sharing
Every participant would like to share their photos with their family and friends. They often shared their photos by passing around envelopes of prints and some left prints lying around in high traffic areas of the house for people to look. Other ways of photos sharing are as follows: (i) camera phone; (ii) TV; (iii) personal computer; (iv) email; (v) upload to web site; and (vi) PDA devices.
What features do they want? Among the features that participants want are as follows: (i) zoom; (ii) flash; (iii) better resolution; (iv) digital camera for children: inexpensive and rugged.

Davis (2004) introduced a mobile media annotation system and built a framework ("MMM" for "Mobile Media Metadata") that enables photo annotation at the time of photo capture using Nokia 3650 camera phones over the AT&T Wireless GSM/GPRS service. The system stores the photos and metadata and assists the user in annotation on the camera phone by providing guesses about the content descriptions of the captured photos. Fifty five participants were given camera phones to test the system for four months and the findings concluded that mobile camera phones with the MMM system have made a new approach that can reduce user effort by facilitating metadata capture at the time of image capture, adding some metadata automatically and leveraging networked collaborative metadata resources (Wilhelm et al. 2004, p.1403).

Later on, Van House et al. (2005) conducted an empirical study of the uses of a developed prototype camera phone application for Mobile Media Metadata 2 (MMM2) proposed by Davis et al. (2005). Sixty participants (40 students, 20 researchers) were given Nokia 7610 camera phones. The camera phones were loaded with the MMM2 application. From the findings, Van House and her colleagues saw similar social uses of personal digital photos that related to prior work on photographic practices (Van House et al. 2004) and added a new social use as an additional social use:

(6) functional: self and others. The camera phones with a good quality of images sometimes are used (e.g. sent images of a clock to remind his workgroup) instead of writing, copying or scanning.
2.3 User interface support for digital photo retrieval

The user interface is an important component of information retrieval systems because it connects user and computer and allows interaction for users to the organised information resources such as digital photos (Chowdhury 2004, p.227). Retrieving digital photo information involves a number of stages and at each stage, numbers of actions are taken and decisions are made. The user interface of the retrieval system may provide support in performing these actions. This has been widely discussed in the literature (Marchionini 1995; Spink and Saracevic 1998; Hearst 1999). Most Interfaces assume an interaction cycle consisting of query formulation, retrieval and inspection of retrieval results, and then either stop or reformulate the query until a suitable result set is found. Shneiderman et al. (1997) proposed an interface design framework for the information retrieval process which can be divided into four major phases: (1) formulation; (2) action; (3) review of results; and (4) refinement.

- Formulation
  Formulation is triggered by an information need and several decisions are made regarding what to search for and the search variants.

- Action
  In the action phase, usually a search button needs to be pressed to conduct the search process. The dynamic queries where there is no search button; the results set is continuously displayed and updated as phases of the search are changed.

- Review of the results
  The third phase is the review of the results, in which users read messages and view results set by selecting some available output attributes such size of the display and sequencing of the retrieved items.
• Refinement

The fourth phase is refinement, in which the interfaces provide facilities for modifying and refining queries.

More recently, Sheniderman and Plaisant (2005, p.566) extended the ideas by proposing a new phase that added the fifth phase of the framework:

• Use

The fifth phase is use, in which the search results can be saved or used as input to other programs (e.g. visualisation, email)

2.3.1 Photo annotation

Photo annotation is the process which involves labelling the semantic content of photos with a set of keywords or other semantic information (Suh and Bederson 2004, p.1). The type of data that is used in photo annotation is important because it has a direct impact on the way in which photo retrieval is performed in locating the photos from the photo database.

Usually personal digital photos taken by digital cameras only include limited metadata such as time and date. Further metadata are needed in order to allow more retrieval options. Photograph users, for example, may record names, dates, places, and events. However, other types of data could be used for photo annotation, depending on the origin of the photos (Rodden 2001, p.26). For example, Toyama et al. (2003) developed a system that capitalises on geographical location metadata in digital photographs. The system is capable of creating appropriate context maps for a slideshow of location-tagged digital photographs and was developed on top of the World Wide Media eXchange (WWMX) protocol and services.
Digital photos with no metadata will be difficult to find or identify (Besser 2003, p.3). The following types of information could be associated with the digital photos (Del Bimbo 1999, p.2):

- Content-independent metadata is related to the digital photo content, but does not describe it directly. Examples of such data are: author's name, date, location, ownership, etc.

- Data which refers to the visual content of digital photos. It can be divided into two types:
  
  o Content-dependent metadata refers to low/intermediate-level features (colour, texture, shape, motion, etc.).

  o Content-descriptive metadata refers to content semantics. It is concerned with relationships of image entities with real-world entities or temporal events, emotions and meaning associated with visual signs and scenes.

The 'cost' of photo collections rises steeply as soon as the photo users attempt to annotate their photos collection (Kustanowitz and Shneiderman 2005, p.1). Annotating photos can be a time-consuming, tedious, error-prone data entry task activity and can discourage most owners of personal photo libraries (Shneiderman and Kang 2000, p.88). Researchers have explored techniques for improving the process of photo annotation. For example Shneiderman and Kang (2000) developed a direct annotation method that focuses on labelling names of people in photos. "The user can simply select a name from a manually entered name list and drag and drop it onto the photo to be annotated. Although it avoids most of the typing, it is still a manual method that involves many drag and drop operations" (Shneiderman and Kang 2000, p.88).
To address this manual annotation problem, there has been a large amount of research done on image annotation that can be explored for personal digital photo annotation. Kustanowitz and Sheiderman (2005, p.2) grouped the annotation techniques into three categories: Manual, Semi-Automated, and Automated.

- Automated annotation happens with no user involvement, for example automated annotation is done inside the digital camera by applying a time stamp.
- Manual annotations involve a process when the users enter some descriptive keywords manually when the photos are loaded or registered or browsed. However it is slower but more precise than automatic annotation.
- Semi-automatic annotation is a combined technique of the manual and automatic annotation approaches.

Table 2.1 shows the annotation techniques from the perspective of human effort and machine assistance.

Table 2.1: Annotation techniques (Kustanowitz and Shneiderman 2005, p.2)

<table>
<thead>
<tr>
<th>Human Effort</th>
<th>Machine Assistance</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Manual</strong></td>
<td><strong>Add structured annotations, with sufficient semantic information to be useful for retrieval. Save annotations in a database.</strong></td>
</tr>
<tr>
<td><strong>Semi-Automated</strong></td>
<td><strong>Add freestyle annotations or captions. Potentially work with machine’s output in an iterative fashion.</strong></td>
</tr>
<tr>
<td><strong>Automated</strong></td>
<td><strong>Verify machine’s accuracy and make corrections as needed.</strong></td>
</tr>
</tbody>
</table>
2.3.2 User visual access modes

Information retrieval systems vary in terms of design, objective, characteristic, content and users. Most image retrieval systems provide browsing and searching modes of access (Flickner et al. 1995; Smith and Chang 1997; Nakazato et al. 2003). Jomni and Yetongnon (2001, p.4) classified the user visual access modes into three modes of access which can be used separately or combined in a digital photo retrieval system:

- **Free mode**, or search by specification. This mode of access requests precise information that has been used to describe semantic information of the image (such as date of photo capturing, author, etc.), image attributes (such as colour) and image object properties (such as shape).

- **Guided mode**, or search by example. This mode of access requires the image retrieval system to extract one or more query parameters of the image (e.g. colour, shape) from image example and searches for the other images that have similarity parameters.

- **Navigational mode** or browsing. This mode of access is an exploratory mode which allows the user to navigate through display pages (e.g. results set pages) to find the desired images.

The above modes of access indicate that various approaches could be applied to image retrieval systems. Therefore, many different types of image retrieval user interfaces can be found. These approaches have been further classified into image retrieval strategies.
2.3.3 Retrieval strategies

In general retrieval strategies can be divided into five categories (Lai 2000, p.19): query by text, query by image example, query refinement, browsing and navigation.

- Query by text is the most common method used in image retrieval systems and requires very little learning time. The system does the matching process based on the words entered by users (Chuah et al. 1997, p.88). Users can also retrieve images with high-level concepts such as object names in the image or the location where the photo was taken. However, in order to make a text-based approach effective, the users have to annotate all of the digital images and it can be an extremely tedious task for home users.

- One of the most popular methods of visual retrieval is Query by Example (QBE) (Kato 1992). Basically the users are required to select an image example and ask the system to retrieve visually similar images from the database. With this method, users are able to convey their visual queries through the image example submitted. However, this method may be difficult to use if the users are unable to find an appropriate photo example that represents what is desired. In such cases, they have to look in detail through the library to find an appropriate photo example. Compared to query languages such as SQL, QBE is more transparent to the user who has to compose a query (Chuah et al. 1997, pp.88-89). Several studies have further classified a subtype of query by example. For example Verdrig (1997) has classified three subtypes of query by example: query by external example, query by internal example and query by sketch.
Query by external example is the form of query by example in which example images are provided from external sources (such as the internet or elsewhere). The user can then ask the system for similar images in the database. The influence of query by external example however, decreases if it is hard for the user to come up with an example image or photo.

Query by internal example is when example images are selected from the retrieval system's image database. Users can select a query image from the available image collection. Users are not required to generate or capture any image. However, the user still has to find an appropriate example image from the database, which can be tedious and time consuming.

Query by sketch can be further divided into two types (Chuah et al. 1997, p.89): Free-hand sketch and object manipulation sketch. Employing free-hand sketching, users freely draw the query using a mouse, pen or other input devices. In object manipulation sketching, users may construct sketches from available primitive objects (such as palettes). The sketch is then analysed, and key features are extracted (such as colour, shape) for the purpose of matching. However, in the real world, sketches have a high form of abstraction. Users are required to draw only the important parts of the image they desired. The advantage of query by sketch is that users can specify their information need by indicating only the most important particulars of an image while the disadvantage of query by sketch is that it requires a user with some artistic capabilities (Chuah et al. 1997, p.89).
Chapter 2

- Query refinement or relevance feedback is a process that allows users to improve query formulations by changing the original queries (Salton 1989, p.307). Some systems attempt to apply the concept of query refinement by allowing users to re-enter their feedback by indicating positive and negative relevance of the original query results (Del Bimbo 1999, p.8).

- In browsing, users may look through the entire collection. Browsing allows users to recognise a relevant image when they see it. After they have received results, there may be some images that need to be browsed visually before making the final image selection (Bederson 2001, p.71). Browsing has the capability to help users in determining their information need and provide an overview of the data by presenting it to them. Based on these capabilities, Verdrig (1997) distinguishes two forms of browsing in the domain of image retrieval: visual browsing and query by association.

  o Visual browsing is a form of browsing in which all images in collection are inspected by the user one by one. This method implies that the users already know that the image exists. Query formulation is not required. However the disadvantage of visual browsing is that the users may have to inspect all of the images before they can find the image they desire or realise that the image is absent in the collection. This limitation has made visual browsing mainly suitable for a quick scan of a small amount of images.

  o Query by association is the retrieval of data items by selection of a data item that is associated with the data items to be retrieved involving the hypermedia concept. Employing query by association, users can navigate through a collection of images based on the semantics of the images. The ‘query’ can be created simply by a mouse click on the data item (such as an image or an abstraction).
However, query by association is more suitable for a retrieval system where images can be grouped according to their content.

- Navigation allows users to navigate around the multi-dimensional "information space" to seek specific images by accessing the specific user interface of the system (Lai 2000, p.25). For example "El Niño" query models proposed a display space as either a two dimensional surface or a three dimensional volume place that shows the results of user queries (Santini and Jain 1999). Figure 2.3 illustrates the 2-D displays and 3-D displays in "El Niño" navigation user interface.

![2D display](image1.png) ![3D display](image2.png)

Figure 2.3: Illustration of 2-D and 3-D display (Santini and Jain 1999, p.525)

The retrieval strategies presented above can be implemented by the digital photo retrieval system user interface by providing several interfaces to the same digital photo database. In practice, most image retrieval systems consist of a combination of some of the retrieval strategies described above. In the following sections, several systems related to digital photos are presented.
2.3.4 Studies related to digital photo applications

In recent years, digital photo systems have been developed in both commercial and academic organisations for both public and personal image collections. Some of them are described below.

QBIC (Query by Image Content) was developed at the IBM Almaden Research Center and was among the first commercial system of image retrieval on the basis of their visual content (Flickner et al. 1995). QBIC pronounced "cubic" allow users to make queries of large image databases based on visual image content properties such as colour percentages, colour layout, and textures occurring in the images. The information on the current version of QBIC is available at (http://www.qbic.almaden.ibm.com).

Photobook developed at MIT media laboratory (Pentland et al. 1996) is a set of interactive tools for browsing and searching images and image sequences. It runs under the UNIX/Linux operating system. Photobook allows the user to browse large image databases using both text annotation information and by image content.

Multimedia Analysis and Retrieval Systems (MARS) developed at the University of Illinois (Huang et al. 1996) is an image retrieval system that supports similarity and content based image retrieval based on a combination of their colour, texture, shape and layout properties. The user interface was written in JAVA applets and accessible over the World Wide Web using the Netscape browser. The user can specify queries to retrieve images based on a single property or a combination of image properties. In addition the interface allows users to combine image properties as well as text annotations in specifying user queries. A few papers have reported improving MARS features with relevance feedback (Porkaew and Chakrabarti 1999; Chakrabarti et al. 2000).
Some studies deal with user interface agents that assist users by proactively looking for opportunities for image annotation and image retrieval. For example Lieberman et al. (2001) from Eastman Kodak have developed Aria: An Agent for Annotating and Retrieving Images. The prototype intelligent agent was designed to assist users by proactively looking for opportunities for image annotation and retrieval. ARIA was applied to an email program by continuously monitoring the user’s typing, searching for images that might be relevant to the content of the message and displaying them beside the message window. These features could reduce user interface overheads and lead to better annotated image libraries with less missed opportunities for image use.

Nakazato et al. (2003) proposed a flexible image retrieval interface for CBIR system namely query by groups. In this system the users can interactively compare different combinations of query examples by dragging and grouping images onto the workspace. The concept of “image groups” is also applied to annotating and organising a large number of images.

Platt et al. (2003) from Microsoft Research have proposed Photo Table of Contents (PhotoTOC): Automatic Clustering for Browsing Personal Photographs. PhotoTOC is a browsing user interface that helps users find digital photographs in their own collection of hundreds or thousands of photographs. PhotoTOC uses metadata provided by digital cameras to provide a simpler, more intuitive, time-based user interface. PhotoTOC was developed by design iteration on an earlier clustering Microsoft AutoAlbum user interface (Platt 2000). Graham et al. (2002) proposed two photo browsers (calendar browser and hierarchical browser) to accommodate thousands of photos. The browsers were supported by cluster analysis of the times (when photo was taken). Results from a user study show that the time element and appropriately summarising photo collection can lead to significant improvements in the time taken to search for photos.
Certain systems support multiple platforms for organising digital photo collections. Gargi et al. (2002) proposed a system for managing and searching a personal digital photo collection across multiple devices. With Java technology the system allows users to organise and view digital photos across local and remote computers. Users can search the photo collection based on time (event), face detection and visual similarity augmented by digital camera metadata. The system also allows users to customise photo collection views.

Mitsubishi Electric Research Laboratory (MERL), Cambridge Research Laboratory and University of Illinois described the Personal Digital Historian (PDH) system that supports an interactive multi-person environment. The system enables informal storytelling, using personal digital data such as photos, audio and video in a face-to-face social setting (Shen et al. 2003). The system enables natural face-to-face conversation through the user’s desktop or handheld devices. In addition PDH supports table top display metaphor browsing and based on 4Ws (who?, where?, when?, what?) for organisation and navigation.

Some studies emphasise support for photo annotation tasks. For example, PhotoFinder allows users to annotate photos with free text fields plus date and location fields stored in a database. The system also provides Boolean query and direct manipulation photo labelling (Kang and Shneiderman 2000). Microsoft Research has proposed a system (MiAlbum) for home photo management using manual and semi-automatic image annotation approaches. MiAlbum system was implemented using Microsoft Visual C++. The system allows users to retrieve photos by keyword searching and query by example (Wenyin et al. 2000). Recently Microsoft Research proposed a system, MediaBrowser (Drucker et al. 2004). The system integrates a variety of visualisation and interface techniques (e.g. grid view, time, gallery view and cluster view) that allows user to select, filter and name large number of photos and videos. Some systems take advantage of face recognition (extract faces from digital photo) for automatic photo labelling (Girgensohn et al. 2004).
A few systems utilise audio annotations for digital photos. Show&Tell which runs on a Sun Ultra, provides two stages of functionalities i) audio annotation in addition to visual-based methods and ii) querying that provides point and click querying synchronized with speech on the annotated digital image (Srihari and Zhang 2000). Mills et al. (2000) from AT&T Laboratories proposed a research prototype named AT&T Shoebox. The prototype is an application for organising, annotating, indexing, searching, and displaying collections of personal digital photos obtained either from digital cameras or by scanning photographs taken with conventional cameras. AT&T Shoebox provides three primary features: (a) thumbnail-based browsing tool for organising, labelling, and viewing photos; (b) audio annotation whereby users can speak about their photos when loading digital photos into the system; and (c) Image analysis and indexing algorithms which allow the user to search for photos based on their visual content. Chen et al. (2003) proposed a system named SmartAlbum which is capable of using speech syntax to annotate photos based on event (when?), location (where?), people (what?) and time (when?). The system utilises keyword based textual queries. Prior to retrieval, each photo was annotated in both textual and audio form (each audio file is less than 20 seconds). It also allows the users to retrieve photos based on the following considerations; i) pull out all photos with faces in them; ii) photos with a specific number of faces; and iii) determine the specified location of faces in the photos to be retrieved (Tan et al. 2002).

Recently, some studies have utilised Global Positioning System (GPS) and cellular technologies in adding photo location (the latitude and longitude where the photograph was taken) information to digital photos. For example, Naaman et al. (2003) described LOCALE, a system that allows metadata sharing for digital photos with geographic coordinates. Later on, Naaman et al. (2004) proposed PhotoCompas that is capable of automatically generating a meaningful organisation for personal photo collections. The system utilised the time and geographic location stamp embedded in digital photos to produce three different
types of browsing output: event segmentation, location (geographical coordinates)
hierarchy and suggested names.

More recently, some related systems were developed to support users to manage
their digital photo collection using small screen display devices such as PDA and
mobile phone. Davis (2004) introduced a Mobile Media Metadata (MMM) system.
The system provides opportunities for mobile phone users by: (1) enabling photo
annotation at the time of capture; (2) leveraging contextual metadata and
networked metadata resources; and (3) enabling iterative metadata refinement on
the mobile imaging device. Later on, Mobile Media Metadata 2 (MMM2) was
developed to address three central problems in camera phone use: (1) getting
photos off the phone; (2) finding and managing photos; (3) sharing photos (Davis
et al. 2005). The MMM2 system allows users to add captions to photos and
videos at the time of capture. It also allows users to send their photos from the
phone to other MMM2 phones or via email. MMM2 also automatically uploads all
photos and videos to the MMM2 phone network database, where each user has
their own private MMM spaces.

Multimodal interfaces for digital image retrieval have received attention recently.
Käster et al. (2003) introduced a system for interactive standalone content based
image retrieval. The system utilised a multimodal user interface that allows
multimodal interaction combining mouse, keyboard, speech and touch screen.
From usability experiments, they found that participants well appreciate the
multimodal interfaces for image retrieval.

A few digital photo systems provide slide show and digital album generator to the
users, for example Apple iPhoto2 (Apple) and Adobe Photoshop Album (Adobe).
Both systems provide facilities for users to convert their photo collection to web
format.
Several systems provide web based digital photo management services, for example Ulead iMira (Ulead 2005) and Fotomill (Inspiral 2004). Both systems allow users to organise personal photos anywhere and at anytime and offer remote admin through a standard web browser. In addition the systems allow users to share photos with other users. Both systems only provide query by keywords for user to retrieve their photos.

The above studies indicate that past and present research has focused on digital photo systems from different views. Many of the studies have concentrated on retrieval and annotation issues. Not much research however, has been done regarding user interface support for the retrieval systems. None of the systems utilised web based multimodal user interface for personal digital photo retrieval.

2.4 Multimodal user interface

Multimodal user interface is currently an important research area in human computer interaction. There are three groups of modality definitions in the study of human computer interaction (Dannenberg and Blattner 1992, p.24). The first definition of modality refers to the sense by which information is perceived (sensory modality). For example, Braille embossed on paper may be perceived through more than one modality, touch and sight. The second definition of modality refers to a state of interaction such as "insert mode" in a text editor. Finally the third definition of modality has been used to refer to the style of interaction. For example, keyboard entry, speech entry or a formal command language are considered different modes. In this study the terms modes and modality considered is the third definition, the style of interaction.

Interaction styles provide a behavioural view of how the user communicates with the system (Hix and Hartson 1993, p.6). In practice, there are many interaction styles which people can use to interact with a system. Shneiderman (1998, p.213) categorised the major types of interaction styles as command mode, menu
selection, form fill in, direct manipulation and natural language. Users can interact with the system using input devices such as keyboard, mouse, pen, touch screen and microphone. The use of these input devices can involve different operations such as pressing, clicking, tapping, touching and speaking.

A growing interest in multimodal user interface has recently emerged and has inspired more transparent, flexible, efficient, and powerful human-computer interaction (Oviatt 2003, p.286). Humans naturally express themselves and communicate through multiple modalities. For example, hearing is the sense of sound perception, sight or vision describes the ability to detect objects within the visible range and tactile modality involves the sense of pressure. Users must be able to use the system effectively and efficiently to perform actions and achieve results in the chosen workplace or at home. However, humans have some limitations in terms of cognitive, visual, speech, and motor ability concerning the quantity and quality of information they can process (Weinschenk and Barker 2000, p.187). Among the limitations are (summarised mainly from Weinschenk and Barker (2000, pp.189-191)):

- **Human visual processing limitations**
  - People do not read everything on a screen, page or window.
  - People will not be able to find information if it is on a screen that is too full and cluttered.
  - People get unfocused easily with visual stimuli such as graphics or too much information.
  - People have trouble reading fonts that might be too small or in too many different styles or sizes.

- **Human motor processing limitations**
  - People cannot hit targets that are too small on the screen.
  - People may have difficulty double clicking.
  - People do not always realise they can or should drag and drop.
People do not like to be constantly switching between a mouse and a keyboard.

People must have time to get used to pointing devices such as a mouse.

- **Human speech processing limitations**
  - Listening to synthesized speech requires more processing capacity than listening to natural speech. Synthesized speech is harder for humans to encode. However once encoded, synthesized speech is stored as efficiently as natural speech (Sanders and McCormick 1993, p.216).
  - Speech is slow for presenting information, is transient and therefore difficult to review or edit, and interferes significantly with cognitive tasks (Shneiderman 2000, p.63).
  - The background noise and variations in users' speech could also interrupt the accuracy of speech interpretation (Shneiderman 2005, p.375).

A multimodal user interface could provide the user with more than a single mode of interaction but for a user interface to be easy to use, it cannot overload any of these limitations.

Many tasks in system applications are suited to a multimodal user interface (Weinschenk and Barker 2000, p.194). Theoretically, multimodal user interface can be classified into two main categories which are (Balbo et al. 1993, pp. 269-270): i) Exclusive multimodal user interface and ii) Synergic multimodal user interfaces.

A user interface is exclusively multimodal, if multiple modalities are available to the user and an input (or output) expression is built up from one modality only. For example, to activate or open a Windows application, the user might choose...
among double-clicking an icon, using a keyboard shortcut, or say "open window". Therefore at a given time, an input expression uses one modality only.

A user interface is synergically multimodal, if multiple modalities are available to the user and an input (or output) expression is built up from multiple modalities. As an example of a synergic multimodal user interface, the user of a graphic editor such as Talk and Draw (Salisbury et al. 1990) can say "put that there" while pointing at the object to be moved and showing the location of the destination with the mouse or a data glove. In this formulation, the input expression involves the synergy of two modalities. Speech events, such as "that" and "there", call for complementary input events, such as mouse clicks and/or data glove events, interpretable as pointing commands.

Sienel et al. (2004, pp.1-2) listed six advantages of multimodal interaction: (1) User can select the preferred modality of interaction at any time; (2) can be extended to selection of the preferred device (multi-device); (3) user is not tied to a particular channel's presentation flow; (4) improves human-machine interaction by supporting selection of supplementary operations; (5) interaction becomes a personal and optimised experience; and (6) multimodal output is an example of multimedia where the different modalities are closely synchronised.

2.4.1 Speech interface and information retrieval

For many years, graphical user interfaces (GUIs) have dominated the human computer interface for most information retrieval systems. Graphics elements have become the dominant medium in which computer based information is delivered to people. Many information retrieval systems with GUIs allow users to operate using mouse and keyboard to type their text input query formulations. Generally a user may be able to type from 13 to 41 wpm (words per minute) and a good typist may type from 61 to 90 wpm. However people speak much faster than they type. For example, a speaker with normal speech rate produces about 150
wpm (Rebman et al. 2003, p.513). Thus speech recognition could generate text faster than using keyboard. In fact, the accuracy for PC-based speech recognition systems typically ranges from 80 to 99% (Alexander 1999, p.65). However for video material with frequent overlap of music and speech and large scale vocabulary (64000 words), the Informedia Digital Video Library studies revealed lower speech recognition accuracy, showing the error rate to be over 30%. (Witbrock & Hauptmann 1998; Hauptmann et al. 2002; 2003). This explains, a large-vocabulary systems with adverse environment (overlap of music and speech) have significantly lower recognition accuracy. Therefore in order to obtain higher accuracy, system applications that incorporate speech recognition should consider the opportunities and obstacles of speech recognition technologies (Shneiderman 2005, p.376).

With recent progress in speech technologies, a number of speech based methods have been explored in information retrieval. These can be classified into two categories which are (1) spoken document retrieval (SDR) and (2) spoken query retrieval (SQR) (Fujii et al. 2002, p.94). In the recent information retrieval literature, the retrieval of SDR using textual query has been well studied and documented. Various approaches in indexing the spoken document have been promoted by the SDR track of TREC (Garofolo et al. 1999, p.1). In contrast there have been relatively fewer studies in the retrieval of textual documents using spoken query (Chang et al. 2002, p.533).

Nowadays, information retrieval over the World Wide Web is no longer restricted to pure textual information. Millions of digital photos are available and users can find these digital photos by searching or following available hyperlinks by browsing through the World Wide Web. However, most of the available web pages only allow users to interact using graphical user interfaces which primarily uses a visual medium that requires a keyboard and mouse to navigate, and this can discourage several types of users, for example, those who lack motor skills to use a keyboard and mouse and find navigation troublesome. Therefore, an
alternative user interface such as graphical user interface with speech (S/GUI) or a multimodal user interface could give more flexibility to the users as speech input does not require use of hands, and thus enables users to carry out other actions and move freely. In addition, speech input also offers physically disabled people opportunities to use system applications more efficiently (Lahtinen and Peltonen 2005, p.87).

Speech interfaces can be divided into two categories (1) auditory user interfaces and (2) graphical user interfaces with speech. This research is concerned with (2) graphical user interface (GUI) with speech, or S/GUI (for Speech/GUI) (Weinschenk and Barker 2000, p.10). Sanders and McCormick (1993, p.169) noted when speech and visual modes are more effective to be implemented:

Speech mode is effective when:
- Information is short and simple
- Information is needed immediately (does not have to be remembered)
- Information is temporal in nature (refer to events over time)
- The message is a critical warning
- A verbal response is required
- The visual system of the person is already overextended
- The environment is not conducive to a visual display (for example the lighting is insufficient)
- The person needs to stay “dark adapted”
- The person needs to be moving continually

Visual mode is effective when:
- Information is complex and long
- Information needs to be remembered
- Information deals with spatial relationships; for instance maps
- The person’s audition is overextended
• The environment is noisy or creating noise in environment is not acceptable (Sanders and McCormick 1993, p.169).

Although speech recognition might be unstable across different groups of users, environments and time; ‘speech is the bicycle of user interface design and has an important role but can only carry a light load’ (Shneiderman 2005, p.375). Shneiderman (2005, p.376) noted that speech systems have some opportunities and obstacles:

Opportunities:
• When users have vision impairments
• When the speaker’s hands are busy
• When mobility is required
• When the speaker’s eyes are occupied
• When harsh or cramped conditions preclude use of keyboard

Obstacles to speech recognition:
• Increased cognitive load compared to pointing
• Interference from noisy environments
• Unstable recognition across changing users, environments and time.

Obstacles to speech output:
• Slow pace of speech output when compared to visual displays
• Ephemeral nature of speech
• Difficulty in scanning/searching

Sun Java (1998) noted that speech is well suited to some tasks. The following Tables 2.2 lists the characteristics that determine when speech input and output are appropriate choices.

52
Table 2.2: When is speech input and output appropriate? (Sun 1998)

<table>
<thead>
<tr>
<th>When is speech input appropriate?</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Use When...</strong></td>
<td><strong>Avoid When...</strong></td>
</tr>
<tr>
<td>1. No keyboard is available (e.g., over the telephone, at a kiosk, or on a portable device).</td>
<td>1. Task requires users to talk to other people while using the application.</td>
</tr>
<tr>
<td>2. Task requires the user's hands to be occupied so they cannot use a keyboard or mouse (e.g., maintenance and repair, graphics editing).</td>
<td>2. Users work in a very noisy environment.</td>
</tr>
<tr>
<td>3. Commands are embedded in a deep menu structure.</td>
<td>3. Task can be accomplished more easily using a mouse and keyboard.</td>
</tr>
<tr>
<td>4. Users are unable to type or are not comfortable with typing.</td>
<td></td>
</tr>
<tr>
<td>5. Users have a physical disability (e.g., limited use of hands).</td>
<td></td>
</tr>
</tbody>
</table>

When is speech output appropriate?

<table>
<thead>
<tr>
<th><strong>Use When...</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Task requires the user's eyes to be looking at something other than the screen (e.g., driving, maintenance and repair).</td>
</tr>
<tr>
<td>2. Situation requires grabbing users' attention</td>
</tr>
<tr>
<td>3. Users have a physical disability (e.g., visual impairment).</td>
</tr>
<tr>
<td>4. Interface is trying to embody a personality.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Avoid When...</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Large quantities of information must be presented.</td>
</tr>
<tr>
<td>2. Task requires user to compare data items.</td>
</tr>
<tr>
<td>3. Information is personal or confidential.</td>
</tr>
</tbody>
</table>

2.4.2 Studies related to multimodal user interface and applications

Numerous multimodal user interface systems and prototypes have been tested, built and have demonstrated the usefulness of multimodal interaction in various research domains. None, however, take into account the effects of the multimodal interfaces on the user search performance in an information retrieval system, particularly for a web based personal digital photo retrieval system. There have been a few studies comparing the effectiveness of speech input with other input channels in some applications domains.

Schmandt et al. (1990) performed an investigation using speech input and mouse input to control windows navigation in an X Windows system while allowing keyboard and mouse input for other tasks, such as direct manipulation interface in
application programs. The results indicated that there was no significant difference in speed between speech input and mouse input in navigating between exposed windows. Speech, however, was better than mouse when windows were partially or completely obscured. In addition, users of speech leaned to use more windows and allow more and a greater degree of windows overlap.

Pausch and Leatherby (1991a) conducted two different studies of the Macintosh graphical editor program named MacDraw (Claris, version 1.9.6). In the first study, the control group used mouse input modality while the experimental group used speech input to enter commands and mouse for pointing and selecting graphical objects. The finding showed that the experimental group that used speech and mouse input modalities decreased task completion time by 21.23 % as compared with mouse interaction. In the second study (Pausch and Leatherby 1991b), they performed a user study with the same program but used accelerator keys (keystrokes which have been bound to application command). The results showed a novice group of users who had not memorised the keyboard commands worked 9.92% faster than when using mouse input, whereas an advanced group of users who had memorised the command worked 14.51% faster. The findings concluded that voice or speech input provides a significant reduction in task completion time for graphical editor application when compared to the traditional alternatives.

Karl et al. (1993) conducted an experiment and demonstrated the advantages of using speech activated commands over mouse activated commands for word processing applications when, in both cases, the keyboard is used for text entry and the mouse for direct manipulation. Sixteen participants were trained to issue eighteen voice activated commands to perform four simple word processing tasks. The results showed that performance times for all tasks were significantly faster when using speech to activate commands as opposed to using the mouse. On average, the reduction in task time due to using speech was 18.67%. Overall, the participants reacted positively to using speech input and preferred it over the mouse for command activation, however, they were concerned about recognition
accuracy, the interference of background noise, inadequate feedback and slow response time (Karl et al. 1993).

Molnar and Kletke (1996) compared the effect on performance and satisfaction of a menu and a front-end voice interface to a Lotus 1-2-3 spreadsheet software package. The experimental task required subjects to solve a total of three business problems using the spreadsheet software package. The first group used Lotus 1-2-3 through the use of menus while the second group used Lotus 1-2-3 with a front end voice interface. The findings showed that the voice interface users overall required more time to complete the tasks and had less favourable attitudes about the software than the menu users. The mouse and menu options appeared to be more user-friendly than having to remember the appropriate voice commands.

Interactive maps are a promising application for multimodal interfaces. For example, a multimodal system for dynamic interactive maps which allows users to use speech and pen either separately or in a multimodal way to perform map-based tasks has been developed (Oviatt 1996). Oviatt et al. (1997) found that spoken and written modes in the system provide complementary semantic information rather than redundant information. Several experimental systems have investigated the use of multimodal user interface in the map and command post (Myers et al. 2002). Their studies show that using a multimodal map interface is more efficient than a conventional direct manipulation interface.

Karat et al. (1999) conducted a user performance evaluation and satisfaction in completion of a set of text creation tasks using three commercially available continuous speech recognition systems. These systems were IBM ViaVoice 98 Executive, Dragon Naturally Speaking Preferred 2.0, and L&H Voice Xpress Plus. The study also compared user performance on similar tasks using keyboard input. The findings indicated that users took a longer time in completing the tasks when
using speech recognition systems and several participants commented that keyboard entry seemed "much more natural" than speech for entering text.

Christian et al. (2000) compared voice control over mouse control for web browsing. A total of 18 subjects were used and all of them had significant experience using computers and web browsers, but none had any experience with voice browsers. Conversa web browser produced by Conversational Computing (http://www.conversa.com) was used in the study. The results of the experiment suggested speech recognition for input is less precise than using the mouse, and web hyperlinks that sound similar could accidentally be activated when using voice control. In addition voice commands do introduce cognitive overhead and add approximately 50% to the performance times for simple navigation tasks that are focused on rapid navigation through multiple links. Other additional results suggested that motor impaired users who speak English without an accent will be able to use voice control to navigate the World Wide Web. They did not need to be trained in the speech recognition software for their specific voice.

Information kiosks show potential applications for multimodal interfaces. A number of experimental systems have investigated multimodal user interfaces by adding speech input to interactive information kiosks (Raisamo 1998; Lamel et al. 2002). For example, Lamel et al. (2002) conducted a user evaluation of the Multimodal Multimedia Service Kiosk (MASK) prototype. The MASK has a touch screen for tactile input, loud speakers and microphones for speech input. The kiosk is able to provide train timetable and fare information, and simulated ticket purchases. With 200 subjects, user assessment trials were carried out to assess the users' performance with the prototype. The time to complete the transaction with the MASK kiosk is reduced by about 30% compared to that required for the standard kiosk, and the transaction success rate is 85% for novices and 94% once familiar with the system.
Multimodal interface for digital image retrieval has received attention recently. Käster et al. (2003) introduced a system for interactive standalone content based image retrieval. The system utilised a multimodal user interface that allows multimodal interaction combining mouse, keyboard, speech and touch screen. A database of 1250 digital images taken from the ArtExplosion image collection was considered in the study. Twenty participants were divided into four groups of five each with different input modalities: i) mouse; ii) mouse and speech; iii) touch screen; and iv) touch screen and speech were evaluated. The findings show that the participants using a touch screen device performed best and fastest while users of speech and touch screen were the slowest and least successful. The usability experiments which investigated the impact of multimodal interfaces in image retrieval showed that users well appreciate the multimodal interfaces for image retrieval.

More recently, Lahtinen and Peltonen (2005) presented an approach to develop speech interfaces to Unified Modeling Language (UML) tools. VoCoTo (VoiceControlTool) prototype was developed to evaluate the possibilities of a speech controlled user interface in the software engineering domain. The prototype is integrated with an existing Rational Rose CASE-tool, where the user can use speech control to widen the range of available tasks, or to make the usage faster and easier. Based on the performed tests, users who use the VoCoTo prototype completed all the class diagrams tasks faster. They also found that UML is a favourable domain for speech recognition and it can be used to enhance the usage of UML CASE-tools.

2.4.3 User interface engineering

This section deals with guidelines and techniques for designing multimodal user interfaces. Research into published sources of multimodal user interface guidelines, framework for multimodal interface, prototyping and evaluation paradigms are discussed.
2.4.3.1 Multimodal user interface guidelines

Many different researchers have provided guidelines for designing user interfaces. For example, International Business Machine (IBM) published the Common User Access (CUI) guidelines. Microsoft (1995) published guidelines for Microsoft Windows environment software. Marcus et al. (1995) provides graphical user interface (GUI) guidelines to facilitate the design of application under multiple GUI interface environments. Shneiderman (1998, pp.124-150) offered eight golden rules: (1) Strive for consistency; (2) Enable frequent users to use shortcuts; (3) Offer informative feedback; (4) Design dialog to yield closure; (5) Offer simple error handling (6) Permit easy reversal of actions; (7) Support internal locus of control and (8) Reduce short-term memory load.

A number of researchers have proposed guidelines related to multimodal user interface design. Weinschenk and Baker (2000, p.184) compared the work done by Shneiderman (1998) and Nielsen (1994) and added to their contributions. From these sources they outline 20 design principles which can be used for multimodal user interface design. The design principles include user control, human limitation, modal integrity, accommodation, linguistic clarity, aesthetic integrity, simplicity, predictability, interpretation, accuracy, technical clarity, flexibility, fulfilment, cultural propriety, suitable tempo, consistency, user support, precision, forgiveness and responsiveness. Reeves et al. (2004, pp.57-59) discuss six main categories of guidelines to establish principles for multimodal interaction design. The categories are: i) Requirements specification; ii) Designing multimodal input and output; iii) Adaptivity; iv) Consistency; v) Feedback; and vi) Error prevention/handling.

Although there are many different sets of user interface guidelines, some researchers have cautioned against relying too much on guidelines. Some of the guidelines can be viewed as being contradictory (Myers 1994, p.78). Tetzlaff and Schwartz (1991, p.329) noted that designers had significant difficulty in
interpreting the design guidelines. Grudin (1989, p.1164) argued that too much emphasis on user interface guidelines can divert focus from obtaining a real understanding of the users' tasks.

Technically, to design an effective multimodal user interface, Sun Java notes that the development of a multimodal user interface has to consider three design issues which include (Sun 1998): i) Feedback & Latency ii) Prompting and iii) Handling Errors.

- Feedback & Latency - In speech-only systems, performance delays can cause confusion for users. Speech interaction sometimes causes people to forget what they have just said. In multimodal interface however, graphic display can show the user the state of the recogniser that speech-only interface cannot.

- Prompting - In multimodal systems, prompts can be spoken or printed. It depends greatly on the content and context of the application. The computer should not speak out loudly if privacy is an issue. In contrast, some spoken output can enable eyes-free interaction and can provide the user with the feeling of having a conversation. With a screen, prompts usually involve providing the user with a list of applicable spoken commands. Another strategy is to let users speak the text they see on the screen, for example, the menu text or button text.

- Handling Errors - With multimodal interfaces, the system can display speech recognition results and make it easier for error detection. For example, if a speech recognition error occurs, users can see what the recogniser thinks was said and correct any errors. An application should not assume that users will always catch errors. Therefore a filtering process for unexpected input is helpful. This mechanism also can be one of
the indications for users to switch to a different input modality if the modality (such as speech recognition input) is not working reliably.

2.4.3.2 Framework for multimodal interface

A number of researchers have proposed frameworks for multimodal interface. Krahnstoever et al. (2002) describes a framework for designing a natural multimodal system which specifically combines speech and gesture. Several applications related to the retail and entertainment industry are described that have been implemented using this framework. Flippo et al. (2003) describe a multimodal interface framework that is reusable across applications and modalities. Using the frameworks, a multimodal interface for map application was created.

Despite the fact that multimodal interfaces are not yet common in the World Wide Web environment, the World Wide Web Consortium (W3C) has produced a document of Multimodal Interaction Framework. The framework identifies the major components for every multimodal system and each component represents a set of related functions. This Framework is intended as a basis for developing multimodal applications in terms of markup, scripting, styling and other resources (Larson et al. 2003). Figure 2.5 illustrates the basic components of the W3C multimodal interaction framework.

![Figure 2.5: The W3C multimodal interface framework (Larson et al. 2003)](Image)
The input component uses multiple input modes such as audio, speech, handwriting, keyboarding, and other input modes. The output component uses one or more modes of output, such as speech, text, graphics, audio files, and animation. The interaction manager maintains the interaction state and context of the application and responds to inputs from component interface objects and changes in the system and environment. Then the interaction manager manages these changes and coordinates input and output across component interface objects. The session component provides an interface to the interaction manager while the system and environment component enables the interaction manager to respond to changes in device capabilities, user preferences and environmental conditions (Larson et al. 2003).

To support multimodal interaction in the World Wide Web environment, The SALT forum group (Cisco Systems, Comverse, Intel, Philips, Scansoft and Microsoft) has published the Speech Application Language Tags (SALT) specification that enables multimodal and telephony-enabled access to information, applications, and Web services from PCs, telephones, tablet PCs, and wireless personal digital assistants (PDAs). SALT is an established standard which extends the capabilities of existing mark-up languages such as HTML, XHTML, and XML. Multimodal access with SALT enables users to interact with an application in a variety of ways: they will be able to input data using speech, a keyboard, keypad, mouse and/or stylus, and produce data as synthesized speech, audio, plain text, motion video, and/or graphics. Each of these modes will be able to be used independently or concurrently (SALT Forum 2002).

W3C has recently produced the First Public Working Draft of Multimodal Architecture and Interfaces for review by W3C Members and other interested parties (Barnett 2005). The document describes the architecture of the Multimodal Interaction (MMI) framework and the interfaces between its constituents. Meanwhile, The W3C Multimodal Interaction working group has aims to develop specifications to enable access to the Web using multimodal interaction through
EMMA: Extensible MultiModal Annotation markup language. Components that generate EMMA markup: (1) Speech recognizers; (2) Handwriting recognizers; (3) Natural language understanding engines; (4) Other input media interpreters (e.g. DTMF, pointing, keyboard) and (5) Multimodal integration component (Johnston et al. 2005).

2.4.3.3 Prototypes

A prototype can be used in order to test various aspects of the design as well as illustrate ideas or features of a system. It also can be use to get user feedback at the early stages of system development (Wikipedia 2005). According to Preece et al. (2002, p.241) "prototype is a limited representation of a design that allows users to interact and explore its suitability". Prototypes can range from extremely simple sketches to a complex piece of software (Preece et al. 2002, pp. 240-241). For software interface design including web applications, prototypes can be used to observe the content, aesthetics, and interaction techniques from the perspective of designers, clients, and users (Walker et al. 2002, p.661). Prototyping can be divided into low-fidelity prototyping and high-fidelity prototyping (Rudd et al. 1996, p.76).

A low fidelity prototype is one that is sketchy, incomplete but does look very much like the final system. It is usually used in the early stages of the development, during conceptual design for example, in order to quickly produce the prototype and test the concept. Low fidelity prototypes are inexpensive and can be built fast, however, they have limitations in showing navigation and flow, overlook some design problems and leave low level design decisions to programmers (Preece et al. 2002, p.243). A high fidelity prototype is quite close to the final system with lots of detail and functionality (Preece et al. 2002, p.245). Users are able to examine in detail and make strong conclusions about how behaviour will relate to use of the final system. Further advantages and disadvantages of the types of prototyping are listed in Table 2.3.
Table 2.3: Relative effectiveness of low-fidelity prototype vs. high-fidelity prototypes (Rudd et al. 1996, pp. 84).

<table>
<thead>
<tr>
<th>Prototype type</th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low-fidelity prototype</td>
<td>• Lower development cost.</td>
<td>• Limited error checking</td>
</tr>
<tr>
<td></td>
<td>• Evaluate multiple design concepts.</td>
<td>• Poor detailed specification to code.</td>
</tr>
<tr>
<td></td>
<td>• Useful communication device.</td>
<td>• Facilitator-driven</td>
</tr>
<tr>
<td></td>
<td>• Address screen layout issues.</td>
<td>• Limited utility after requirement</td>
</tr>
<tr>
<td></td>
<td>• Useful for identifying market requirements.</td>
<td>established.</td>
</tr>
<tr>
<td></td>
<td>• Proof-of-concept.</td>
<td>• Limited usefulness for usability</td>
</tr>
<tr>
<td></td>
<td></td>
<td>tests.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Navigational and flow limitations</td>
</tr>
<tr>
<td>High-fidelity prototype</td>
<td>• Complete functionality.</td>
<td>• More expensive to develop</td>
</tr>
<tr>
<td></td>
<td>• Fully interactive.</td>
<td>• Time-consuming to create</td>
</tr>
<tr>
<td></td>
<td>• User-driven.</td>
<td>• Inefficient for proof-of-concept</td>
</tr>
<tr>
<td></td>
<td>• Clearly defines navigational scheme.</td>
<td>designs.</td>
</tr>
<tr>
<td></td>
<td>• Use for exploration and test.</td>
<td>• Not effective for requirements</td>
</tr>
<tr>
<td></td>
<td>• Look and feel of the final product.</td>
<td>gathering.</td>
</tr>
<tr>
<td></td>
<td>• Serves as a living specification.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Marketing and sales tool.</td>
<td></td>
</tr>
</tbody>
</table>

A few studies compared the effectiveness of low-fidelity prototype and high-fidelity prototype. Rettig (1994) identified inherent high-fidelity prototyping problems as: (1) They take too long to build; (2) Reviewers and testers tend to comment on superficial aspects rather than content; (3) Developers are reluctant to change something they have crafted for hours; (4) A software prototype can set expectations too high and (5) Just one bug in a high-fidelity prototype can bring the testing to halt. Therefore he argues that more projects should use low-fidelity prototyping. Walker et al. (2002, p.661) compared user testing with low-fidelity and high-fidelity prototypes in both computer and paper media through task-based user tests of sketched (low-fidelity) and HTML (high-fidelity) website prototypes. They found that low-fidelity and high-fidelity prototypes are equally good at uncovering usability issues. Designers should choose whichever medium and level of fidelity suit their practical needs and design goals. High-fidelity prototyping is useful for selling ideas and testing out technical issues. However low-fidelity
should be actively encouraged for exploring content and structure issues (Preece et al. 2002, p.243).

2.4.3.4 Evaluation paradigms

Various evaluation paradigms have been created and promoted. Each paradigm has particular methods and techniques associated with it. Preece et al. (2002, pp.340-343) identify four core evaluation paradigms that can be used for user interface evaluation: (1) “quick and dirty” evaluation; (2) field studies (3) predictive evaluation; and (4) usability testing;

(1) Usually the ‘Quick & dirty’ evaluation is descriptive and informal. It describes the common practice in which evaluators informally get feedback from users or consultants to confirm that their ideas are in-line with users’ needs and are liked. One of its advantages is that the evaluation session of “quick and dirty” can be done at any time on a short timescale. The emphasis is on fast input to the design process rather than carefully documented findings.

(2) The aim of field studies evaluation is to understand what users do naturally (natural behaviour) and emphasise how the technology could impact on them. During the evaluation the evaluator observes users performing their usual job tasks in the context of their natural settings.

(3) In predictive evaluation, users are not required to be present. During the evaluation session, the experts apply their knowledge of typical users which is often guided by heuristics aspects.
(4) Usability testing is a means for measuring users' performance on prepared tasks in a controlled setting for which the system was designed. The key to usability testing is to observe users using the system in as realistic situation as possible to discover the users' performance with the proposed system. During usability testing, user activities which are based on tasks are watched and recorded. Normally after the users have finished all of the tasks, user satisfaction questionnaires and interviews can be used to draw out users' opinions regarding their experiences with the proposed system features.

Usability testing is a research tool that can be used to focus on the observation of actual user behaviour. Its roots are in classical experimental methodology. The range of usability testing is considerable, from experiments with large sample sizes and complicated test design to very informal qualitative studies with only a single participant. Each usability testing approach has different goals and purposes (Rubin 1994, pp. 25-26). The issue of the number of participants in usability testing is one of intense interest and discussion in the usability community (Dumas and Redish 1999, p.135). To feel comfortable in seeing problems, three participants for each group is probably an absolute minimum (Nielsen and Molich 1990, p.249). Virzi (1992, p.457) found that 80% of usability problems were detected with between 4 and 5 participants and 90% were detected with 10 participants. For web site usability, some studies argue that more than five participants are needed to find 85% of the problems (Spool and Schroeder 2001, p.285). Using a mathematical model, Nielsen (2000) stated that 5 users are enough to catch 85% of the problems on practically any web site. The mathematical formula only holds for comparable participants who will be using the web site in fairly similar ways. Additional participants are only needed when a website has several highly distinct groups of users.
In conducting a research study however, the number of participants must be of sufficient size to measure statistically significant differences between groups (Spyridakis 1992, p.28). For a web application with different types of users (e.g., novices and experts), a minimum of six participants per condition was recommended to be utilised (Koyanl et al. 2003 p.7). Dumas and Redish (1999, p.136) proclaim that the number of participants in a test depends on how many sub groups are covered, how much time and money can be used and how important it to be able to compute statistically significant results.

Table 2.4 shows the characteristics of each evaluation paradigm adapted from Preece et al. (2002, p.344).
Table 2.4: Characteristics of evaluation paradigms (Preece et al. 2002, p.344)

<table>
<thead>
<tr>
<th>Issues</th>
<th>Evaluation paradigms</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>“Quick and dirty”</td>
</tr>
<tr>
<td>Roles of users</td>
<td>Natural behaviour</td>
</tr>
<tr>
<td>Who controls</td>
<td>Evaluators take minimum control</td>
</tr>
<tr>
<td>Location</td>
<td>Natural environment or laboratory</td>
</tr>
<tr>
<td>When used</td>
<td>Any time you want to get feedback about a design quickly.</td>
</tr>
<tr>
<td>Type of data</td>
<td>Qualitative, informal descriptions</td>
</tr>
<tr>
<td>Feedback into design by..</td>
<td>Sketches, quotes, descriptive report.</td>
</tr>
<tr>
<td>Philosophy</td>
<td>User-centred, highly practical approach</td>
</tr>
</tbody>
</table>
2.5 Summary

The literature review for this study can be summarised from the following perspectives: (a) image retrieval evolution; (b) user studies in image collections; (c) user interface support for digital photo retrieval; and (d) multimodal user interface.

Firstly, studies of image retrieval evolution found that the early generation of image retrieval involved storing the images in a database with textual metadata including various annotations that describe the content of the images. The most significant achievement in most existing image retrieval systems is the efficiency of the image retrieval engine. Currently there are a number of methods used in image retrieval systems. The first method is text-based. This method is still most important and the dominant technique in image retrieval. The second method is by query based on visual content. Most current image retrieval studies emphasise developing a more efficient image retrieval system engine. Not much research however, has been done to support multimodal user interface elements in image retrieval. Until now, reported research aimed at providing this has been limited.

Secondly, most studies on user queries in image collections have been concerned with problems involved with what the user needs, how people search for images, what attributes of an image are to be indexed and how images are used. Characteristics of user queries on image collection domains have been reported by Enser and McGregor (1992), Keister (1994), Hasting (1994), Jorgensen (1998), Ornager (1997), Enser & Armitage (1997), Markkula & Sormunen (1998, 2000), Rasmussen & Choi (2002), Eakins et al. (2004) and Othman (2005). Past and present research has focused on the needs of specific image user groups, searching behaviour and image attributes. The studies above confirm that some image users have very specific needs. Most related studies suggested that image retrieval systems should provide searching of image content by four main categories such as Who?, What?, When?, and Where?. Content based retrieval
(CBIR) was suggested to be most appropriate at the browsing stage. However, there has been very little attention given to how people retrieve their personal digital photos. Much of the work is concerned with the technology for managing and sharing personal digital photos.

Studies of CBIR show that, it is useful in certain contexts but for most users it has not yet become a highly important means of searching (Eakins et al. 2004). This is clearly related to its limited capabilities in using colour, shape and texture attributes that are not immediately obvious to general users. But it is certain that this technology is maturing and so should be strongly considered in the development of image retrieval systems.

Thirdly, from the studies of user interface support for digital photo retrieval, it was found that there is a spectrum of digital photo systems ranging from programs that just transfer images from the camera to the computer to those that manage large collections of images. Some of the systems provide limited visual content based image retrieval and query by keywords. Only a few provide a web based platform. None of the systems used multimodal user interface for retrieving personal digital photo in a web based environment.

Personal digital photo retrieval systems with text and visual applications are naturally hand and eye busy applications. Furthermore, the user has to understand and follow application menu flows or steps. These tasks may involve some problems for some users. There are risks of spelling errors and is time consuming in typing textual annotations for keyword searching. In the case of visual retrieval tasks, users have to browse their digital photo collection in order to select their visual example. This is not necessary with speech input or commands. If the user knows what to say, he or she simply speaks the input command, such as the photo file reference without having to scroll or browse every angle of the pages to get their visual example digital photo.
Finally, reviews of multimodal user interface, including speech interface and information retrieval; studies related to multimodal user interface and application; and user interface engineering related to multimodal systems found that theoretically a multimodal user interface should provide a flexible interface for digital photo retrieval.

In the case of multimodal user interfaces, the two most relevant studies for this thesis are Rodden and Wood (2003) and Käster et al. (2003). Although these two studies have some similarity, Rodden and Wood (2003) used speech annotation but did not include speech retrieval in a study of personal digital photos. Käster et al. (2003) made use of speech retrieval for stand-alone content based image retrieval for expert digital image users.

In the studies reported, speech has been mainly used for annotation in recent implementations. This is in an attempt to simplify the process of adding text to images for organisation purposes. Searching, however, is done by typing in the term to match the speech annotation (Chen et al. 2003). Apart from Käster et al. (2003) (who implemented speech retrieval in a stand-alone CBIR system aimed at the professional image searcher) no system has been developed and evaluated that uses speech for searching personal digital photos on the web which provides this research with an opportunity to create such a system.

In the case of multimodal applications in general, results sometimes conflict. For example, some early studies showed that speech modality decreased task completion time (Pausch and Leatherby 1991a; Karl et al. 1993; Lamel 2002; Lahtinen and Peltonen 2005) whereas others (Molnar and Kletke 1996; Karat et al. 1999; Christian et al. 2000) needed more time to complete the tasks. It is clear that there is still insufficient data to indicate the success of multimodal user interfaces specifically S/GUI in task performances. Therefore, it is attractive to explore the needs, analyse and develop a flexible user interface for a web based personal digital photo retrieval system. Consequently, the goal of this study is to
fill some of the many gaps that remain in the field of multimodal user interface and web based digital photo retrieval system.
CHAPTER 3

METHODOLOGY

3.0 Introduction

This chapter discusses the research methodology, research design and procedures used in this thesis.

- Section 3.1 discusses the research methodology, including several related system development lifecycles.

- Section 3.2 discusses the research design and procedures, including a small scale evaluation, user tasks description, low fidelity prototype and the prototype evaluation.

- Section 3.3 discusses the assumptions underlying this research.

3.1 Research methodology

Research in general can be divided into several types. Functionally, it can be divided into: descriptive, historical, correlational, developmental, ex-post-facto and experimental (Sahibuddin 1999, p.139). Descriptive research is concerned with determining the nature and degree of existing conditions. Historical research is concerned with determining, evaluating and understanding past events for the purpose of better prediction of the future. Correlational research is concerned with discovering the degree of relation between two or more variables. Developmental research is concerned with the patterns of growth as a function of time. Ex-post-
facto research investigates possible cause and effect relationships by observing some existing consequences and searching back through the data for plausible causal factors. Experimental research is a scientific investigation in which one or more independent variables are manipulated and the dependent variable or variables are observed. The nature of this research is experimental and is explained further in the research design in Section 3.2.

3.1.1 System development lifecycle models

Throughout the study, the aim was to design and develop an appropriate multimodal user interface which offers a flexible user interface for a web based personal digital photo retrieval system. The appropriate framework for this type of research is System Development Life Cycle (SDLC) methodology. "SDLC is the overall process of developing information systems through a multi step process from investigation of initial requirements through analysis, design, implementation and maintenance" (Kay 2002). Although there are many different SDLC models and methodologies, each, in general, consists of a series of defined steps or stages and these are intended to complement each other. Several system development lifecycles in the related field have been reviewed. Among them are the following:

- Waterfall (Royce 1987).
- Spiral (Boehm 1988).
- Star (Hartson and Hix 1989).
- Logical user-centred interaction design (LUCID) (Kreitzberg 1996, p.119).
- Usability Engineering Lifecycle (Mayhew 1999).
- Lifecycle model for interaction design (Preece et al. 2002).

For many years, the Waterfall model formed the basis of most software development. The model was introduced in 1970 by W. W. Royce and is seen as
flowing steadily through the phases of requirements: analysis, design, implementation, testing (validation), integration, and maintenance (Royce 1987). Waterfall development, however, does not allow for much reflection or revision. Once an application is in the testing stage, it is very difficult to go back and change something that was not well-thought out in the concept stage.

The Spiral model assumes multiple iterations of the waterfall model and is intended to solve the problem of adapting to change more readily. The spiral model was defined by Boehm in 1988. Each phase starts with a design goal and ends with the client reviewing the progress thus far. The efforts of analysis and engineering are applied at each phase of the project (Boehm 1988). Nevertheless when strictly applied, there is a possibility that later iterations will invalidate earlier work. Furthermore, not all analysis and design occurs before construction. The spiral model is favoured for large, expensive, and complicated projects (Satzinger et al. 2004, p.683).

The Rapid Application Development (RAD) approach attempts to take a user-centred view through user involvement in the Joint Application Development (JAD) phase and is a response to the inappropriate nature of the waterfall lifecycle models. Through JAD workshop platform, users and developer are able to draw up the system requirements (Wood and Silver 1995, p.190). Tools and techniques such as prototyping, fourth-generation programming languages, CASE tools and object oriented analysis, design and development have been frequently associated with the RAD approach (Satzinger et al. 2004, p.672). However, RAD has not provided flexible processes for studying the users and understanding their needs. User participation sessions are limited during the JAD workshop and iteration process is limited to the design and build phase.

STAR methodology does not specify any ordering of activities. The activities are highly interconnected and extremely flexible. Therefore it is difficult to control the development process. Logical user-centred interaction design (LUCID)
methodology is intended to be prescriptive and involved with rigid business roles which are not appropriate in this study (Kreitzberg 1996, pp.119-120).

In the Usability Engineering Lifecycle (UEL), the iteration process between the design/ testing/ development phase and the requirements analysis phase occurs only after the conceptual model and the detailed designs have been developed, prototyped and evaluated one at a time. However, there is a need to return to identifying user needs even at the level of the conceptual model and when the detailed designs have been developed.

The development of the FlexPhoReS prototype is based on the system development Lifecycle model for interaction design proposed by Preece et al. (2002). The model was deemed most relevant to the scenario and scope of the study. The methodology offered a flexible and transparent way for interactive user interface development. Other reasons why the Lifecycle model for interaction design was chosen are as follows:

a) The model is based on software engineering and human computer interaction lifecycles.

b) It applies a user-centred software design.

c) It incorporates iteration and encourages a user focus.

d) It is focused on the "front-end" of software: assessing user requirements and developing a look, feel and navigational flow which supports the functional requirements of the system.

e) It is not intended to be prescriptive. Therefore the flexibility of the model is suitable for this study.

There are four basic activities in the Lifecycle interaction design model (Preece et al. 2002, p.186): (1) Identifying needs and establishing requirements; (2) Developing alternative designs; (3) Building interactive versions of the designs; and (4) Evaluating designs. In the first activity, identifying needs and establishing
requirements is used to understand as much as possible about the users and their requirements. In an attempt to meet the needs and requirements that have been identified, alternative designs are generated. Then interactive versions of the designs are developed and evaluated. From the evaluation activity, it may be necessary to return to identifying needs and requirements, or it may be possible to go straight into redesigning activity. The development ends with an evaluation activity that ensures the final version meets the usability criteria. Figure 3.1 shows the Lifecycle model for interaction design proposed by Preece and colleagues.

![Lifecycle model for interaction design](image)

Figure 3.1: Lifecycle model for interaction design (Preece et al. 2002, p.186)

This model was found more relevant to the scope of this study compared to other system development lifecycle models, even though it has some limitations. In general, user interface design is a multidisciplinary field coming from human computer interaction (HCI) which involves the study of the interaction between people (users) and computers. The development of the user interface may employ several different groups of active participants such as the system user, developer and content provider. The Lifecycle model for interaction design encourages a user focus and is not intended to be prescriptive, however when compared with
RAD, the model has not always provided clear processes for developing a system that involves groups of active participants in the design and development process, but this is more appropriate for a large complex system unlike the current study.

3.2 Research design

The research was conducted in three stages. The first stage involved developing the model for FlexPhoReS for personal digital photo collections. The second stage concentrated on the development of the prototype based on the model proposed (this included the Lifecycle model’s second and third activities). The third stage evaluated the prototype to find its acceptability, subjective satisfaction and search performance with the digital photo users.

For the modelling stage, identifying users’ needs and establishing requirements were addressed. The aim was to understand as much as possible about the users and produce a stable set of requirements from the needs identified. In order to achieve the aim, data were drawn from the following activities:

- Studies related to image retrieval evolution, user studies in image collections, user interface support for digital photo retrieval and multimodal interface (reported in detail in Chapter 2).

- A data gathering exercise involving structured interviews was carried out with a small group of digital photo users on the subject of how they organise and retrieve their digital photos (reported in detail in Section 3.2.1 and in Chapter 4).

The data gathering exercise conducted was used to gather information and to provide additional input into the construction of the model. Both the literature review studies and interviews formed the basis of the conceptual model, through task description of the proposed system model. The model produced, in turn, was
used to design the prototype. The initial design was sketched on paper and then mocked-up for an interactive version of a low fidelity prototype.

For the second stage, an interactive version of the prototype was built based on the task descriptions. The prototype evolved over several iterations. Initially a low fidelity prototype was generated and evaluated informally in order to get some feedback from users. This low fidelity prototype was used to clarify users' needs and requirements related to the content and structure, screen design and learning how to use the system. The informal evaluation was input to the redesign and development of the high fidelity prototype which was implemented using the five enabling technologies identified in Chapter 5.

The third stage of this research concentrated on the evaluation of the prototype. After conducting a pilot test, usability testing was applied in order to measure the users' search performance, subjective satisfaction and acceptability of the prototype. Controlled experiments were carried out in the evaluation session to compare two or more conditions to see if digital photo users perform better in one condition than in the other. There was only one group of digital photo users involved. The group was exposed to one treatment, namely evaluating the FlexPhoReS prototype. Data was gathered after the participants interacted with the prototype system model and filled in an evaluation form questionnaire. The results from the third stage were analysed to get the information on search performance, subjective satisfaction and the acceptability of the prototype and thus the system model.

3.2.1 Stage 1: Pre-development needs evaluation exercise

A small-scale needs evaluation exercise was undertaken to inform the development of the prototype. A structured interview-style data collection method was more appropriate than administering a questionnaire, as the participants were encouraged to verbalise their views, feelings, and experiences fully. Also,
because of their varying knowledge of digital photo technology, it was predicted that they might have problems understanding certain technical terms related to the subject. The approach and scale was similar to that undertaken by Rodden (1999). A total of seven digital photo users, individuals between twenty and forty years of age who had more than 100 personal digital photos were selected to participate in the study. All of the respondents used a digital camera to capture their digital photos and were familiar with the use of the Internet and personal computers, especially with Windows based applications. Three respondents were experienced in using a digital photo management system included in the Windows XP operating system, while other respondents had not used any specific digital photo management system to manage their digital photos. In addition, each was familiar with the content of their photos which related to various aspects of their life and memories. In the interview sessions, each was asked about their current practice and problems. They then gave their opinions as to how they retrieve and organise their personal digital photos. All of the respondents were asked the same set of questions (Appendix 1) in the interview sessions. The structured interview schedule was divided into five sections. The first elicited the respondent's background knowledge on using computers, the internet and digital photo capture devices. The second section focused on digital photo management system experiences. The remaining three sections focused on their current practice in storing, sharing and retrieving their digital photos. In these sections a number of questions were asked for the participants' agreement or disagreement in relation to security, annotation, visual similarity and speech interface. These questions were intended to stimulate the participants' minds and make a preliminary investigation in this area in order to explore the potential acceptability of the proposed prototype features. The interviews were recorded in order to gather a detailed sense of what users said during the interview sessions. The results are reported in Chapter 4 Section 4.1.1.1.
3.2.2 Stage 2: User Task description

A use case task description was used to describe user tasks with the prototype. It focused specifically on the interaction between the user and the prototype system. The tasks were chosen by identifying users' needs from stage 1 and establishing requirements which in turn generated a list of possible user interactions in retrieving digital photos. The use case also was described graphically through a use case diagram which simplified the interaction between the user and the proposed system.

3.2.3 Low fidelity prototyping development and evaluation

As part of the Lifecycle model for interaction design, at an early stage of development, researchers recommend that an initial evaluation should be done using a low fidelity prototype (Preece et al. 2002, p.243). Through a low fidelity prototype evaluation, potential usability problems can be detected at a very early stage in the design process before any programming code has been written. Based on the task description, the FlexPhoReS paper prototype (low fidelity prototype) was developed. The intention was to clarify requirements and enable draft interaction designs, screen designs and learning how to use the system to be simulated and tested. Rough screen design was used for the initial design of each individual screen. Each screen was pinned on large A1 paper with a logic flow diagram as for interaction design (Appendix 3). Microsoft PowerPoint with audio output was used in designing the low fidelity FlexPhoReS prototype (Appendix 4).

As recommended for usability testing (Chapter 2.4.3.4), three subjects participated in the low fidelity prototype informal evaluation. All of them were digital photo users aged between 17 and 40, computer literate and experienced in using Internet and web applications. The participants were asked to carry out a realistic task. Each participant simulated pointing and clicking using a pencil,
simulated typing by writing on paper and used their voice for speech input while the researcher clicked an appropriate screen for viewing the outcomes of the user task and explained what happened for each selected task. After completing the tasks, they were asked to give feedback related to the content and structure, screen design and learning how to use the system through a structured interview. A copy of the structured interview questions and the low fidelity FlexPhoReS prototype are available in Appendix 2 and Appendix 4.

3.2.4 Stage 3: FlexPhoReS evaluation

When the prototype was built the purpose of the evaluation was to measure the acceptability and subjective satisfaction of the prototype user interface and to understand whether the participants could use the prototype user interface that employed two input modalities (1) mouse and keyboard; and (2) mouse and speech, equally well in performing photo retrieval tasks from the photo collection. As described in Section 2.4.3.4, usability testing was chosen as the most appropriate evaluation method (Preece et al. 2002, p.344) for this prototype.

3.2.4.1 Pilot test

The objective of the pilot test was to "debug" the equipment, software, materials and procedures that were used for the full evaluation experiment. Two subjects participated in the pilot study. Both were digital photo users, and had experience in designing and evaluating a variety of software interfaces and services. Based on their feedback and observations, several changes were made to the evaluation materials and procedures. It was decided:

- To provide additional time for free exploration of the prototype in order to let users familiarise themselves with the FlexPhoReS prototype user interface.
• For both input modalities, to provide additional tasks for retrieving photos by browsing.

• For both input modalities, to provide additional tasks for retrieving photos by keyword searching.

• To use different photo examples with each input modality for retrieving photos by visual example.

• To make several other minor adjustments to the evaluation procedures, such as reminding the users to wear the headphones during the evaluation session.

3.2.4.2 Participants

The major evaluation was carried out with a total of 20 participants of digital photo users during the months of February and April 2005 (see Appendix 8). All of the participants had digital photo collections. None had taken part in any previous related FlexPhoReS tests. This was because previous experience would have familiarised them with the evaluation tasks. All of the participants had considerable experience with computers and were experienced in using web based digital photo retrieval features. Some of them had used these system features very often (more than 5 times per month).

3.2.4.3 Experimental procedures

A notice announcing the evaluation was placed in the ‘For Sale and Wanted’ Notice Board on the Loughborough University website asking for volunteers for the FlexPhoReS evaluation (http://www.lboro.ac.uk/news/notices/index.html). All interested participants were contacted through email or by telephone. The test took place in a special laboratory set up for this purpose in the Department of Information Science at Loughborough University. Participants came at a fixed time
for the evaluation. They were asked to fill in a recruitment questionnaire which assessed their computer and related system experience, including their experience with online digital photo retrieval features, as well as the number of digital photos in their collection, age, gender and educational background. A copy of the recruitment questionnaire is available in Appendix 5.

At the start of the evaluation, each participant was encouraged to go through a brief speech recognition training profile in order to maximise speech recognition accuracy. Each participant was then given a short description of the experimental procedures (Appendix 6) of the session that would be followed. With a user ID and password, each participant was given 20 minutes for free exploration of the prototype. During this period, each participant was encouraged to talk about what they were looking at and what they were thinking. They could also ask any related questions about the prototype user interface. The objective was to familiarise them with the prototype interface so they felt comfortable in performing the actual tasks.

After the participants filled in their recruitment questionnaire and went through the free exploration of the FlexphoReS prototype, the set of tasks was given to them and they were asked to work on their own. If any task took more than 15 minutes to complete, they were asked to stop and proceed to the next task. If they felt that they were unable to complete a task and wanted to move on, this was allowed. They were also free to consult the online help available in the prototype. After completing all the tasks, they were asked to fill in a questionnaire on photo retrieval tasks, subjective satisfaction, suitability and flexibility of the prototype user interface. If they had any questions, they could ask at any time during the test session.
3.2.4.4 The tasks

There were 10 tasks which were divided into two different categories. Each participant was required to complete all the tasks. The first set of tasks (Task 1 to 6) was used when participants were using mouse and keyboard input modalities (see Table 3.1).

<table>
<thead>
<tr>
<th>Task</th>
<th>Description</th>
</tr>
</thead>
</table>
| **Task 1** | Login into FlexPhoReS system  
A. Open FlexPhoReS http://localhost/cgi-bin/main.asp or press F2  
B. Use the given user id and password to login into the system |
| **Task 2** | Accessing and learning the general system help  
A. Please access the prototype Help component.  
B. Read and try to understand the contents.  
C. On that page, please tick the box when you understand the help contents |
| **Task 3** | i) Accessing and learning the 'What can I say' component for control/browse  
A. Please access the 'What can I say' component for control/browse.  
B. Read and try to understand the contents.  
C. On that page, please tick the box when you understand the contents  

ii) Accessing and learning the 'What can I type/say' component for keyword searching  
A. Please access the 'What can I type/say' component for keyword searching.  
B. Read and try to understand the contents.  
C. On that page, please tick the box when you understand the contents  

ii) Accessing and learning the 'What can I type/say' component for visual example searching  
A. Please access the 'What can I type/say' component for visual example searching.  
B. Read and try to understand the contents.  
C. On that page, please tick the box when you understand the contents |
| **Task 4** | Retrieve photos by browsing  
A. Please retrieve photos based on 'snowing' (event)  
B. Please retrieve photos based on 'castle' (event)  
C. Please retrieve photos based on 'swimming' (event) |
| **Task 5** | Retrieve photos by keyword searching  
A. Please retrieve photos using 'castle'  
B. Please retrieve photos using 'snow'  
C. Please retrieve photos using 'camp' |
| **Task 6** | Retrieve photos by visual example searching  
A. Please retrieve other photos which have a visual similarity to the photo given below: |
Task 1 consisted of logging into FlexPhoReS system. Tasks 2 required each participant to access and study the general system help, while Task 3 required participants to access more specific help components for each of the retrieval strategies. Task 4 required participants to retrieve photos by browsing. Task 5 required participants to retrieve photos by keyword searching and Task 6 by visual example searching.

The second set of tasks (Task 7 to Task 10) was used when participants were using mouse and speech input modalities (see Table 3.2).

Table 3.2: Mouse and speech input modalities tasks

<table>
<thead>
<tr>
<th>Task</th>
<th>Description</th>
</tr>
</thead>
</table>
| 7    | Login into FlexPhoReS system  
|      | A. Open FlexPhoReS http://localhost/cgi-bin/main.asp or press F2  
|      | B. Use the given user id and password to login into the system  |
| 8    | Retrieve photos by browsing  
|      | D. Please retrieve photos based on 'snowing' (event)  
|      | E. Please retrieve photos based on 'castle' (event)  
|      | F. Please retrieve photos based on 'swimming' (event)  |
| 9    | Retrieve photos by keyword searching  
|      | D. Please retrieve photos using 'castle'  
|      | E. Please retrieve photos using 'snow'  
|      | F. Please retrieve photos using 'camp'  |
| 10   | Retrieve photos by visual example searching  
|      | A. Please retrieve other photos which have a visual similarity to the photo given below: |
|      | ![Photo](image_url)  |

Task 7 required participants to login into the FlexPhoReS system. Task 8 was to retrieve photos by browsing. Task 9 involved retrieving photos by keyword searching and Task 10 involved retrieving photos by visual example searching.
3.2.5 Data collection

The evaluation used a combination of data collection methods. These were: computer screen recording and a questionnaire.

3.2.5.1 Computer screen and speech recordings

TechSmith's Camtasia version 3.0.2, show-and-tell communication software (TechSmith 2002), was used to capture each user's entire action with the FlexPhoReS prototype session. Because the Camtasia is essentially 'invisible' it was not expected to influence users' normal tasks and searching behaviour. It recorded how each participant was using the FlexphoReS prototype. After capturing the session, the recordings were analysed.

3.2.5.2 Questionnaire for suitability, flexibility and user subjective satisfaction

After completing all ten tasks, participants were asked to fill in a questionnaire about their interaction with the prototype. The questionnaire was designed partly by taking questions from Questionnaire for User Interface Satisfaction (QUIS) (Shneiderman and Plaisant 2005, pp.152-161). A copy of the questionnaire for this evaluation is available in Appendix 7. The QUIS instrument was modified to suit the current evaluation. It measured satisfaction, suitability and flexibility attributes on a 9-point scale. The questions covered included: overall reaction to the system, reaction to photo retrieval with mouse and keyboard input modalities and with mouse and speech input modalities, suitability and flexibility of the FlexPhoReS prototype as a digital photo retrieval system. There were also some open-ended questions, where the participants could list down other related aspects of the user interface. The data collected through the questionnaire and with some open-ended questions provided valuable information in addition to the statistical data which could be analysed both qualitatively and quantitatively.
3.2.6 Variables studied

The following variables were tested in the FlePhoReS evaluation on search performance, suitability, flexibility and subjective satisfaction of the FlexPhoReS prototype among the groups of digital photo users.

3.2.6.1 The independent variables

This study included two major independent variables:

a) FlexPhoReS prototype with mouse and speech input modalities to search photos

b) FlexPhoReS prototype with mouse and keyboard input modalities to search photos

3.2.6.2 The dependent variables

The two major groups of dependent variables studied were the search performance variables and users' subjective satisfaction. Other groups of dependent variables studied were suitability and flexibility of the prototype user interface.

Search performance

The search performance variable was calculated based on search task completion time: the total time taken to complete the search tasks. These times were extracted from the users' interaction with the system, and the computer screen recording software.
Subjective satisfaction and acceptability (suitability and flexibility)

Prototype evaluation of subjective satisfaction, suitability and flexibility with the prototype user interface was measured by questionnaires given to the participants upon completion of the experiment. Responses to the open-ended items in the questionnaire were analysed to find out both positive and negative aspects about the prototype user interface.

3.2.7 Data analysis techniques

The quantitative data collected through screen recording software and the questionnaire were analysed using the Statistical Package for Social Science (SPSS). Frequency counts were performed on data to obtain descriptive measures.

3.2.7.1 Paired-sample t-test

The paired-sample t-test compares the mean scores for the same group of people under two different conditions, whether one condition is significantly higher than the other condition. Assumptions underlying the use of paired-sample t-test involve the dependent variable having a normal distribution (Pallant 2005, p.205). It is therefore important to verify the data for anomalies before conducting a test. For data with reasonably 'normal' distribution, paired-sample t-tests were run to see the search performance difference between using mouse and keyboard input modalities, and mouse and speech input modalities. Otherwise a non-parametric statistical test, Wilcoxon Signed Rank tests was run (Pallant 2005, p.286). A null hypothesis indicated the two conditions (first condition: using mouse and keyboard input modalities, second condition: using mouse and speech input modalities) were not significantly different. If the test shows significance, the null hypothesis is rejected to conclude that there is a difference between both input modalities used.
3.2.7.2 Wilcoxon Signed Rank test

A Wilcoxon Signed Rank test also known as the 'Wilcoxon Matched Pairs Signed Ranks test' is a non parametric statistical test used when subjects are measured under two different conditions (Pallant 2005, p.292). This test was conducted to compare the search performance difference between using mouse and keyboard input modalities, and mouse and speech input modalities.

3.3 Assumptions in the research

There are a number of assumptions underlying this research. In the first phase:

- Data gathered from the literature review on user studies in image collections studies; image retrieval evolution; user interface support for digital photo retrieval; and multimodal interface are sufficient as an important basis for the modelling of FlexPhoReS.

- A small group of digital photo users participated in the initial study in order to provide additional input of user requirements. Data gathered from this group would be sufficient to assist the construction of the model.

In the third phase:

- It was assumed that an evaluation by one group of digital photo users would be sufficient to determine the effectiveness of the prototype.

- With a small group of personal digital photo users (20 participants) and limited functionalities of the FlexPhoReS prototype, it is difficult to generalise the results to the general population who have different characteristics. Nevertheless, the evaluation results, to some extent, could give some indication on how web based personal digital photo retrieval with
multimodal user interface would be accepted and additionally reflects the ways in which personal digital photo users would use multimodal user interface to retrieve their web based personal digital photo collections.

3.4 Summary

In summary, the aim of this research was to design and develop an appropriate multimodal user interface which offers a flexible user interface for a web based personal digital photo retrieval system. There were three stages in this research. The first stage involved developing a model for Flexible Photo Retrieval System (FlexPhoReS). The second stage concentrated on the development of the prototype based on the model proposed and the third stage evaluated the prototype model to find its acceptability, subjective satisfaction and the search performance differences among digital photo users. Inferential statistics method was used to analyse the evaluation results.
CHAPTER 4

MODELLING FLEXPHORES

4.0 Introduction

This chapter describes the model of the FlexPhoReS prototype. The FlexPhoReS system model aims to be flexible in terms of user interaction styles when retrieving photos from digital photo collections in a web based environment.

• Section 4.1 discusses the proposed model in relation to the literature reviewed and the small-scale user study outcomes.

• Section 4.2 discusses the task description which includes the use case diagram of the user tasks.

• Section 4.3 presents the process model of the FlexPhoReS system.

4.1 Proposed model design

The system model proposed is based on the literature review and data collected from a small-scale user study. The model aims to meet the goals outlined in the overall purpose of this research given in the previous chapter.
4.1.1 The proposed model

Looking at previous and current photo retrieval models, very few of the systems make use of multimodal interaction styles for digital photo retrieval processes and none in a World Wide Web based environment. The proposed model in this research is based on data collected from a small-scale user study and incorporates some of the findings reported in the literature (Flickner et al. 1995; Shneiderman et al. 1997; SALT Forum 2002).

4.1.1.1 Small-scale user study outcomes

To gain user input on the design of the proposed system, an exploratory study through a structured interview was undertaken to investigate how digital photo users retrieve and organise their growing collections of digital photos? The structured interview schedules are provided in Appendix 1. From the interview session, among the data gathered were the following points (Ismail and O'Brien 2004, p.1046):

- There was a ready acceptance of this digital photo technology by all participants even though there was sometimes frustration in making sure the photos were always findable.

- When asked about difficulties in finding photos in their collection all the participants agreed that it was easy as long as the collection was small. "Small" is a relative term for personal collections and varies with the individual. Typically collection size among these participants was more than 100 personal digital photographs.

- Another difficulty in finding the photo(s) needed is the typically unhelpful filenames automatically assigned to images by the digital camera software. Usually these names do not help as they rarely give any idea of the content
but are frequently just a numerical string (img001.jpg). Difficulty is therefore defined as a need to browse excessively which is a problem when collection get larger.

- The most popular retrieval approach was to retrieve by events and time, followed by place, subject and content. Participants usually place their digital photos collection in a computer directory folder which has been named based on events and time as this is easy to remember.

- It was found that the consequences of participants' directory structure in the storing and annotation process affects their retrieval approach, e.g. sometimes they store by events and sometimes by time and this can lead to a lack of consistency.

- Most of the participants found that browsing thumbnail photos mitigates, to some extent, their problems in finding photos. It gives the ability to scan a screen full of images quickly and in this way helps refine their search. Viewing thumbnails can be a reasonably sophisticated retrieval process as within the set of images displayed, a person can make lateral links and choices when reacting to the images, rather than to a mere series of filenames.

- Browsing folders in hierarchical arrangements is effective especially when supported by thumbnail images but reveals its limitations as the collection increases. Browsing or searching with text via a set of user-oriented categories would give more specific access points and therefore better retrieval.

- It was clear from the participants that they would like to have more information associated with their photos as they perceive this will improve their retrieval success. Inevitably this extra information is thought of as
essentially textual; better descriptions, extra notes, more useful file and folder names. As textual retrieval by 'keyword' is seen as a powerful tool, participants recognised that more text would improve their ability to make more complex links and as a result they could find specific photos more quickly.

- All participants said that they did wish to share their digital photos with their family, friends and relatives, and that sharing is one of the main reasons why they moved to digital photo collections. The most popular sharing mechanism amongst these participants was to publish to a web site which anyone can access, followed by sending through email, sending CDROMs and posting printed digital photos. Email was not rated top as there can be a problem with space limitations on some email servers.

- Most of the participants said that they send photos through email when they want to send specific photos to specific people. In order to overcome the problem of sending multiple photos, they attempt to publish their digital photos to a web site which offers more space and allows anyone to access their photos as long they know the web site address. However, all of the respondents said that this mechanism is not always appropriate for sharing personal and confidential photos. They agreed that publishing to a web site which can only be access by authorised users would be a better mechanism for sharing their personal digital photos. All of the respondents agreed that digital photos should be as secure as possible to withstand a variety of threats from unauthorised users or hackers.

- People recognise their limitations, so it was not surprising that participants were happy to think they might be able to search using attributes of the images not involving text. All seven had very little knowledge of non-text content based image retrieval but were sympathetic to the idea of finding photos based on visual similarity. In fact, they practise this by their
observation of photos while browsing but find it time-consuming. There was an enthusiastic response to the prospect of effective automatic content based retrieval.

- All participants agreed it would be a good idea to have speech interface embedded in the system as an additional or alternative interface. Most of the participants said that speech interface could help them retrieve their digital photos collection and control the applications menu. They did not, however, want to lose keyboard and mouse facilities and would welcome both interface modes. Two participants commented that although speech interface could help them retrieve their digital photos, there are some limitations, such as speech noise and hardware requirements (sound card and microphone). In addition, people use different kinds of speech structure and dialect in their speech. The application with speech interface should consider those limitations in order to use both interfaces significantly.

The outcomes of the interviews informed the FlexPhoReS prototype system development and task description. The number of interview questions that asked for participants' agreement or disagreement in relation to the proposed prototype features (security, annotation, visual similarity and speech interface) were all answered positively. This could have been predicted and might have produced less subjective results if the questionnaire had been designed with more open ended questions. But as an exploratory study, the results confirmed that participants wanted more helpful and user friendly user interfaces incorporating the proposed features that would save time and were easy to understand.
4.2 Task description

The descriptions of user tasks with the 'use case' were developed to express and envision the user tasks for photo retrieval. A 'use case' for retrieving digital photos with chosen interaction modes in the World Wide Web based environment is as follows:

1. The system prompts user for a valid user id and password.
2. The system provides user with the interaction modes option.
3. The user logs into the system with chosen interaction modes.
4. The system checks user authentication.
5. The system prompts the user with the retrieval strategies.
6. The user retrieves photos using the chosen retrieval strategy with chosen interaction modes.
7. The system searches and displays the search results based on user input.
8. The user browses the retrieval results.
9. The user refines the search if necessary.
10. The user logs out of the system.

Alternative courses:

4. If the user id or password are invalid.
   4.1 The system displays an error message and speech prompt.
   4.2 The system returns to step 1.
8. If inappropriate photos are found and displayed
   8.1 The user can do retrieval refinement.
   8.2 The system returns to step 5.

Figure 4.1 shows the 'use case' diagram for the proposed prototype system showing five use cases and one actor.
4.3 Process model

The tasks description and the 'use case' diagram for the proposed photo retrieval system serve as the basis for the process model of FlexPhoReS. The process model (Figure 4.2) is used to describe the structure of the prototype system and represents the process flow.
Different users with different profiles retrieve their photos through the user interface. This stage consists of a set of user tasks and interaction modes which define the communication between the user and the photo retrieval system. Within the multimodal interaction modes, users can interact with the system either using mouse and keyboard or using mouse and speech input modalities. They can also
switch between these input modalities to suit their style and interest. With the chosen interaction mode (step 1a or 1b) initially the user logs into the system with a valid user id and password (step 2).

The login information entered is compared with the information stored in the authorisation database which includes detailed information about users such as name, user identification (id), password and numbers of photos in the collection. To retrieve photos, users not only must understand how to use the interaction devices but also give careful consideration to understanding the retrieval strategies. Photo browsing and navigation, query by text or keywords and query by visual example are among the typical retrieval strategies employed in the retrieval process. With the chosen input modalities (step 3a or 3b), the photo retrieval tasks starts with the selection of appropriate retrieval strategies which represent the user query formulations (step 4a to 4f and step 5a to 5i). These query formulations are triggered by the user's information need.

Once the search begins, the users are expected to wait until the search process is completed (step 6a to 6f). Then the retrieval results are displayed which enable users to view (step 7). Users can stop retrieving or exit if they are satisfied with the retrieval results (step 10). However, in some cases, users need to reformulate the search statement and perform a new search (step 8 to 9). To support this retrieval process, all of the digital photos must be indexed based on their photo features.
4.4 Summary

In this chapter, the proposed model (in relation to the literature reviewed and small-scale user study outcomes) was presented. The outcomes were used to develop the user task as illustrated in the 'use case' diagram where the emphasis is on user system interaction. The last section dealt with the FlexPhoReS proposed system process model itself. The process model was described in relation to the tasks description. The tasks description and the process model will, in turn, be the base of components and architecture of FlexPhoReS which is covered in the next chapter along with other FlexPhoReS prototype issue.
CHAPTER 5

FlexPhoReS PROTOTYPE

5.0 Introduction

The FlexPhoReS prototype is based on the model developed in Chapter 4. This chapter discusses the development of the prototype.

- Section 5.1 talks about the enabling technology used in the prototype design and development.
- Section 5.2 deals with the prototype design and development.
- Section 5.3 discusses the architecture of FlexPhoReS prototype.
- Section 5.4 presents the components of FlexPhoReS prototype.

5.1 Enabling technology

The five types of technology that are of interest to this research are the World Wide Web (WWW), Speech technologies and SALT (Speech Application Language Tags), Matrix Laboratory (MATLAB), Active Server Pages (ASP) and JavaScript. These areas provide the technology to enable this research to accomplish its goal.
5.1.1 The World Wide Web

The World Wide Web has become one of the most essential tools for the dissemination of global information and for global communication and was introduced by Centre European pour La Recherche Nucleaire (CERN) in 1990. (Deitel et al. 2001, p.10). World Wide Web is a set of information accessible using computers and networking, each unit of information identified by a Universal Resource Identifier (URI) (Berners-Lee 1994, p.1). The development of web based photo retrieval systems requires knowledge of how the World Wide Web works. Basically, the user needs to know how to use a web browser, which in turn interacts with the World Wide Web (see Figure 5.1) (Sahibuddin 1999, pp.191-192).

The web browser's basic functions are handled through HTML which can interact with executable content or gateway programs. A HTML document can also be used as an interface and sends data to a gateway program in the web server. FlexPhoReS system can be implemented on the internet, an intranet or extranet. In terms of enabling technology, there is not much difference between these three
environments. In the internet setup, the web server is accessible from anywhere else in the Internet and security access control is minimal. In an intranet setup, the web server and its client sit behind a firewall. In an extranet setup the server and the majority of its clients sit behind a firewall. Access from the Internet is limited to certain people or groups by setting up the server configuration and the firewall configuration.

5.1.2 Speech technologies and SALT

The two key underlying technologies behind speech technologies are speech recognition (SR) and text to speech synthesis (TTS).

i) Speech recognition (SR)

Speech recognition (or speech-to-text) includes the technologies that enable computer systems to identify the sound of a human voice (Weinschenk and Barker 2000, p.98). It involves capturing and digitizing sound waves, converting them to basic language units or phonemes, constructing words from phonemes, and contextually analyzing the words to ensure correct spelling for those words that sound alike. Then the application processes the words and compares them with the application grammar which is a structured collection of words or phrases that the application recognises. Figure 5.2 illustrates speech recognition description and process flow. An algorithm is used to segment the word and determine which letter "produces" which sound. For example "h" in "hello" produces the "h" phoneme, the "e" produces the "eh" phoneme, the first "l" produces the "l" phoneme, the second "l" nothing, and "o" produces the "oe" phoneme. To ensure accurate recognition, the application is encouraged to create or access a "speaker profile" that includes a detailed map of the user's speech patterns used in the matching process during recognition (Microsoft 2002).
ii) Speech synthesis (TTS)

The term speech synthesis refers to the technologies that enable computer systems to output simulated human speech (Weinschenk and Barker 2000, p.103). Speech Synthesis (or text-to-speech) is the process of converting text into spoken language which generates the digital audio for playback. It involves breaking down the words into phonemes. Figure 5.3 illustrates speech synthesis description and process flow.

Figure 5.3: Speech synthesis or text-to-speech process flow (Microsoft 2002)
iii) Speech Application Language Tags (SALT)

Speech Application Language Tags (SALT) is a speech interface markup language released by the SALT Forum in July 2002 (SALT forum 2002). It consists of a small set of XML elements, with associated attributes and Document Object Model (DOM) object properties, events and methods, which apply a speech interface to web pages. To support speech input and speech output, SALT can be used with HTML, XHTML and other standards to write speech interfaces for both voice-only (e.g. telephony) and multimodal applications. Figure 5.4 shows the SALT architecture.

![SALT Architecture Diagram](image)

Figure 5.4: SALT Architecture (SALT forum 2002).

There are three top-level elements in SALT; listen, prompt and dtmf. The listen element is used to configure the speech recognizer, executes recognitions and handles speech input events. The prompt element is used to configure the speech synthesizer and plays out prompts. Its content may be simple text, speech output
markup, variable values, links to audio files, or any mix of these. The dtmf element is used to configure and control key press input and other events in the telephony applications.

### 5.1.3 MATLAB

MATLAB stands for Matrix Laboratory. It is a high-performance language for technical computing that integrates computation, visualisation, and programming in an easy-to-use environment where problems and solutions are expressed in familiar mathematical notation. MATLAB features a family of application-specific solutions called toolboxes that consist of comprehensive collections of MATLAB functions (M-files). It extends the MATLAB environment to solve particular classes of problems including image processing, control systems, neural networks, fuzzy logic and many others. For a web environment, The MATLAB® Web Server enables users to create MATLAB applications that use the capabilities of the World Wide Web to send data to MATLAB for computation and to display the results in a Web browser (MathWorks 2004).

![Figure 5.5: Matlab web server configuration (MathWorks 2004).](image)
Figure 5.5 shows the configuration, a Web browser runs on client workstation, while MATLAB, the MATLAB Web Server (matlabserver), and the Web server daemon (httpd) run on another machine (MathWorks 2004).

5.1.4 Active Server Pages (ASP)

Active Server Pages (ASP) is Microsoft's server-side scripting technology that is used to create dynamic web sites. Active Server Pages uses an ASP engine that is part of the Microsoft's Web Server. Therefore, the processing of ASP code takes place on the web server. Figure 5.6 shows how an ASP page is processed and displayed. When a user requests an ASP page, (step 1) the ASP engine processes the page (step 2) and then sends the result as HTML code to the user's Web browser (step 3). This allows the ASP pages to be viewed and interpreted by every web browser (step 4). Basically ASP provides several built-in objects as a method for communicating with a web browser, gathering data sent by an HTTP request and distinguishing between users.

![Diagram of ASP page process and displayed](Wall 2000)
The most common ASP objects used are Request, Respond and Server. Request object is used to access the information passed by a 'get' or 'post' request. This information usually consists of data provided by a client in an HTML form. The Respond object sends information such as HTML to the client. The Server object provides access to methods and properties on the server (Deitel et al. 2001, p.832).

5.1.5 JavaScript

JavaScript was created by Brenden Eich of Netscape Communications and was first made available in 1995 as part of Netscape Navigator 2.0 web browser (McDuffie 2003, p.10). JavaScript is an object based scripting language specifically designed to make web pages dynamic and interactive (Hoque 1997, p.9). JavaScript can be embedded in HTML pages and does not use a server to run its code (except for Server-side JavaScript). Most of today's web browsers are JavaScript enabled. There are three forms of JavaScript; Core JavaScript, Client-side JavaScript and Server-side JavaScript (McDuffie 2003, pp.10-11). Core JavaScript includes the operators, control structures, build-in functions and objects that make JavaScript a programming language. Client-side JavaScript extends the JavaScript Core to provide access to web browser and web document objects via Document Object Model (DOM). Server-side JavaScript is another extension of Core JavaScript that provides access to database. Nevertheless, Client-side JavaScript is the most popular form of JavaScript.

5.1.6 Enabling technology summary

The World Wide Web in combination with MATLAB, Speech technologies and SALT, ASP and JavaScript technologies provide a platform for a digital photo retrieval system across the World Wide Web network. They also provide the flexibility in arranging the information and the document. Adding speech
technology as an additional interaction style provides the flexibility for a user to choose the modes of interaction to retrieve their digital photos.

5.2 Designing the prototype

Based on the task description and in conjunction with the process model, the interface flow diagram that illustrates the system interaction was sketched. Figure 5.7 shows the interface flow diagram with an embedded multimodal user interface that provides difference modalities for user input tasks and lets the user select the interface modes to perform a particular interaction. This, in turn, is based on the initial design of a low fidelity prototype and was proposed to get fast feedback from users on the initial prototype design. The initial design was sketched on paper and then mocked-up in Microsoft PowerPoint.
User Login

Type user id and password followed by clicking the submit button

Say user id and password

Valid?

Valid

Main Menu

1. Retrieval Strategies
2. Logout

Retrieval Strategies screen:

User Input

1. Query by text or keyword
2. Query by image content

Select

Query by text or keyword
(e.g. photo event, time, place, person, extra notes)

Query by photo content
(visual similarity/photo example)

Browsing & Navigation

Select

Type an appropriate keyword followed by clicking the submit button

Tap and say an appropriate keyword

Type an appropriate photo file name followed by clicking the submit button

Tap and say an appropriate photo file name

Click an appropriate hyperlink

Tap and say an appropriate word (e.g. photo event, time, place, person, extra notes)

PROCESSING USER QUERY

Retrieval results:
Set of digital photos

Refinement?

Yes

No

Display the output

Query refinement

End session

Figure 5.7: FlexPhoReS interface flow diagram
A series of screen shots of the low fidelity prototype version was developed and used to clarify user requirements related to the screen design and learning how to use the system. Figure 5.8 shows a series of screen shots of the low fidelity prototype demonstrating the flows of how a user can use the proposed system. Initially the system prompts the user for a valid user id and password through Screen A. Within the screen, the system provides interaction mode options which allow the user to choose either mouse or keyboard or mouse and speech as input modalities. By using a valid user ID and password, the user logs into the system with the chosen interaction modes.

The system checks the user authentication based on the input. If the user id or password is invalid, the system displays an error message and speech prompts (Screen B) and advises the user to retry again. If the system received a valid input, the system will prompt the user with the retrieval strategies option for photo retrieval (Screen C). From screen C, the user can retrieve photos using chosen retrieval strategies with chosen interaction modes.

Finally the system searches and displays the search results based on user input. Screen D shows the retrieval results of photo browsing by photo events, places, time and subjects. Screen E shows the retrieval results of keyword searching and Screen F shows the retrieval results of photo by visual example searching. In some cases, users need to reformulate the search statement and perform a new search or retrieval refinement before they can end the retrieval session.

Screen design involves colour combination, fonts, arrangement of information and sequence of screens after some tasks. In the informal low fidelity prototype user evaluation session, three digital photo users were asked to carry out a realistic task by simulating pointing and clicking using a pencil, simulating typing by writing on paper and using their voice for speech input while the researcher clicked an appropriate screen for viewing the outcomes of the user task.
Start

User id and password

Check User ID and Password

Valid

Not Valid

Error message

Check User ID and Password

Valid

Not Valid

Keywords Searching

Visual Example Searching

Browsing

Yes

No

Refinement

End

Figure 5.8: Series of screens shot (in PowerPoint) of the low fidelity prototype
Having finished the session, they were asked some questions about the screen design and the learning process of how to use the system. Details of the series of PowerPoint screens of the low fidelity prototype can be found in Appendix 4.

**What was learned?**

In general, all three users found that the screen design for the system was clear and adequate. They also found that the system was easy to understand and use. The following summary (Table 5.1) shows the problems participants experienced and the actions taken, if any in the further development of the prototype.

**Table 5.1: Problems and recommendation**

<table>
<thead>
<tr>
<th>Issues</th>
<th>Actions taken</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Users generally did not immediately know how to use a multimodal interface, tap and talk input method. However, once they were used to using the tap and talk input method they had no difficulty.</td>
<td>No action taken</td>
</tr>
<tr>
<td>2 It was not clear to two of them how they could do the searching by visual example. There was no indication to guide them when they should choose a photo for visual example searching.</td>
<td>Put an indication on the visual retrieval menu to let user know that they should choose a photo for the searching.</td>
</tr>
<tr>
<td>3 For retrieving photos by example, most of the users expected to click the selected photo rather than to type in the photo reference name in the provided textbox to start searching.</td>
<td>Enable both input interactions which allow users to click on the photo reference name and type in the photo reference name into the provided textbox.</td>
</tr>
</tbody>
</table>

After conducting the informal low fidelity prototype user evaluation session, several modifications were made. The modifications were based on the issues and actions taken in Table 5.1 and were an input to the development of the FlexPhoReS high fidelity prototype.
5.3 FlexPhoReS architecture

The FlexPhoReS architecture utilises World Wide Web technology in order to achieve its objective. Figure 5.9 shows the architecture of the web based FlexPhoReS system. The system architecture consists of two sections, namely, client and server. All FlexPhoReS programs and data files including the user's photo repository, profiles, dialogues, grammars, prompts and retrieval engine are stored and located in the web server. No information is kept on the client side.

![Figure 5.9: Architecture of FlexPhoReS](image)

The client machines run the web browser and the server machine runs the web server (see Figure 5.10). Microsoft Internet Information Services (IIS) web server and MATLAB web server were used to deploy FlexPhoReS in the World Wide Web environment. Microsoft Internet Information Services (IIS) web server is a platform to enable information publishing on the Internet while the MATLAB web server is used to power the visual retrieval function in FlexPhoReS. It comprises a combination of M-files, Hypertext Markup Language (HTML) and Active Server Pages (ASP).
The system is initiated when the user enters a FlexPhoReS URL in a web browser; the Web server opens the FlexPhoReS application's default page. The Web server sends HTML, SALT, and JavaScript to the client machine. SALT markup in the pages that the Web server sends to the client can trigger the speech recognition and text-to-speech synthesis engine. For text-to-speech synthesis, the prompt element is used to specify the content of the audio output. Speech recognition, or speech-to-text, involves capturing and digitizing the sound waves. Then FlexPhoReS processes the words and compares them with the FlexPhoReS grammar (XML tag suite) which is a structured collection of words or phrases that the FlexPhoReS recognises and attempts to match human patterns of speech. In FlexPhoReS, the listen element is used for speech recognition. Listen element contains one or more grammar elements, which are used to specify possible user inputs. Figure 5.11 shows a snapshot of SALT elements used in the FlexPhoReS.
Different scenarios of speech recognition can require subtle differences in behaviour from a speech recognizer. For a multimodal application, a single mode of recognition was chosen as a type of recognition scenario (SALT 2002, p.39). Single mode recognition is typically used for 'tap and talk' scenarios. In this mode, the return of a recognition result is under the control of an explicit stop call from the application. Figure 5.12 illustrates the single mode listen behaviour of speech recognition events timeline that used in FlexPhoReS prototype system.

Figure 5.11: SALT pages elements

```html
<xhtml>
<input ...
</xhtml>

```salt:listen```
```salt:prompt```
```salt:grammar```
```script```
  RunSpeech()
```</script>```
```</xhtml>```

Figure 5.11: SALT pages elements

Figure 5.12 shows the Stop() call in action and the possible resulting listen events of onreco or onnoreco. Onreco is the event handler that is fired when the recognizer has a successful recognition result, while onnoreco is the event...
handler that fired when the recognizer was unable to return a complete recognition result.

The prompt element is used to specify the content of audio output. A prompt may be managed through a model of prompt queuing. Figure 5.13 shows the queuing of two prompts. Prompt 2 is added to the tail of the same sub queue, after prompt 1. Therefore prompt 2 will be played back immediately after prompt 1 is finished. Prompts also can be specified and played individually.

![Prompts queuing](image)

Figure 5.13: Prompts queuing (SALT 2002, p.22).

The FlexPhoReS prototype currently has 30 digital photos and has been annotated manually by 'what?' (event), 'where?' (place), 'who?' (subject/people) and 'when?' (time). These photos are part of the researcher’s family personal digital photo collection.

Typically, personal collections of photos include images of family, friends, activities such as outings, holidays, places and ‘things’ that were encountered as part of these activities (Taylor et al. 1982). Although as individuals our lives and experiences are different, our photo collections are very similar in that they represent memories of events, people, time and places that we wish to preserve and share. Birthdays, holidays, social events and activities such as wedding etc. represent the core of personal collections. In this respect, most personal collections are similar (Rodden 1999; Gargi 2002; Van House 2004).
Unlike institutional image collections, personal photo users do not engage in making detailed annotations to their personal photo collection (Rodden and Wood 2003; Van House 2004). Evidence shows that most personal photo users do minimal photo annotation noting only the event, date, location or people based on their own personal meaning (Rodden 1999; Ismail and O’Brien 2004; Van House 2004). Commercial photo management systems while usually allowing ‘keyword’ annotations, generally prompt users to store folders by date and to use rating system (Canon digital photo system) or tagging their personal photo based on categories such as event, places, people, time and others (Adobe 2006) (Figure 5.14). Some related systems adopted photo browsing that represent based on this similar model of organising personal photos (Shen et al. 2003, Chen et al 2003).

![Adobe Photoshop Album](http://www.dpreview.com/news/0301/03010601photoshopalbum.asp)

**Figure 5.14:** Adobe Photoshop Album (Adobe 2006)

For the FlexPhoReS prototype it was decided to choose a small sample of photos that typically illustrated this type of collection. The selection came from the researcher's personal collection. It consisted of 30 digital photos representing events and activities such as a trip to castles, a trip to York, playing in the snow, a trip to the seaside, a camping trip, etc. Additionally due to the digital photo copyright issue, where all of the digital photos materials are protected by copyright law (TASI 2006), personal digital photo collection in the FlexPhoReS were only based on researcher's personal collection. However given this limitation the users were able to simulate photo searching tasks based on FlexPhoReS photo browsing, keywords searching and visual example searching.

Figure 5.15 gives an overview of the FlexPhoReS user interface which has the following abilities:

- Browse photos by event (what?), by place (where?), by people/subject (who?) and by time (when?) from user photo repository web database.

- Control or navigate through the system. For example to logout from system, go to main page, retrieve system help, and go to next page and previous pages.

- Search photos from user photo repository web database by text or keywords.

- Search photo by visual example from user photo repository web database.
Chapter 5  

Microphones for mouse and speech input modalities

Control/Browse  
**What can I say?**
Please say your command

Search by keyword  
**What can I type/say?**
Please enter your keywords
Submit  Reset

Search by visual example  
**What can I type/say?**
Please choose your photo example
Submit  Reset

Browse photos by
- Events
- Places
- Subjects/People
- Times

Figure 5.15: FlexPhoReS user interface

With a multimodal web browser, the system user interface provides multimodal interaction which allows the use of either i) mouse and speech input modalities or ii) mouse and keyboard input modalities to control the system and perform the photo retrieval tasks. Users can select the input modalities for interaction that best suits their style and needs.

The FlexPhoReS prototype currently only enables a limited number of words for speech interaction. The following words and their meanings are the words that
can be use for speech interaction with the prototype in the research experiment (Table 5.2).

**Table 5.2: Words and their meanings for speech interaction**

<table>
<thead>
<tr>
<th>Retrieval strategies</th>
<th>Words for speech interaction</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Browsing</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>words used for speech interaction to invoke browsing category ()</td>
<td>camping (event)</td>
<td>Trip to camping</td>
</tr>
<tr>
<td></td>
<td>beaches (event)</td>
<td>Playing on the beaches</td>
</tr>
<tr>
<td></td>
<td>castle (event)</td>
<td>Trip to castles</td>
</tr>
<tr>
<td></td>
<td>swimming (event)</td>
<td>Swimming in the pool</td>
</tr>
<tr>
<td></td>
<td>strawberry (event)</td>
<td>Picking strawberries</td>
</tr>
<tr>
<td></td>
<td>snowing (event)</td>
<td>Playing in the snow</td>
</tr>
<tr>
<td></td>
<td>park (event)</td>
<td>Playing in the park</td>
</tr>
<tr>
<td></td>
<td>boating (event)</td>
<td>Boating on the river</td>
</tr>
<tr>
<td></td>
<td>browse (event)</td>
<td>Browse all events</td>
</tr>
<tr>
<td><strong>Keyword searching</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>april</td>
<td></td>
</tr>
<tr>
<td></td>
<td>camp</td>
<td></td>
</tr>
<tr>
<td></td>
<td>camping</td>
<td></td>
</tr>
<tr>
<td></td>
<td>castle</td>
<td></td>
</tr>
<tr>
<td></td>
<td>snow</td>
<td></td>
</tr>
<tr>
<td></td>
<td>snowing</td>
<td></td>
</tr>
<tr>
<td></td>
<td>york</td>
<td></td>
</tr>
<tr>
<td><strong>Visual example searching</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>boat one</td>
<td>unique and refer to photos file name</td>
</tr>
<tr>
<td></td>
<td>boat two</td>
<td></td>
</tr>
<tr>
<td></td>
<td>camping one</td>
<td></td>
</tr>
<tr>
<td></td>
<td>camping two</td>
<td></td>
</tr>
<tr>
<td></td>
<td>park one</td>
<td></td>
</tr>
<tr>
<td></td>
<td>park two</td>
<td></td>
</tr>
<tr>
<td></td>
<td>park three</td>
<td></td>
</tr>
<tr>
<td></td>
<td>park four</td>
<td></td>
</tr>
<tr>
<td></td>
<td>pool one</td>
<td></td>
</tr>
<tr>
<td></td>
<td>sand one</td>
<td></td>
</tr>
<tr>
<td></td>
<td>sand two</td>
<td></td>
</tr>
<tr>
<td></td>
<td>snow one</td>
<td></td>
</tr>
<tr>
<td></td>
<td>snow two</td>
<td></td>
</tr>
<tr>
<td></td>
<td>snow three</td>
<td></td>
</tr>
<tr>
<td></td>
<td>strawberry one</td>
<td></td>
</tr>
<tr>
<td></td>
<td>strawberry two</td>
<td></td>
</tr>
<tr>
<td></td>
<td>strawberry three</td>
<td></td>
</tr>
<tr>
<td></td>
<td>york one</td>
<td></td>
</tr>
<tr>
<td></td>
<td>york two</td>
<td></td>
</tr>
<tr>
<td></td>
<td>york three</td>
<td></td>
</tr>
<tr>
<td><strong>Control</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>about</td>
<td>To control the application which is always visible on the web application page.</td>
</tr>
<tr>
<td></td>
<td>go back</td>
<td></td>
</tr>
<tr>
<td></td>
<td>go forward</td>
<td></td>
</tr>
<tr>
<td></td>
<td>help</td>
<td></td>
</tr>
<tr>
<td></td>
<td>home</td>
<td></td>
</tr>
<tr>
<td></td>
<td>log out</td>
<td></td>
</tr>
<tr>
<td></td>
<td>profile</td>
<td></td>
</tr>
<tr>
<td></td>
<td>what can I say?</td>
<td></td>
</tr>
<tr>
<td></td>
<td>what can I type?</td>
<td></td>
</tr>
</tbody>
</table>

121
Browsing commands

In FlexPhoReS, photo browsing was based on four different categories of browsing (see Figure 5.23). Users could browse photos by clicking on the categories of Event, Place, Subject/People and Time. Each category of photos were already associated by hypertext with retrieval words that link to the related user's photo collection (see Figure 5.24 and Figure 5.15). Users simply choose to browse their photo categories by clicking on the retrieval words hyperlink and FlexPhoReS displays the set of photos (result set) based on the chosen hyperlink word.

When using mouse and speech input modalities, users have to click the specific (blue) microphone to invoke Browse mode and identify the appropriate browsing categories by using speech. To browse the photos, there are four possibilities: Event, Place, Subject/People and Time. Speech recognition gets more difficult when the application grammar and vocabularies are large or have many similar-sounding words (Shneiderman 2005, p.375). At the recognition stage, due to performance limitations, speech recognition is also unstable when recognizing too many combinations of vocabularies or long words input. The users therefore need a simple word to invoke the correct browse category in order to avoid any grammar collision with other photo browsing retrieval categories (Hunt 2004). For example, if the user desires to search for photos of York, he/she has to say the terms "York" and also the category term "Place". In the same way, if the user wishes to retrieve photos of snowing, the user has to say the word "snowing" and the category term "Event".

According to WordNet the definition of event refers to something that happens or activity at a given place and time (WordNet 2006). Some commentators noted that events might have to take on some of the character of objects (Bunnin, N and E. P. Tsui-James 2006). However, individuals define an event in subjective way (Wikipedia 2006). In FlexPhoReS usability testing, the term 'event' was used to
indicate a category which included various kinds of activities and objects. It may not have been the ideal term to employ because not all the retrieval words listed under 'Event' are objectively events. Some of them are objects and other entities. This did not cause any confusion in the evaluation as the users were able to simulate their photo browsing tasks during the evaluation including pilot test and final evaluation session as participants saw it simply as a tag to evoke the appropriate browse category.

For the prototype it was decided to base the browsing retrieval tasks (Table 3.1 and Table 3.2) on the categories of 'Event' (What?) and 'Place' (Where?). Further development would involve adding more annotation for speech retrieval especially in the categories that have fewer annotations at present (When?) or currently not used in the system (Who?) evaluation.

**Control commands, Keyword searching and Visual example searching**

FlexPhoReS limited number of words for speech interaction were intended to support the retrieval strategies (Section 5.4.4 and Figure 5.23) and application control. This control commands vocabulary was designed based on the strategy to permit the user to speak words that could be seen on the application menu buttons, hyperlinks and the list of examples of what the user can say. This strategy was based on Java Speech API Programmer's Guide (Sun 1998) that demonstrated all of the limited control commands used in the FlexPhoReS including "go back", "go forward", "help", "home", "log out", "profile", "what can I say?", "what can I type?" were visible on the application screen.

For retrieving photos by keyword, the words used in the retrieval process with speech input is similar to text input by keyboard. The specific microphone (red microphone) with the permitted words were used to support consistent handling of the speech input vocabularies for keyword searching. Those words were able to replicate digital photo retrieval by keyword searching that support the 4W's
Chapter 5

FlexPhoReS Prototype

("camping" - What? (event), "april" - When? (time), "person's name" - Who? (people), "York" - Where? (place) and photo description/extra note (Other) that related to the digital photo collection.

To retrieve photos by visual example, the specific microphone (black microphone) with the permitted words for speech interaction were based on the photo file name (unique and visible on top of each thumbnail size photo) to provide consistent handling of the speech input vocabularies.

5.4 FlexPhoReS components

This section describes the components of FlexPhoReS. The components are described according to their function. There are six key components in the FlexPhoReS prototype system:

1. User profiles
2. Repository and access control
3. Multimodal interaction
4. Retrieval strategies
5. Photo features
6. Help

5.4.1 User Profiles

The user profiles component was implemented in order to support the user's environment according to the information in the profile. Settings in the user profile are unique to each user. Changes made to one user's profile do not affect other users or other users' profiles. The FlexPhoReS system stores the user's profile information in the profiles directory that were located in the user photo repository and user profiles database. The user's profile in the FlexPhoReS database
includes information about users such as name, user identification (id), password and numbers in the collection. Figure 5.16 shows the FlexPhoReS user profiles.

Figure 5.16: FlexPhoReS user profiles

5.4.2 Repository and access control

Repository and access control are important components but not central to this research. A repository is a user's photo database, while access control manages the user security aspect of the system model. The access control issue is handled by authorisation of the client during the login into the system. Figure 5.17 shows the FlexPhoReS access control steps (step 1 to step 5b). Access into the system is managed by requiring the user to enter a valid user identification and password. The information entered is compared with the information stored in the authorisation database. Any attempt to access the FlexPhoReS web page will result in a login being displayed on the client's multimodal web browser (step 1, Figure 5.17). The client is required to enter a user id and password (Figure 5.18). The information is passed back to a server program on the web server (Step 2, Figure 5.17). The user id and password are then compared against the authorisation file (Step 3 and 4, Figure 5.17). The FlexPhoReS starting page (Figure 5.19) which contains a valid user's photo repository is displayed when the user id and the password are confirmed valid (Step 5a, Figure 5.17). Otherwise a warning message and speech prompt are generated and displayed to the client (Step 5b, Figure 5.17) which asks the user to re login (Figure 5.20). This in turn will determine the authorised user of the photo repository.
Introduction

Web-based Flexible photo retrieval system (FlexPhoReS) is a prototype system model which allows users more flexibility in performing photo retrieval in their personal photo collection.

It allows users to use speech and graphical user interface (S/GUI) to control and retrieve their personal photo collection.

The aim of FlexPhoReS is to provide a system that will improve the acceptability of a photo retrieval user interface for personal digital photo collections in a web environment.

Figure 5.17: FlexPhoReS access control architecture

Figure 5.18: FlexPhoReS login page

Please sign in to retrieve your digital photo

Enter your ID and password

User ID
What can I help you?
Password
Log in | Reset
Chapter 5

FlexPhoReS Prototype

Figure 5.19: FlexPhoReS starting page

Figure 5.20: FlexPhoReS warning message and speech prompt

127
5.4.3 Multimodal interaction

Multimodal interaction refers to the style of interaction which enables users to use either mouse and keyboard or mouse and speech input modalities for photo retrieval tasks. The multimodal interaction component is implemented in order to provide flexibility with different modalities for user input tasks and lets the user select the interaction modes to perform a particular task. For mouse and speech input, the user can give commands by using the ‘tap’ and ‘talk’ features. ‘Tap’ and ‘talk’ avoids speech detection problems that are critical to noise environment deployment for FlexPhoReS. Users can tap to activate the microphone and speak specific words. There are 3 different microphone buttons with different embedded functions for speech interaction. Figure 5.21 is a snapshot of the ‘tap’ and ‘talk’ features in the FlexPhoReS prototype while figure 5.22 shows that each photo name is visible on top of each displayed photo in order to let users select it for speech interaction.

![Diagram of multimodal interaction components]

**Figure 5.21: ‘Tap’ and ‘talk’ in the FlexPhoReS**
The first microphone button refers to photo browsing and application control functions. The second microphone button refers to searching by keywords function and the third microphone button refers to searching by visual similarity function. Users do not have to click the submit button if they use speech interaction. FlexPhoReS will automatically submit the recognised input data. If the data is not recognised, the system will prompt an error message through speech output and ask the user to re enter the input. This process will continue until the user speaks the recognised input data. 'What can I say?' hyperlink is a medium for the user to know what they can speak if they tap the selected microphone. The hyperlink will also pop up if the user taps the microphone and asks 'what can I say?'. Different microphones should have different outputs and guidelines to the user.

![Figure 5.22: Visible photo name on top of each photo](image)

Each photo file name is visible to let users use it in multimodal interaction.
5.4.4 Retrieval strategies

The retrieval strategies are an important component to this research. Figure 5.23 illustrates the retrieval strategies flow in the FlexPhoReS prototype system which allows users to utilise photo browsing and navigation, query by text or keywords and query by visual example.

Users can interact with their chosen retrieval strategy either by using mouse and keyboard or by using mouse and speech input modalities or switching between these input modalities to suit their style and interest. For example:

- Users can browse and navigate their photos by following the appropriate hyperlinks that are associated with the user's photo repository database.

- Alternatively, if users would like to browse and navigate their photos by using mouse and speech input modalities, they have to click the first (blue) microphone. Then they will hear an audio prompt which asks them to say...
their control or browse commands. Users are advised to wait until the audio prompt has finished. Then they can say any available control or browse command

- If the users know what to say, they do not have to wait until the speech prompt is finished.

- Currently there are limited numbers of available control or browse commands (see Section 5.3). Figure 5.24 shows the browsing and navigation result and categories that can be used in the system.

![Figure 5.24: Browsing and navigation result and categories](image-url)
• For keyword searching, users are able type in any available keywords into the given textbox; followed by pressing the submit button to start the searching.

• Alternatively, if the users prefer to search by keywords by using mouse and speech input modalities, they click on the second (red) microphone. Then they will hear an audio prompt which asks them to say the keyword. Users are advised to wait until the audio prompt has finished. Then they can say any available keywords.

• However, if the users know what to say, they do not have to wait until the speech prompt has finished.

• Currently there are a limited number of available keywords that the users can use in the system (see Section 5.3). Figure 5.25 shows a keyword used and the keywords searching results set page.

Figure 5.25: Searching by keywords
• For visual example searching, users can use their mouse to click on the photo that they wish to choose as an example. Then the photo name will appear in the visual example searching textbox. Users have to press the Submit button to start the searching. They may also type in any photo name into the given textbox that they wish to choose as their photo example for visual searching, followed by pressing the submit button to start the searching. Each photo name is situated on the top of a set of displayed photos (Figure 5.26).

![Photo names diagram]

Figure 5.26: Photo names

• Alternatively, if the users would like to do visual example searching by using mouse and speech input modalities, they have to click the third (black) microphone. Then they will hear an audio prompt which asks them to say their photo example name. Users are advised to wait until the audio prompt has finished. Then they can say any available photo name that they wish to use for searching by visual example. Users are encouraged to browse their photo collection before doing the visual example searching in order to find an appropriate photo example. However, if the users know what to say, they do not have to wait until the speech prompt is finished or browse their photo collection. Figure 5.27 shows visual example searching results set page with the selected photo example.
Figure 5.27: Visual example searching results with photo example.

5.4.5 Photo features

To support retrieval strategies, both the logical and the physical level of photo features are implemented in the FlexPhoReS prototype. The logical level is represented by a set of photo contexts such as: the photo events (what?), photo places (where?), photo times taken (when?), photo subject/person (who?) and extra notes (photo description). The Structure Query Language (SQL) technique was used to support the browsing and searching by text or keywords.

The physical level is represented by intrinsic physical features to support the visual similarity attributes. This research concentrated on the aspect of visual retrieval algorithm that is based on colour features. This technique is widely used in existing visual retrieval systems because colour features are usually very robust.
to noise, photo degradation, changes in size, resolution and orientation (Flickner et al. 1995; Bach et al. 1996).

The algorithm was calculated based on algorithms proposed by the well known IBM's QBIC CBIR systems (Niblack et al. 1993). The colour extractions were performed using a global colour histogram which represents one whole image by a single histogram. An example of a histogram can be seen in Figure 5.28.

Figure 5.28: Sample photo and its corresponding histogram

The equation used in deriving the distance between two colour histograms is the quadratic distance algorithm metric which measures the weighted similarity between histograms:

\[ d^2(x, y) = (x - y)^T A(x - y) \]  

The equation consists of two important terms. The first term consists of the difference between two colour histograms (x-y); or more specifically the difference in the number of pixels in each bin. The second term is the similarity matrix \( A \) where \( A = [a_{ij}] \) and \( a_{ij} \) denotes the similarity of colours with indexes \( i \) and \( j \). By defining colour similarity in Hue, Saturation, and Value (HSV) colour space, the similarity between any two colours \( m_q = (h_q, s_q, v_q) \) and \( m_t = (h_t, s_t, v_t) \) is given by equation (2) (Smith and Chang 1997, p.34).
\[ a_{q,t} = 1 - \sqrt{\left( v_q - v_t \right)^2 + \left( s_q \cos(h_q) - s_t \cos(h_t) \right)^2 + \left( s_q \sin(h_q) - s_t \sin(h_t) \right)^2} \]

(2)

The final result \( d \) (in equation 1) represents the colour distance between two images. The closer the distance is to zero the closer the images are in colour similarity. The further the distance from zero the less similar the images are in colour similarity.

Figure 5.29 illustrates the user retrieval query options from retrieval strategies and shows the photo feature processes in the FlexPhoReS web server photo retrieval engine. Initially the photo retrieval engine will process the retrieval query based on chosen query options (step 1a or 1b or 1c). For query by keywords searching (step 2a) and query by photo browsing and navigation (step 2b), the engine will process the retrieval query using SQL statements and hyperlinks from the user’s photo repository database (step 3a) and send the retrieval results set (step 3b and 7) for viewing. For query by photo example (step 2c), the engine will extract the photo example and generate its corresponding global colour histogram (step 3c). In order to generate the other photos’ global colour histograms, the same photo extraction process is implemented (step 2c(i), 3c(i)). This in turn will be based on the photo histogram comparison process (step 4c and 4c(i)) that generates the similarity values (step 5) for each photo. The engine will send the retrieval results set based on the similarity values (step 6 and 7) for viewing in the user's web browser. In some cases, users need to reformulate the search statement or perform retrieval refinement.
Figure 5.29: Retrieval query options and process

5.4.6 Help

The Help component is provided in order to support the prototype users in how to use the FlexPhoReS system. The implementation of the Help component is based on HTML documents and consists of two categories namely general Help and specific Help.

The general Help provides information about how to use and interact with the system in general. From the main page, the user can access the general Help either by clicking the help button or by tapping the blue microphone followed by
saying 'help' for mouse and speech input modalities. The specific Help provides specific information about what the user can say or type into the specific textboxes. ‘What can I say?’ and ‘What can I type?’ are the sentences and hyperlinks that can be used to access the specific helps. The user can access the specific help by following the specific ‘What can I say?’ or ‘What can I type/say?’ hyperlinks. For mouse and speech input modalities, the user can tap the specific microphone, followed by saying ‘What can I say?’ or ‘What can I type?’.

5.5 Limitations of the prototype

A prototype is a limited version of a full system that can simulate the scenarios of a real world application. It has, therefore a limited number of real system functionalities. FlexPhoReS prototype has a small number of digital photos; a limited number of words that can be use for search tasks; visual retrieval is based only on colour similarity; and it has no Boolean retrieval in keyword searching. However, given these limitations it can replicate most of the search strategies of a typical photo retrieval system.

5.6 Summary

This chapter began with a discussion of the implementation issues. Enabling technologies that are of interest to this research were presented. The initial design and informal evaluation of the low fidelity prototype which was based on FlexPhoReS modelling was discussed. Finally a discussion of the architecture and components of FlexPhoReS prototype itself was given, including an explanation of user profiles, repository and access control, multimodal interaction, retrieval strategies, photo features and help components in relation with the design of the prototype. Discussion on the evaluation of the prototype is given in the next chapter.
CHAPTER 6

USER EVALUATION OF THE PROTOTYPE

6.0 Introduction

This chapter presents the results of the user evaluation of the FlexPhoReS prototype. The participants who took part in this evaluation were digital photo users recruited randomly from various backgrounds at Loughborough, United Kingdom. The purpose was to examine their search performance, satisfaction with and the acceptability of the FlexPhoReS interface. A set of tasks was devised for the study. User interaction with the interface was recorded by screen and audio recording software which provided a clear picture if a user was successful or not in the search task, as well as time taken to complete the tasks.

- Section 6.1 discusses the users' background, gender and experience
- Section 6.2 presents the data analysis
- Section 6.3 presents the test hypotheses
- Section 6.4 presents the results of the user evaluation
- Section 6.5 presents the tests for statistical significance
6.1 User background, gender and experience

Twenty digital photo volunteers took part in the final evaluation. They filled in a recruitment questionnaire which assessed their computer and related system experience including experience with online digital photo retrieval features. The number of digital photos in their collection, their age, gender and educational background were recorded. Appendix 8 gives data about participants' demographic characteristics, and their experience in using computer.

The majority of the participants were male (17 participants) with only three female participants. They comprised ten research students, four postgraduate students, one undergraduate student, two machine operators, one lecturer, one automotive consultant and one teacher. Seventeen participants were below forty five years of age. Most of them had at least a degree at undergraduate level. They all reported to have at least 3 year's computer experience. The majority of them had at least one year's experience in using web applications (i.e. Google/Yahoo image search, eBay etc) that have digital photo or image retrieval features, with only one participant having less than 1 year's experience. The responses showed that twelve participants had used these system features very often (more than 5 times per month) (Figure 6.1).

![Frequency use of online image retrieval features](image)

Figure 6.1: Frequency use of online image retrieval features
In the questionnaires, the participants could report having experience with more than one category of retrieval strategies experience. The complete listing is in Figure 6.2.

![Figure 6.2: Participants' retrieval strategies experience](image)

The response shows that all of the participants had experience in searching by keyword and by browsing. Only two participants had experience in searching by visual example. All of the participants were experienced in using web based application with mouse and keyboard input modalities. None of the participants had previously used any web based application with mouse and speech input modalities.

### 6.2 Data analysis

The methods used in this experiment provided a large amount of data that had to be coded, structured and analysed. The data were analysed based on the participants' search performance and subjective measurements. Search performance measurements were quantitative involving the participants' search tasks completion time that could be seen and counted, while the subjective measurements were both qualitative and quantitative involving the participants'
perception, opinions and judgements of the system. The data gathered were analysed according to the following criteria:

1. Search performance measures
   - Overall search task performance completion time
   - Specific search tasks (browsing, keyword searching and visual example searching) completion time.

2. Subjective measures
   - Subjective satisfaction
   - Acceptability (suitability and flexibility) of the prototype

6.3 Test hypotheses

Based on independent and dependent variables (Section 3.2.6), statistical comparative tests were conducted to determine the differences between different input modalities (mouse and keyboard and mouse and speech input modalities) in search performance and subjective satisfaction among all participants. The following hypotheses were tested. The null hypotheses explored were:

H1 There is no difference between using mouse and keyboard input modalities and using mouse and speech input modalities in search performance.

H2 There is no difference between using mouse and keyboard input modalities and using mouse and speech input modalities in subjective satisfaction.

H3 There is no difference between using mouse and keyboard input modalities and using mouse and speech input modalities in search performance in retrieving photos by browsing.
H4 There is no difference between using mouse and keyboard input modalities and using mouse and speech input modalities in search performance in retrieving photos by keyword searching.

H5 There is no difference between using mouse and keyboard input modalities and using mouse and speech input modalities in search performance in retrieving photos by visual example searching.

6.4 Results of the user evaluation

6.4.1 Task completion time

In the evaluation experiment, all of the participants performed three different task categories namely: login into FlexPhoReS, understanding help components and search tasks. Details of the search tasks used in this evaluation can be found in Chapter 3.2.4.4. The time taken to complete each task was rounded up to the nearest minute. The task completion time included overall task completion time; instances of requested termination by participants and termination as a result of the fifteen minute time limit (Section 3.2.4.3). All of the participants completed the tasks without any major difficulty. Table 6.1 show the description of each task and task completion times of all participants. Table 6.2 shows the average and standard deviations of time taken to complete each task by all participants. Figure 6.3 shows the actual distribution.
Table 6.1: Participants task completion times

<table>
<thead>
<tr>
<th>Task Description</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
<th>19</th>
<th>20</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Login</td>
<td>0.28</td>
<td>0.36</td>
<td>0.33</td>
<td>0.18</td>
<td>0.25</td>
<td>0.22</td>
<td>0.27</td>
<td>0.40</td>
<td>0.40</td>
<td>0.47</td>
<td>0.38</td>
<td>0.37</td>
<td>0.30</td>
<td>0.25</td>
<td>0.22</td>
<td>0.27</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>General help</td>
<td>1.02</td>
<td>0.68</td>
<td>1.12</td>
<td>1.15</td>
<td>1.15</td>
<td>1.15</td>
<td>1.15</td>
<td>1.15</td>
<td>1.15</td>
<td>1.15</td>
<td>1.15</td>
<td>1.15</td>
<td>1.15</td>
<td>1.15</td>
<td>1.15</td>
<td>1.15</td>
<td>1.15</td>
<td>1.15</td>
<td>1.15</td>
<td></td>
</tr>
<tr>
<td>3i Browsing help</td>
<td>0.60</td>
<td>0.50</td>
<td>0.30</td>
<td>0.42</td>
<td>0.57</td>
<td>0.50</td>
<td>0.47</td>
<td>0.83</td>
<td>0.98</td>
<td>1.08</td>
<td>0.38</td>
<td>0.17</td>
<td>0.25</td>
<td>0.50</td>
<td>0.73</td>
<td>0.23</td>
<td>0.58</td>
<td>1.23</td>
<td>0.8</td>
<td></td>
</tr>
<tr>
<td>3ii Keyword searching</td>
<td>0.55</td>
<td>0.27</td>
<td>0.30</td>
<td>0.43</td>
<td>0.30</td>
<td>0.30</td>
<td>0.41</td>
<td>0.60</td>
<td>0.48</td>
<td>0.97</td>
<td>2.15</td>
<td>0.05</td>
<td>0.22</td>
<td>0.13</td>
<td>0.65</td>
<td>0.77</td>
<td>0.92</td>
<td>0.87</td>
<td>1.38</td>
<td>0.13</td>
</tr>
<tr>
<td>3iii Visual example</td>
<td>0.57</td>
<td>0.77</td>
<td>0.68</td>
<td>0.57</td>
<td>0.60</td>
<td>0.57</td>
<td>0.68</td>
<td>0.63</td>
<td>0.48</td>
<td>0.78</td>
<td>0.70</td>
<td>0.26</td>
<td>0.12</td>
<td>0.22</td>
<td>0.17</td>
<td>0.13</td>
<td>0.13</td>
<td>0.22</td>
<td>1.13</td>
<td></td>
</tr>
<tr>
<td>Browsing</td>
<td>0.52</td>
<td>1.05</td>
<td>0.27</td>
<td>0.93</td>
<td>0.65</td>
<td>0.78</td>
<td>1.08</td>
<td>1.37</td>
<td>1.20</td>
<td>0.97</td>
<td>0.75</td>
<td>0.37</td>
<td>0.50</td>
<td>1.55</td>
<td>0.80</td>
<td>1.12</td>
<td>0.85</td>
<td>0.60</td>
<td>0.45</td>
<td></td>
</tr>
<tr>
<td>5 Keyword searching</td>
<td>0.40</td>
<td>0.77</td>
<td>0.38</td>
<td>0.45</td>
<td>0.53</td>
<td>0.55</td>
<td>0.60</td>
<td>0.57</td>
<td>0.68</td>
<td>0.63</td>
<td>0.53</td>
<td>0.40</td>
<td>0.73</td>
<td>0.77</td>
<td>0.62</td>
<td>0.52</td>
<td>0.37</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6 Visual example</td>
<td>0.46</td>
<td>0.87</td>
<td>0.68</td>
<td>0.70</td>
<td>0.60</td>
<td>0.88</td>
<td>0.78</td>
<td>0.60</td>
<td>0.25</td>
<td>0.20</td>
<td>0.25</td>
<td>0.15</td>
<td>1.25</td>
<td>0.22</td>
<td>0.62</td>
<td>0.52</td>
<td>0.37</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Browsing</td>
<td>0.29</td>
<td>0.35</td>
<td>0.38</td>
<td>0.78</td>
<td>0.58</td>
<td>0.59</td>
<td>0.58</td>
<td>0.17</td>
<td>0.20</td>
<td>0.20</td>
<td>0.10</td>
<td>0.17</td>
<td>0.17</td>
<td>0.17</td>
<td>0.17</td>
<td>0.17</td>
<td>0.17</td>
<td>0.17</td>
<td>0.17</td>
<td></td>
</tr>
<tr>
<td>9 Keyword searching</td>
<td>0.30</td>
<td>0.40</td>
<td>0.30</td>
<td>0.60</td>
<td>0.30</td>
<td>0.40</td>
<td>0.30</td>
<td>0.40</td>
<td>0.40</td>
<td>0.30</td>
<td>0.30</td>
<td>0.30</td>
<td>0.30</td>
<td>0.30</td>
<td>0.30</td>
<td>0.30</td>
<td>0.30</td>
<td>0.30</td>
<td>0.30</td>
<td></td>
</tr>
<tr>
<td>10 Visual example</td>
<td>0.27</td>
<td>0.50</td>
<td>0.35</td>
<td>0.40</td>
<td>0.40</td>
<td>0.40</td>
<td>0.40</td>
<td>0.40</td>
<td>0.40</td>
<td>0.30</td>
<td>0.30</td>
<td>0.30</td>
<td>0.30</td>
<td>0.30</td>
<td>0.30</td>
<td>0.30</td>
<td>0.30</td>
<td>0.30</td>
<td>0.30</td>
<td></td>
</tr>
<tr>
<td>Sum</td>
<td>6.15</td>
<td>7.33</td>
<td>5.86</td>
<td>5.87</td>
<td>5.93</td>
<td>6.39</td>
<td>7.05</td>
<td>10.56</td>
<td>4.71</td>
<td>3.70</td>
<td>4.84</td>
<td>6.09</td>
<td>6.99</td>
<td>6.76</td>
<td>5.55</td>
<td>7.74</td>
<td>4.61</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average</td>
<td>0.51</td>
<td>0.61</td>
<td>0.38</td>
<td>0.49</td>
<td>0.55</td>
<td>0.49</td>
<td>0.41</td>
<td>0.49</td>
<td>0.53</td>
<td>0.66</td>
<td>0.88</td>
<td>0.39</td>
<td>0.31</td>
<td>0.40</td>
<td>0.51</td>
<td>0.58</td>
<td>0.56</td>
<td>0.46</td>
<td>0.65</td>
<td>0.38</td>
</tr>
</tbody>
</table>
Table 6.2: Means and standard deviations of task completion time for all participants (n=20)

<table>
<thead>
<tr>
<th>Task</th>
<th>Description</th>
<th>Mean</th>
<th>Standard deviations</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Login (mouse and keyboard)</td>
<td>0.31</td>
<td>0.10</td>
</tr>
<tr>
<td>2</td>
<td>General help</td>
<td>0.86</td>
<td>0.36</td>
</tr>
<tr>
<td>3i</td>
<td>Browsing help</td>
<td>0.56</td>
<td>0.29</td>
</tr>
<tr>
<td>3iii</td>
<td>Keyword searching help</td>
<td>0.41</td>
<td>0.24</td>
</tr>
<tr>
<td>3iii</td>
<td>Visual example searching help</td>
<td>0.58</td>
<td>0.52</td>
</tr>
<tr>
<td>4</td>
<td>Browsing</td>
<td>0.84</td>
<td>0.34</td>
</tr>
<tr>
<td>5</td>
<td>Keyword searching</td>
<td>0.55</td>
<td>0.14</td>
</tr>
<tr>
<td>6</td>
<td>Visual example searching</td>
<td>0.62</td>
<td>0.35</td>
</tr>
<tr>
<td>7</td>
<td>Login (mouse and speech)</td>
<td>0.17</td>
<td>0.06</td>
</tr>
<tr>
<td>8</td>
<td>Browsing</td>
<td>0.56</td>
<td>0.16</td>
</tr>
<tr>
<td>9</td>
<td>Keyword searching</td>
<td>0.37</td>
<td>0.14</td>
</tr>
<tr>
<td>10</td>
<td>Visual example searching</td>
<td>0.33</td>
<td>0.17</td>
</tr>
</tbody>
</table>

Figure 6.3: Average time taken to complete each task
It can be seen that on average participants required the longest time to complete Task 2 while they required the shortest time to complete Task 7.

Login into FlexPhoReS

All participants went through two different login tasks, namely Task 1 and Task 7. In completing Task 1, the participants were asked to use mouse and keyboard input modalities to login into FlexPhoReS. In Task 7, the participants were asked to use mouse and speech input modalities to login into the FlexPhoReS. It can be seen that participants required less time to complete Task 7 than Task 1.

Understanding help components

After logging into the FlexPhoReS system, all of the participants went through a tutorial session to understand system help components. There were four different main help pages in the FlexPhoReS help component including general help (Task 2), "what can I say?" that supports control, navigation and browsing (Task 3i), "what can I say/type?" that supports keyword searching (Task 3ii) and "what can I say/type?" that supports visual example searching (Task 3iii). All of the help pages were used to support participants to accomplish their search tasks. It can be seen that on average participants took a longer time to interact with the general system help components (Task 2). This was understandable as it was their first introduction to the system.

System Errors

The number of system errors that occurred was monitored. No system error occurred when participants used mouse and keyboard input modalities. When using mouse and speech input modalities, there were one hundred and forty speech input actions required in the evaluation experiment. Nine speech input errors relating to the speech recognition system occurred during the experiment.
The errors required participants to say selected words more than once before the system could understand the input commands. The system errors represent only 6.4% of all speech input actions.

**Search tasks**

All participants went through six different search tasks namely, Task 4, Task 5, Task 6, Task 8, Task 9 and Task 10. They were asked to use mouse and keyboard input modalities in completing Tasks 4 (retrieve photos by browsing), Task 5 (retrieve photos by keyword searching) and Task 6 (retrieve photos by visual example). Task 8 (retrieve photo by browsing), Task 9 (retrieve photos by keywords searching) and Task 10 (retrieve photos by visual example), were completed using mouse and speech input modalities.

All participants were able to complete all search tasks successfully and none exceeded the fifteen minute time limit for each task. Tables 6.3 show the search task completion times of all participants. Table 6.4 shows the means and standard deviations of the time taken to complete each search task by all participants while Figure 6.4 shows the actual distribution.

It can be seen that on average participants required the longest time to complete Task 4 (to retrieve photos by browsing using mouse and keyboard). A possible reason this might be due to the fact that this was their first retrieval task when they were least familiar with the prototype user interface.
Table 6.3: Participants search tasks completion

<table>
<thead>
<tr>
<th>I</th>
<th>Task</th>
<th>Description</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
<th>19</th>
<th>20</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>Browsing</td>
<td></td>
<td>0.52</td>
<td>1.05</td>
<td>0.27</td>
<td>1.10</td>
<td>0.83</td>
<td>0.65</td>
<td>0.76</td>
<td>1.08</td>
<td>1.37</td>
<td>1.20</td>
<td>0.97</td>
<td>0.75</td>
<td>0.37</td>
<td>0.50</td>
<td>1.55</td>
<td>0.80</td>
<td>1.12</td>
<td>0.85</td>
<td>0.60</td>
<td>0.45</td>
</tr>
<tr>
<td>5</td>
<td>Keyword searching</td>
<td></td>
<td>0.40</td>
<td>0.77</td>
<td>0.38</td>
<td>0.35</td>
<td>0.47</td>
<td>0.45</td>
<td>0.53</td>
<td>0.55</td>
<td>0.60</td>
<td>0.57</td>
<td>0.77</td>
<td>0.68</td>
<td>0.63</td>
<td>0.53</td>
<td>0.40</td>
<td>0.73</td>
<td>0.77</td>
<td>0.62</td>
<td>0.52</td>
<td>0.37</td>
</tr>
<tr>
<td>6</td>
<td>Visual example</td>
<td></td>
<td>0.68</td>
<td>0.87</td>
<td>0.68</td>
<td>0.75</td>
<td>0.70</td>
<td>0.20</td>
<td>0.73</td>
<td>0.60</td>
<td>0.27</td>
<td>0.88</td>
<td>0.78</td>
<td>0.70</td>
<td>0.25</td>
<td>1.06</td>
<td>0.25</td>
<td>0.15</td>
<td>1.25</td>
<td>0.13</td>
<td>0.22</td>
<td>1.13</td>
</tr>
<tr>
<td></td>
<td>searching</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Browsing</td>
<td></td>
<td>0.88</td>
<td>0.60</td>
<td>0.35</td>
<td>0.83</td>
<td>0.73</td>
<td>0.83</td>
<td>0.48</td>
<td>0.67</td>
<td>0.40</td>
<td>0.48</td>
<td>0.68</td>
<td>0.43</td>
<td>0.35</td>
<td>0.53</td>
<td>0.48</td>
<td>0.53</td>
<td>0.48</td>
<td>0.45</td>
<td>0.53</td>
<td>0.48</td>
</tr>
<tr>
<td>9</td>
<td>Keyword searching</td>
<td></td>
<td>0.30</td>
<td>0.40</td>
<td>0.20</td>
<td>0.30</td>
<td>0.80</td>
<td>0.30</td>
<td>0.28</td>
<td>0.33</td>
<td>0.37</td>
<td>0.32</td>
<td>0.58</td>
<td>0.22</td>
<td>0.30</td>
<td>0.33</td>
<td>0.37</td>
<td>0.32</td>
<td>0.48</td>
<td>0.50</td>
<td>0.42</td>
<td>0.28</td>
</tr>
<tr>
<td>10</td>
<td>Visual example</td>
<td></td>
<td>0.27</td>
<td>0.50</td>
<td>0.47</td>
<td>0.87</td>
<td>0.35</td>
<td>0.22</td>
<td>0.40</td>
<td>0.17</td>
<td>0.15</td>
<td>0.42</td>
<td>0.30</td>
<td>0.17</td>
<td>0.15</td>
<td>0.35</td>
<td>0.17</td>
<td>0.45</td>
<td>0.28</td>
<td>0.13</td>
<td>0.40</td>
<td>0.32</td>
</tr>
<tr>
<td></td>
<td>searching</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sum</td>
<td></td>
<td>3.05</td>
<td>4.19</td>
<td>2.35</td>
<td>4.20</td>
<td>3.88</td>
<td>2.65</td>
<td>3.2</td>
<td>3.4</td>
<td>3.16</td>
<td>3.87</td>
<td>4.08</td>
<td>2.95</td>
<td>2.05</td>
<td>3.32</td>
<td>3.22</td>
<td>2.98</td>
<td>4.38</td>
<td>2.68</td>
<td>2.69</td>
<td>3.03</td>
</tr>
<tr>
<td></td>
<td>Average</td>
<td></td>
<td>0.51</td>
<td>0.70</td>
<td>0.39</td>
<td>0.70</td>
<td>0.65</td>
<td>0.44</td>
<td>0.53</td>
<td>0.57</td>
<td>0.53</td>
<td>0.65</td>
<td>0.68</td>
<td>0.49</td>
<td>0.34</td>
<td>0.55</td>
<td>0.54</td>
<td>0.50</td>
<td>0.73</td>
<td>0.45</td>
<td>0.45</td>
<td>0.51</td>
</tr>
</tbody>
</table>

M&K* = Mouse and keyboard input modalities
M&S* = Mouse and speech input modalities
Table 6.4: Means and standard deviations of time taken to complete each search task for all participants (n=20)

<table>
<thead>
<tr>
<th>I</th>
<th>Task (T)</th>
<th>Description</th>
<th>Mean</th>
<th>Standard deviations</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td></td>
<td>Browsing</td>
<td>0.84</td>
<td>0.34</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>Keyword searching</td>
<td>0.55</td>
<td>0.14</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>Visual example searching</td>
<td>0.62</td>
<td>0.35</td>
</tr>
<tr>
<td>8</td>
<td></td>
<td>Browsing</td>
<td>0.56</td>
<td>0.16</td>
</tr>
<tr>
<td>9</td>
<td></td>
<td>Keyword searching</td>
<td>0.37</td>
<td>0.14</td>
</tr>
<tr>
<td>10</td>
<td></td>
<td>Visual example searching</td>
<td>0.33</td>
<td>0.17</td>
</tr>
</tbody>
</table>

M&K* = Mouse and keyboard input modalities
M&S* = Mouse and speech input modalities

Figure 6.4: Average time taken to complete each task with input modalities
It can be seen that, on average, participants required a longer time to complete search tasks when using mouse and keyboard input modalities (Task 4, Task 5, Task 9). Further statistical analyses were conducted to assess the significance of the differences across different input modalities used (Section 6.5).

6.4.2 Subjective satisfaction with the prototype

Two sets of subjective satisfaction questions were used in the experiment (Appendix 7). Each set included eleven questions, administered to all the participants at the end of the evaluation. Both set of questions were similar to each other but with different titles and purpose. The first set of questions was answered when the participant used mouse and keyboard input modalities and the second set of questions was answered when participants used mouse and speech input modalities. Reliability of the questions was assessed using Cronbach’s Alpha, yielding a value of $\alpha = 0.894$, which indicates that the questions on subjective satisfaction when using mouse and keyboard input modalities was highly reliable. In the case of subjective satisfaction when using mouse and speech input modalities, the Cronbach’s Alpha, yielded a value of $\alpha = 0.948$ indicating that the related questions were also highly reliable. A reliability or consistency test is an essential characteristic of a good evaluation. Cronbach’s Alpha coefficient is the most commonly used statistic to provide an indication of the average correlation among all of the items (questions and participants’ answers) that make up the scale. The values range from 0 to 1 with higher values indicating greater reliability. As a rule of thumb, the Cronbach’s Alpha coefficient of a scale should be above 0.7 (Pallant 2005).

The first six questions (Q1 to Q6) asked the participants to rate their overall satisfaction with FlexPhoReS with mouse and keyboard input modalities and mouse and speech input modalities. Two questions (Q7 and Q8) asked the participants how easy or difficult it was to login and understand help components? The final three questions (Q9, Q10 and Q11) asked the participants to give an
indication of how easy or difficult it was to complete search tasks with different strategies (photo browsing, keyword searching and visual example searching).

All of the participants rated their satisfaction with the FlexPhoReS system on a 9 point Likert-type scale. Item 1 was worded in a negative direction (low optimism) while item 9 was worded in a positive direction (high optimism). Tables 6.5 and 6.6 show the subjective satisfaction scores for participants (1-10) and participant (11-20) when using different input modalities.
Table 6.5: Participants (1-10) Subjective satisfaction scores

<table>
<thead>
<tr>
<th>I</th>
<th>Question</th>
<th>PARTICIPANTS (1-10)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Overall reaction</td>
<td>1 2 3 4 5 6 7 8 9 10</td>
</tr>
<tr>
<td>Q1</td>
<td>Terrible (1-9) Wonderful</td>
<td>6 6 9 3 8 7 8 7 7 7</td>
</tr>
<tr>
<td>Q2</td>
<td>Frustrating (1-9) Satisfying</td>
<td>6 6 9 5 8 8 8 7 7 7</td>
</tr>
<tr>
<td>Q3</td>
<td>Dull (1-9) Stimulating</td>
<td>6 7 9 4 8 7 8 6 7 6</td>
</tr>
<tr>
<td>Q4</td>
<td>Difficult (1-9) Easy</td>
<td>7 7 8 7 8 8 8 6 6 4</td>
</tr>
<tr>
<td>Q5</td>
<td>Inadequate power (1-9) adequate power</td>
<td>7 6 7 4 8 8 9 8 6 7</td>
</tr>
<tr>
<td>Q6</td>
<td>Rigid (1-9) Flexible</td>
<td>6 6 4 1 8 8 8 7 6 7</td>
</tr>
<tr>
<td></td>
<td>Login and understanding help</td>
<td></td>
</tr>
<tr>
<td>Q7</td>
<td>Login into the system: Difficult (1-9) Easy</td>
<td>7 7 6 8 8 9 9 9 7 5</td>
</tr>
<tr>
<td>Q8</td>
<td>Help components: Difficult (1-9) Easy</td>
<td>6 6 6 6 6 8 8 9 8 7 6</td>
</tr>
<tr>
<td></td>
<td>Search tasks with different strategies</td>
<td></td>
</tr>
<tr>
<td>Q9</td>
<td>Photo browsing: Difficult (1-9) Easy</td>
<td>6 6 9 7 8 9 8 8 6 6</td>
</tr>
<tr>
<td>Q10</td>
<td>Keyword searching: Difficult (1-9) Easy</td>
<td>8 6 9 7 9 8 8 7 7 6</td>
</tr>
<tr>
<td>Q11</td>
<td>Visual example searching: Difficult (1-9) Easy</td>
<td>9 8 9 3 9 7 8 7 6 6</td>
</tr>
<tr>
<td></td>
<td>Overall reaction</td>
<td>1 2 3 4 5 6 7 8 9 10</td>
</tr>
<tr>
<td>Q1</td>
<td>Terrible (1-9) Wonderful</td>
<td>7 9 9 3 8 7 9 8 7 9</td>
</tr>
<tr>
<td>Q2</td>
<td>Frustrating (1-9) Satisfying</td>
<td>7 9 9 5 8 7 9 9 9 8 9</td>
</tr>
<tr>
<td>Q3</td>
<td>Dull (1-9) Stimulating</td>
<td>7 9 9 5 8 7 9 8 8 9</td>
</tr>
<tr>
<td>Q4</td>
<td>Difficult (1-9) Easy</td>
<td>7 9 9 7 8 7 8 9 7 9</td>
</tr>
<tr>
<td>Q5</td>
<td>Inadequate power (1-9) adequate power</td>
<td>7 9 9 4 8 7 8 8 6 8</td>
</tr>
<tr>
<td>Q6</td>
<td>Rigid (1-9) Flexible</td>
<td>7 9 6 1 8 6 8 6 6 8</td>
</tr>
<tr>
<td></td>
<td>Login and understanding help</td>
<td></td>
</tr>
<tr>
<td>Q7</td>
<td>Login into the system: Difficult (1-9) Easy</td>
<td>8 9 5 8 9 9 9 9 9 9</td>
</tr>
<tr>
<td>Q8</td>
<td>Help components: Difficult (1-9) Easy</td>
<td>7 9 6 6 9 8 9 8 5 9</td>
</tr>
<tr>
<td></td>
<td>Search tasks with different strategies</td>
<td></td>
</tr>
<tr>
<td>Q9</td>
<td>Photo browsing: Difficult (1-9) Easy</td>
<td>5 9 8 7 8 7 9 9 6 9</td>
</tr>
<tr>
<td>Q10</td>
<td>Keyword searching: Difficult (1-9) Easy</td>
<td>8 9 8 7 8 7 9 9 7 9</td>
</tr>
<tr>
<td>Q11</td>
<td>Visual example searching: Difficult (1-9) Easy</td>
<td>9 9 8 3 8 7 9 9 6 9</td>
</tr>
<tr>
<td></td>
<td>Sum</td>
<td>153 168 171 111 180 166 186 172 147 184</td>
</tr>
<tr>
<td></td>
<td>Average</td>
<td>6.95 7.64 7.77 5.05 8.18 7.55 8.45 7.82 6.68 7.45</td>
</tr>
</tbody>
</table>
Table 6.6: Participants (11-20) Subjective satisfaction scores

<table>
<thead>
<tr>
<th>I</th>
<th>Question</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
<th>19</th>
<th>20</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>Overall reaction</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q1</td>
<td>Terrible (1-9) Wonderful</td>
<td>9</td>
<td>6</td>
<td>6</td>
<td>7</td>
<td>6</td>
<td>6</td>
<td>9</td>
<td>6</td>
<td>8</td>
<td>7</td>
</tr>
<tr>
<td>Q2</td>
<td>Frustrating (1-9) Satisfying</td>
<td>8</td>
<td>6</td>
<td>7</td>
<td>7</td>
<td>5</td>
<td>7</td>
<td>9</td>
<td>6</td>
<td>8</td>
<td>7</td>
</tr>
<tr>
<td>Q3</td>
<td>Dull (1-9) Stimulating</td>
<td>9</td>
<td>7</td>
<td>7</td>
<td>8</td>
<td>5</td>
<td>7</td>
<td>9</td>
<td>6</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>Q4</td>
<td>Difficult (1-9) Easy</td>
<td>8</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>5</td>
<td>8</td>
<td>9</td>
<td>8</td>
<td>8</td>
<td>7</td>
</tr>
<tr>
<td>Q5</td>
<td>Inadequate power (1-9) adequate power</td>
<td>8</td>
<td>5</td>
<td>7</td>
<td>7</td>
<td>6</td>
<td>7</td>
<td>9</td>
<td>7</td>
<td>7</td>
<td>6</td>
</tr>
<tr>
<td>Q6</td>
<td>Rigid (1-9) Flexible</td>
<td>9</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>3</td>
<td>7</td>
<td>9</td>
<td>7</td>
<td>7</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td><strong>Login and understanding help</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q7</td>
<td>Login into the system: Difficult (1-9) Easy</td>
<td>7</td>
<td>6</td>
<td>8</td>
<td>8</td>
<td>7</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>Q8</td>
<td>Help components: Difficult (1-9) Easy</td>
<td>9</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>5</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td><strong>Search tasks with different strategies</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q9</td>
<td>Photo browsing: Difficult (1-9) Easy</td>
<td>9</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>8</td>
<td>7</td>
<td>7</td>
<td>9</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>Q10</td>
<td>Keyword searching: Difficult (1-9) Easy</td>
<td>9</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>8</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>Q11</td>
<td>Visual example searching: Difficult (1-9) Easy</td>
<td>9</td>
<td>7</td>
<td>7</td>
<td>6</td>
<td>6</td>
<td>7</td>
<td>5</td>
<td>9</td>
<td>8</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td><strong>Overall reaction</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q1</td>
<td>Terrible (1-9) Wonderful</td>
<td>9</td>
<td>7</td>
<td>8</td>
<td>8</td>
<td>6</td>
<td>6</td>
<td>9</td>
<td>6</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>Q2</td>
<td>Frustrating (1-9) Satisfying</td>
<td>9</td>
<td>7</td>
<td>8</td>
<td>8</td>
<td>4</td>
<td>6</td>
<td>9</td>
<td>6</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>Q3</td>
<td>Dull (1-9) Stimulating</td>
<td>9</td>
<td>7</td>
<td>8</td>
<td>8</td>
<td>7</td>
<td>5</td>
<td>9</td>
<td>7</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>Q4</td>
<td>Difficult (1-9) Easy</td>
<td>9</td>
<td>8</td>
<td>8</td>
<td>9</td>
<td>7</td>
<td>4</td>
<td>9</td>
<td>8</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>Q5</td>
<td>Inadequate power (1-9) adequate power</td>
<td>9</td>
<td>7</td>
<td>8</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>9</td>
<td>7</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>Q6</td>
<td>Rigid (1-9) Flexible</td>
<td>9</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>4</td>
<td>2</td>
<td>9</td>
<td>7</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td><strong>Login and understanding help</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q7</td>
<td>Login into the system: Difficult (1-9) Easy</td>
<td>8</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>4</td>
<td>5</td>
<td>9</td>
<td>9</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>Q8</td>
<td>Help components: Difficult (1-9) Easy</td>
<td>9</td>
<td>8</td>
<td>8</td>
<td>7</td>
<td>6</td>
<td>5</td>
<td>9</td>
<td>8</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td><strong>Search tasks with different strategies</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q9</td>
<td>Photo browsing: Difficult (1-9) Easy</td>
<td>9</td>
<td>8</td>
<td>8</td>
<td>7</td>
<td>7</td>
<td>5</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>Q10</td>
<td>Keyword searching: Difficult (1-9) Easy</td>
<td>9</td>
<td>8</td>
<td>8</td>
<td>7</td>
<td>7</td>
<td>5</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>Q11</td>
<td>Visual example searching: Difficult (1-9) Easy</td>
<td>9</td>
<td>8</td>
<td>8</td>
<td>7</td>
<td>7</td>
<td>5</td>
<td>8</td>
<td>9</td>
<td>9</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td><strong>Sum</strong></td>
<td>192</td>
<td>151</td>
<td>164</td>
<td>165</td>
<td>132</td>
<td>132</td>
<td>190</td>
<td>170</td>
<td>175</td>
<td>167</td>
</tr>
<tr>
<td></td>
<td><strong>Average</strong></td>
<td>8.73</td>
<td>6.66</td>
<td>7.45</td>
<td>7.50</td>
<td>6.00</td>
<td>6.00</td>
<td>8.64</td>
<td>7.73</td>
<td>7.95</td>
<td>7.59</td>
</tr>
</tbody>
</table>

Mouse and keyboard input modalities

Mouse and speech input modalities

Sum: 192, Average: 8.73
Tables 6.5 and 6.6 provide the results for subjective satisfaction with FlexPhoReS. It can be seen that most of the questions achieved positive scores (above 4.5) with participant 4 having the lowest average subjective satisfaction score. With no experience and limited knowledge of visual example searching, participant 4 did not perform well to complete visual example searching using mouse and speech input modalities (Task 10). It can be seen that this participant required the longest time (0.87 minutes) to complete the task (Table 6.3) and therefore is reflected with a less positive satisfaction score.

To further investigate the difference in subjective satisfaction when using different input modalities, Table 6.7 provides a summary of the subjective satisfaction mean scores when using different input modalities. Figure 6.5 shows the distribution of the mean for subjective satisfaction when using mouse and keyboard input modalities as compared with mouse and speech input modalities.

Table 6.7: Means of subjective satisfaction when using different input modalities

<table>
<thead>
<tr>
<th>Participants (n=20)</th>
<th>Q1</th>
<th>Q2</th>
<th>Q3</th>
<th>Q4</th>
<th>Q5</th>
<th>Q6</th>
<th>Q7</th>
<th>Q8</th>
<th>Q9</th>
<th>Q10</th>
<th>Q11</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mouse and keyboard input modalities</td>
<td>6.90</td>
<td>7.05</td>
<td>6.90</td>
<td>7.15</td>
<td>6.95</td>
<td>6.50</td>
<td>7.80</td>
<td>7.30</td>
<td>7.55</td>
<td>7.65</td>
<td>7.00</td>
</tr>
<tr>
<td>Mouse and speech input modalities</td>
<td>7.55</td>
<td>7.60</td>
<td>7.75</td>
<td>7.90</td>
<td>7.55</td>
<td>6.45</td>
<td>7.95</td>
<td>7.65</td>
<td>7.85</td>
<td>8.05</td>
<td>7.75</td>
</tr>
</tbody>
</table>
Figure 6.5 reveals that on average, participants were more satisfied when using mouse and speech input modalities. Question 6 (rigidity/flexibility) was the exception to these findings with a lower score for both modalities.

Further statistical analysis was conducted to determine the significant difference in subjective satisfaction when using different input modalities (Section 6.5).

6.4.3 Acceptability of the prototype

The Set of questions concerning acceptability was divided into two categories namely (i) suitability and (ii) flexibility. The first four questions relating to suitability (Q1 to Q4) asked the participants to give an indication of the suitability of FlexPhoReS with different input modalities. Two questions related to suitability (Q5 and Q6) asked the participants to give an indication of the flexibility of FlexPhoReS. Reliability of the questions was assessed using Cronbach's Alpha, yielding a value of $\alpha = 0.882$, which indicates the questions were highly reliable.
All of the participants rated acceptability on a 9 point Likert-type scale for optimism. Item 1 was worded in a negative direction (low optimism) while item 9 was worded in a positive direction (high optimism). The score for each item ranged from 1 (strongly disagree) to 9 (strongly agree). Tables 6.8 and 6.9 show the results for acceptability for all of the participants. It can be seen that all of the related questions achieved positive direction score (i.e. above 4.5).

Table 6.8: Acceptability of the prototype for participants (1-10)

<table>
<thead>
<tr>
<th>Questionnaires (Q)</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1 Suitability of FlexPhoReS system with mouse and keyboard input modalities for photo retrieval tasks</td>
<td>7</td>
<td>6</td>
<td>9</td>
<td>7</td>
<td>9</td>
<td>8</td>
<td>8</td>
<td>7</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>Q2 Suitability of FlexPhoReS system with mouse and speech input modalities for photo retrieval tasks</td>
<td>7</td>
<td>9</td>
<td>9</td>
<td>7</td>
<td>9</td>
<td>7</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>7</td>
</tr>
<tr>
<td>Q3 Suitability of FlexPhoReS system with both mouse and keyboard and mouse and speech input modalities for photo retrieval tasks</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>7</td>
<td>9</td>
<td>8</td>
<td>9</td>
<td>9</td>
<td>6</td>
<td>9</td>
</tr>
<tr>
<td>Q4 Both input modalities are complementary each other to retrieve photos</td>
<td>7</td>
<td>8</td>
<td>8</td>
<td>4</td>
<td>9</td>
<td>7</td>
<td>9</td>
<td>9</td>
<td>5</td>
<td>9</td>
</tr>
<tr>
<td>Q5 FlexPhoReS offers better flexibility than a system that only offers one style of input modalities either the mouse and speech support or the mouse and keyboard support.</td>
<td>7</td>
<td>9</td>
<td>9</td>
<td>7</td>
<td>9</td>
<td>7</td>
<td>8</td>
<td>8</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>Q6 FlexPhoReS offers better flexibility in the World Wide Web environment then on the stand alone environment.</td>
<td>7</td>
<td>9</td>
<td>9</td>
<td>7</td>
<td>9</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>Sum</td>
<td>42</td>
<td>49</td>
<td>53</td>
<td>39</td>
<td>54</td>
<td>44</td>
<td>50</td>
<td>50</td>
<td>40</td>
<td>47</td>
</tr>
<tr>
<td>Average</td>
<td>7.00</td>
<td>8.17</td>
<td>8.83</td>
<td>6.50</td>
<td>9.00</td>
<td>7.33</td>
<td>8.33</td>
<td>8.33</td>
<td>6.67</td>
<td>7.83</td>
</tr>
</tbody>
</table>
Table 6.9: Acceptability of the prototype for participants (11-20)

<table>
<thead>
<tr>
<th>Questionnaires (Q)</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
<th>19</th>
<th>20</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1 Suitability of FlexPhoReS system with mouse and keyboard input modalities for</td>
<td>9</td>
<td>6</td>
<td>8</td>
<td>6</td>
<td>7</td>
<td>6</td>
<td>9</td>
<td>8</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>photo retrieval tasks</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q2 Suitability of FlexPhoReS system with mouse and speech input modalities for</td>
<td>9</td>
<td>6</td>
<td>8</td>
<td>7</td>
<td>7</td>
<td>6</td>
<td>8</td>
<td>8</td>
<td>9</td>
<td>7</td>
</tr>
<tr>
<td>photo retrieval tasks</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q3 Suitability of FlexPhoReS system with both mouse and keyboard and speech input</td>
<td>9</td>
<td>6</td>
<td>9</td>
<td>6</td>
<td>7</td>
<td>6</td>
<td>8</td>
<td>8</td>
<td>9</td>
<td>7</td>
</tr>
<tr>
<td>modalities for photo retrieval tasks</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q4 Both input modalities are complementary each other to retrieve photos</td>
<td>9</td>
<td>6</td>
<td>9</td>
<td>5</td>
<td>7</td>
<td>7</td>
<td>8</td>
<td>8</td>
<td>9</td>
<td>6</td>
</tr>
<tr>
<td>Q5 FlexPhoReS offers better flexibility than a system that only offers one style</td>
<td>9</td>
<td>7</td>
<td>8</td>
<td>7</td>
<td>5</td>
<td>7</td>
<td>9</td>
<td>7</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>of input modalities either the mouse and speech support or the mouse and keyboard</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>support.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q6 FlexPhoReS offers better flexibility in the World Wide Web environment than on</td>
<td>9</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>5</td>
<td>7</td>
<td>9</td>
<td>7</td>
<td>9</td>
<td>8</td>
</tr>
<tr>
<td>the stand alone environment.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sum</td>
<td>54</td>
<td>39</td>
<td>50</td>
<td>39</td>
<td>38</td>
<td>39</td>
<td>51</td>
<td>46</td>
<td>52</td>
<td>45</td>
</tr>
<tr>
<td>Average</td>
<td>9.00</td>
<td>6.50</td>
<td>8.33</td>
<td>6.50</td>
<td>8.63</td>
<td>6.50</td>
<td>8.50</td>
<td>7.67</td>
<td>8.67</td>
<td>7.50</td>
</tr>
</tbody>
</table>

6.5 Tests for statistical significance

Statistical significance tests were performed to see if the null hypotheses could be rejected or accepted. Level of measurement was observed and normality tests were conducted to check the normality of data before conducting the appropriate tests (parametric tests or non parametric tests). For the normality test, histograms for each distribution of scores supported by an inspection of the normal probability plots (labelled Normal Q-Q Plots) and the results of the Kolmogorov-Smirnov statistic were used to assess the normality of distributions (Appendix 9i to 9iv). The statistical significance tests were carried out based on comparison between
different input modalities in search performance and subjective satisfaction among all participants.

6.5.1 Comparison between using mouse and keyboard input modalities and using mouse and speech input modalities in search performance

Tests were conducted to check the normality of data before conducting the test. The result is shown in Appendix 9i, where it can be seen that the distribution of scores for both groups was reasonably ‘normal’. A Pair sample t-test was conducted to see the difference between using mouse and keyboard and mouse input modalities and speech input modalities in search performance. Table 6.10 shows a summary of the results.

Table 6.10: The Pair sample t-test for search performance using different input modalities

<table>
<thead>
<tr>
<th></th>
<th>MK* Search P*</th>
<th>MS* Search P*</th>
<th>t-value</th>
<th>df</th>
<th>2-tailed sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Search performance using different input modalities among the experienced.</td>
<td>2.01 0.52</td>
<td>1.26 0.31</td>
<td>6.13 19</td>
<td>0.000</td>
<td></td>
</tr>
</tbody>
</table>

Search P* = Search Performance.
MK* = Mouse and keyboard input modalities; MS* = Mouse and speech input modalities.

The result showed that there was significant difference in search performance between using mouse and keyboard input modalities and using mouse and speech input modalities. The null hypothesis (H1) is rejected as the 2-tailed sig. value in Table 6.10 is 0.000 (significance level < 0.05).
6.5.2 Comparison between using mouse and keyboard input modalities and using mouse and speech input modalities for subjective satisfaction

A Wilcoxon Signed Ranks test was conducted to see the difference between using mouse and keyboard and mouse and speech input modalities in subjective satisfaction. Table 6.11 shows a summary of the results.

Table 6.11: Wilcoxon Signed Ranks test for subjective satisfaction when using different input modalities.

<table>
<thead>
<tr>
<th>Subjective satisfaction (mouse &amp; speech) - Subjective satisfaction (mouse &amp; keyboard)</th>
<th>N</th>
<th>Mean Rank</th>
<th>Sum of Ranks</th>
<th>Z</th>
<th>Asymp. sig.(2-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Negative ranks</td>
<td>2(a)</td>
<td>12.50</td>
<td>25.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positive ranks</td>
<td>15(b)</td>
<td>8.53</td>
<td>128.00</td>
<td>-2.44</td>
<td>0.015</td>
</tr>
<tr>
<td>Ties</td>
<td>3(c)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>20</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

a subjective satisfaction (mouse and speech) < subjective satisfaction (mouse and keyboard)
b subjective satisfaction (mouse and speech) > subjective satisfaction (mouse and keyboard)
c subjective satisfaction (mouse and speech) = subjective satisfaction (mouse and keyboard)

The result showed that there was significant difference in subjective satisfaction between using mouse and keyboard input modalities and using mouse and speech input modalities. The null hypothesis (H2) is rejected (significance level < 0.05).

6.5.3 Comparison between using mouse and keyboard input modalities and using mouse and speech input modalities in photo browsing search performance.

Tests were conducted to check the normality of data before conducting the test. The result is shown in Appendix 9ii, where it can be seen that the distribution of scores for both groups was reasonably 'not normal'. A Wilcoxon Signed Ranks test was conducted to see the difference between using mouse and keyboard and
mouse and speech input modalities in photo browsing search performance. Table 6.12 shows a summary of the results.

Table 6.12: Wilcoxon Signed Ranks test for photo browsing search performance using different input modalities.

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Mean Rank</th>
<th>Sum of Ranks</th>
<th>Z</th>
<th>Asymp. sig. (2-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Photo browsing (mouse and speech) – Photo browsing (mouse and keyboard)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Negative ranks</td>
<td>15(a)</td>
<td>12.00</td>
<td>180.00</td>
<td>-2.80</td>
<td>0.005</td>
</tr>
<tr>
<td>Positive ranks</td>
<td>5(b)</td>
<td>6.00</td>
<td>30.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ties</td>
<td>0(c)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>20</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

a Photo browsing (mouse and speech) < Photo browsing (mouse and keyboard)
b Photo browsing (mouse and speech) > Photo browsing (mouse and keyboard)
c Photo browsing (mouse and speech) = Photo browsing (mouse and keyboard)

The result showed that there was significant difference in photo browsing search performance between using mouse and keyboard input modalities and using mouse and speech input modalities. The null hypothesis (H3) is rejected (significance level < 0.05).

6.5.4 Comparison between using mouse and keyboard input modalities and using mouse and speech input modalities in keyword searching search performance.

Tests were conducted to check the normality of data before conducting the test. The result is shown in Appendix 9iii, where it can be seen that the distribution of scores for one of the groups was reasonably 'not normal'. A Wilcoxon Signed Ranks test was conducted to see the difference between using mouse and keyboard and mouse and speech input modalities in keyword searching search performance. Table 6.13 shows a summary of the results.
Table 6.13: Wilcoxon Signed Ranks test for keyword searching search performance using different input modalities.

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Mean Rank</th>
<th>Sum of Ranks</th>
<th>Z</th>
<th>Asymp. sig.(2-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Keyword searching</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(mouse and speech)</td>
<td>19</td>
<td>10.18</td>
<td>193.50</td>
<td>-3.31</td>
<td>0.001</td>
</tr>
<tr>
<td>- Positive ranks</td>
<td>1</td>
<td>16.50</td>
<td>16.50</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Ties</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Total</td>
<td>20</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Keyword searching</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(mouse and keyboard)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a Keyword searching (mouse and speech) &lt; Keyword searching (mouse and keyboard)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b Keyword searching (mouse and speech) &gt; Keyword searching (mouse and keyboard)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c Keyword searching (mouse and speech) = Keyword searching (mouse and keyboard)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The result showed that there was significant difference in keyword searching search performance between using mouse and keyboard input modalities and using mouse and speech input modalities. The null hypothesis (H4) is rejected (significance level < 0.05).

6.5.5 Comparison between using mouse and keyboard input modalities and using mouse and speech input modalities in visual example searching.

Tests were conducted to check the normality of data before conducting the test. The result is shown in Appendix 9iv, where it can be seen that the distribution of scores for one of the groups was reasonably 'normal'. A Pair sample t-test was conducted to see the difference between using mouse and keyboard and mouse and speech input modalities in visual example searching search performance. Table 6.14 shows a summary of the results.
Table 6.14: Pair sample t-test for visual example searching search performance using different input modalities.

<table>
<thead>
<tr>
<th></th>
<th>MK* Search P*</th>
<th>MS* Search P*</th>
<th>t-value</th>
<th>df</th>
<th>2-tailed sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Mean</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>S.D.</td>
<td>S.D.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Visual example searching search performance using different input modalities.</td>
<td>0.62</td>
<td>0.33</td>
<td>3.85</td>
<td>19</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>0.35</td>
<td>0.17</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Search P* = Search Performance. MK* = Mouse and keyboard input modalities; MS* = Mouse and speech input modalities.

The result showed that there was significant difference in visual example searching search performance between using mouse and keyboard input modalities and using mouse and speech input modalities. The null hypothesis (H5) is rejected (significance level < 0.05).

6.6 Participants with significant experience in using online retrieval system

Of the twenty participants, twelve of them had considerable experience of using web applications (i.e. Google/Yahoo image search, eBay etc) that have digital photo or image retrieval features, having stated that they use these features more than 5 times per months (Figure 6.1). The task completion times for this group were examined to explore their search performance to complete each search task using mouse and keyboard and mouse and speech input modalities. The results showed that on average, this group (average score= 3.14) performed better than other participants (average score= 3.45) who have less experience of using digital photo or image retrieval features. It might have been predicted that the participants with more experience perform better.
6.7 Summary

In this chapter, the data from the evaluation of the prototype was presented. It was evaluated by 20 participants consisting of personal digital photo searchers. Logging into the system, understanding related help components and search performance data (in time taken to complete search tasks) was gathered using screen and voice recording software during the evaluation session. Data was also gathered after the evaluation session using questionnaires which were aimed at measuring subjective satisfaction and acceptability of the prototype when using mouse and keyboard input modalities and mouse and speech input modalities. Among the methods employed were percentage, mean, standard deviation and statistical techniques to compare groups. The percentage, mean and standard deviation were used to describe the behaviour of the respondents. Statistical techniques used to compare groups involved parametric and non parametric tests to test for significant differences between groups. Table 6.15 shows the list of the null hypotheses accepted or rejected.

Table 6.15: List of null hypotheses explored

<table>
<thead>
<tr>
<th>Null Hypotheses (H) explored</th>
<th>Sig.</th>
<th>R/A</th>
</tr>
</thead>
<tbody>
<tr>
<td>H1  There is no difference between using mouse and keyboard input modalities and using mouse and speech input modalities in search performance.</td>
<td>0.000</td>
<td>R</td>
</tr>
<tr>
<td>H2  There is no difference between using mouse and keyboard input modalities and using mouse and speech input modalities in subjective satisfaction.</td>
<td>0.015</td>
<td>R</td>
</tr>
<tr>
<td>H3  There is no difference between using mouse and keyboard input modalities and using mouse and speech input modalities in search performance in retrieving photos by browsing.</td>
<td>0.005</td>
<td>R</td>
</tr>
<tr>
<td>H4  There is no difference between using mouse and keyboard input modalities and using mouse and speech input modalities in search performance in retrieving photos by keyword searching.</td>
<td>0.001</td>
<td>R</td>
</tr>
<tr>
<td>H5  There is no difference between using mouse and keyboard input modalities and using mouse and speech input modalities in search performance in retrieving photos by visual example searching.</td>
<td>0.001</td>
<td>R</td>
</tr>
</tbody>
</table>

Sig. = Level of significance;  
R/A = R (Rejected) or A (Accepted)
CHAPTER 7

RESEARCH FINDINGS AND DISCUSSION

7.0 Introduction and research findings

This chapter presents and discusses the research findings. FlexPhoReS was motivated by the growth of digital photo technology and the expectation that a multimodal interface could be a more flexible, efficient and powerful method of human-computer interaction (Oviatt 2003, p.286). As stated in Chapter 1:

The aim of this research was to design and evaluate a flexible user interface for a web based personal digital photo retrieval system.

The above aim set the direction of the research and was broken down into a set of research objectives, which determined the tasks that were to be accomplished by this research. The findings are categorised into three phases.

7.1 Phase One

Objective: To design a model of a flexible user interface for a web based personal digital photo retrieval system.

In this phase, two major activities were conducted, namely, the development of FlexPhoReS model and a small-scale user study that served as an input to the development of the model along with the material from the literature review.
Literature review

An extensive survey of the literature was completed. A large number of references in refereed journals, theses, books and conference publications were consulted. The type of literature found, covered the general issues of user queries in image collections, image retrieval systems, user interface support for digital photo retrieval and multimodal interfaces. The review also helped to answer the following research question:

• What are the user interface properties and components suitable for a personal digital photo retrieval system?

Studies of user queries in image collections confirm that people would like to have their photos categorised and ordered to make searching and browsing easier. Some image users have very specific needs. However, most studies suggested that image retrieval systems should provide searching of image content by four main categories which are Who?, What?, When? and Where?. Most studies also suggested that content based retrieval could be applied at the browsing stage.

Studies of image retrieval system evolution (and some existing image retrieval system models) revealed that the features of stand-alone and networked image retrieval systems could also be applied to web based personal digital photo retrieval systems. It was found that there was a spectrum of related systems ranging from software that simply transfers digital photos from the digital camera to the computer, to those that can manage large collections of digital photos. Some of the systems provide limited content based retrieval (CBIR) features and query by keywords. None, however, used a multimodal user interface for a web based personal digital photo retrieval system.
From the perspective of user interface support for digital photo retrieval, the review suggested that a hybrid approach, which incorporates both text-based search and visual based search, offers a more attractive approach to personal digital photo retrieval.

The studies on multimodal interfaces revealed that, theoretically, a multimodal interface should provide a flexible user interface for a web based personal digital photo retrieval system.

Pre-development small-scale user study

The small-scale user study which was conducted helped to answer the following research question:

- What are the requirements of personal digital photo users for photo retrieval tasks?

In general, participants were happy at the prospect of more intuitive systems that might help them to browse and search more effectively. The findings showed that it was easy to retrieve photos when the collection was small. It became difficult, however, to locate photos when the number increased. The findings also brought to light the difficulty of remembering digital photo file names that are automatically created by the digital camera. As a result, it is not always easy for people to remember photos when they need them. They may have to browse through all of their collection before they can select the appropriate photos. These findings are consistent with earlier results obtained from photo retrieval user behaviour studies by Rodden (1999) and Markkula and Sormumen (2000). Allowing browsing via a set of user-oriented categories would give more specific access points to the participants.
The findings also showed an enthusiastic response to the prospect of automatic CBIR. Although the participants had very limited knowledge of content based similarity and its use in photo retrieval, a few participants noted that they do organise their digital photos collection based on visual similarity from their own observation but this was seen as complex and time consuming.

Among the features that the users would like to have are retrieval by photo events (what?), time (when?), places (where?), photo subject/person (who?) and visual example searching.

The data from the small scale user study also confirmed that a web based photo retrieval environment could be a better choice than a stand-alone environment. The World Wide Web environment can offer security and allows users to access their photos anytime and anywhere which would give more flexibility and would be transparent to the users.

As regards multimodal interaction in photo retrieval applications, the small scale user study also revealed that participants were positive towards using speech interaction in addition to the more normal mouse and keyboard input modalities.

As a whole, the small-scale user study was successful in extracting the extra information needed from the participants for the development of the prototype.

**FlexPhoReS model**

Task descriptions with a ‘use case’ diagram were used to convey and envision the user tasks through the photo retrieval process model. The model provides a flexible user interface which allows users to retrieve their photo collections through either mouse and keyboard input modalities or mouse and speech input modalities. Users across the World Wide Web network will be able to use the system with the same user interface through a web browser.
Chapter 7

Research Findings and Discussion

7.2 Phase Two

Objective: To develop a prototype of a flexible user interface for a web based personal digital photo retrieval system.

The phase two research objective was concerned with the implementation of the FlexPhoReS model. This phase saw the development of the prototype and answers the following research question:

- How to design and demonstrate the ability of a flexible user interface (for a web based personal digital photo retrieval system) which allows users to utilise mouse and keyboard input modalities and mouse and speech input modalities to retrieve digital photos through a World Wide Web environment?

The prototype was constructed based on the process model designed in Phase One. The prototype interface was implemented incorporating photo retrieval based on photo features explained in Phase One. The initial design of the prototype interface was sketched on paper and then mocked-up for a low fidelity prototype using Microsoft PowerPoint with audio output to clarify user requirements related to evaluating the screen design and learning how to use the system. The findings showed that the screen design for the low fidelity prototype system was clear, adequate and easy to understand. However, some modifications were made based on some issues that emerged and recommendations (explained in Chapter 5.2) and these were input into the development of the high fidelity prototype.

Another important finding was that the implementation of the prototype using web server services resulted in easier external integration. The multiplicities of platforms that support the World Wide Web environment allow the prototype to
run through a web browser at any time, anywhere in heterogeneous environments, and hence achieved improved external integration.

7.3 Phase Three

Objective 3: To evaluate the prototype in order to measure users’ search performance, subjective satisfaction and acceptability of the system.

In Phase Three, a user evaluation of the FlexPhoReS prototype was conducted with a group of digital photo users. Search performance data (time taken to complete search tasks) was gathered using screen and voice recording software. Questionnaires were designed to measure subjective satisfaction of the prototype in using mouse and keyboard input modalities and mouse and speech input modalities. The questionnaires were also designed to establish whether the prototype was suitable and flexible enough for the photo retrieval process. The evaluation of the prototype helped answer the following research question:

- Will the use of mouse and speech input modalities improve users’ search performance (in time taken to complete search tasks) and receive subjective satisfaction from digital photo users compared to mouse and keyboard input modalities?

The evaluation was carried out based on the dependent and independent variables identified (Section 3.2.6).
7.3.1 Comparison between different input modalities in search performance and subjective satisfaction among all participants

Search performance

A comparison between different input modalities in search performance was undertaken. Table 7.1 shows the average time taken to complete each search task, overall search performance, as well as the percentage reduction in time when using mouse and speech input modalities as compared with using mouse and keyboard input modalities.

Table 7.1: Average time taken to complete search tasks and percentage reduction with different input modalities

<table>
<thead>
<tr>
<th>Participant Group</th>
<th>Input modalities with FlexPhoReS</th>
<th>Specific search tasks (minute)</th>
<th>Overall search task performance (minutes)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Photo browsing</td>
<td>Keyword searching</td>
</tr>
<tr>
<td>All</td>
<td>Mouse and keyboard</td>
<td>0.84</td>
<td>0.55</td>
</tr>
<tr>
<td></td>
<td>Mouse and speech</td>
<td>0.56</td>
<td>0.37</td>
</tr>
<tr>
<td>Percent reduction</td>
<td></td>
<td>33.33%</td>
<td>32.73%</td>
</tr>
</tbody>
</table>
Table 7.1 and Figure 7.1, show that, on average participants needed less time to complete search tasks when they used mouse and speech input modalities compared with mouse and keyboard input modalities. Participants took 1.26 minutes to complete search tasks with mouse and speech input modalities whereas they took 2.01 minutes with mouse and keyboard input modalities. On average, the reduction in search performance time due to using mouse and speech input modalities was 37.31%.

Participants also needed less time to complete photo browsing (average reduction: 33.33%), keyword searching (average reduction: 32.73%) and visual example searching search tasks (average reduction: 46.77%) when they used mouse and speech input modalities compared with mouse and keyboard input modalities. The result implies that among the different retrieval strategies, visual example searching shows a higher reduction percentage when participants used mouse and speech input modalities.
General findings from these data show that mouse and speech input modalities are faster. There are various issues to be considered here. Participants used mouse and speech input modalities last in their session. They already had the possibility of 20 minutes free exploration and also had performed similar tasks with mouse and keyboard which will have built up some familiarity with the user interface. Speaking a word or phase is generally quicker than typing the same text (Rebman et al. 2003, p.513). Nevertheless the difference in time taken for the tasks with different input modalities is substantial in all three search types and demonstrates participants were able to carry out the tasks using speech without any problems.

**Subjective satisfaction**

Comparisons between the different input modalities in participants' subjective satisfaction scores with FlexPhoReS were collected and analysed. Average scores of data collected through the subjective satisfaction questionnaires across both input modalities are shown in Table 7.2 and Figure 7.2.

Table 7.2: Means (average scores) of subjective satisfaction for FlexPhoReS across both input modalities.

<table>
<thead>
<tr>
<th>Subjective satisfaction question</th>
<th>Mouse and keyboard</th>
<th>Mouse and speech</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall reactions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Terrible (1-9) Wonderful</td>
<td>6.90</td>
<td>7.55</td>
</tr>
<tr>
<td>Frustrating (1-9) Satisfying</td>
<td>7.05</td>
<td>7.60</td>
</tr>
<tr>
<td>Dull (1-9) Stimulating</td>
<td>6.90</td>
<td>7.75</td>
</tr>
<tr>
<td>Difficult (1-9) Easy</td>
<td>7.15</td>
<td>7.90</td>
</tr>
<tr>
<td>Inadequate power (1-9) adequate power</td>
<td>6.95</td>
<td>7.55</td>
</tr>
<tr>
<td>Rigid (1-9) Flexible</td>
<td>6.50</td>
<td>6.45</td>
</tr>
<tr>
<td><strong>Login and understanding system help</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Login into the system: Difficult (1-9) Easy</td>
<td>7.80</td>
<td>7.95</td>
</tr>
<tr>
<td>Help components: Difficult (1-9) Easy</td>
<td>7.30</td>
<td>7.65</td>
</tr>
<tr>
<td><strong>Search tasks with difference retrieval strategies</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Photo browsing: Difficult (1-9) Easy</td>
<td>7.55</td>
<td>7.85</td>
</tr>
<tr>
<td>Keyword searching: Difficult (1-9) Easy</td>
<td>7.65</td>
<td>8.05</td>
</tr>
<tr>
<td>Visual example searching: Difficult (1-9) Easy</td>
<td>7.00</td>
<td>7.75</td>
</tr>
<tr>
<td><strong>Average:</strong></td>
<td>7.16</td>
<td>7.64</td>
</tr>
</tbody>
</table>
Figure 7.2: Means (average scores) of subjective satisfaction for FlexPhoReS across difference input modalities

Although all of the participants were satisfied with both input modalities, the subjective satisfaction score related to overall reaction indicated that mouse and speech input modalities were preferred in almost all cases. The differences in average score (0.48) are generally not large.

Participants felt that mouse and speech input modalities were easier for browsing (subjective satisfaction average scores = 7.85), keyword searching (subjective satisfaction average scores = 8.05) and visual example searching (subjective satisfaction average scores = 7.75) compared to mouse and keyboard input
modalities for browsing (subjective satisfaction average scores = 7.55), keyword searching (subjective satisfaction average scores = 7.65) and visual example searching (subjective satisfaction average scores = 7.00). These results were expected as the search performance results (Table 7.1) showed that participants required a longer time to complete all of the search tasks when using mouse and keyboard input modalities.

Participants on average felt that logging in using mouse and speech input modalities was easier than mouse and keyboard. This result was expected as they required a longer time to login into FlexPhoReS system when using mouse and keyboard input modalities (Table 6.3).

However participants believed that mouse and keyboard input modalities were more flexible than mouse and speech input modalities. This result for rigidity/flexibility was expected as during the free exploration period, participants assumed that they could speak any word rather than the limited number of words allowed by the prototype (Section 5.3).

**Tests for significance**

The positive results for mouse and speech input modalities were tested for statistical significance. Table 7.3 indicates some variation in the comparison between different input modalities in search performance and subjective satisfaction among all participants.
Table 7.3: Statistical significance tests for different input modalities in search performance and subjective satisfaction

<table>
<thead>
<tr>
<th>Variable &amp; (Hypothesis)</th>
<th>Statistical test</th>
<th>Level of significance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pair sample t-test</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mouse and keyboard Mean S.D.</td>
<td>Mouse and speech Mean S.D.</td>
</tr>
<tr>
<td>Overall search performance (mouse &amp; keyboard) – Overall search performance (mouse &amp; speech) (H1)</td>
<td>2.01 0.52</td>
<td>1.26 0.31</td>
</tr>
<tr>
<td>Visual example searching (mouse &amp; keyboard) – Visual example searching (mouse &amp; speech) (H5)</td>
<td>0.62 0.35</td>
<td>0.33 0.17</td>
</tr>
<tr>
<td></td>
<td>Wilcoxon Signed ranks test</td>
<td></td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>Mean Rank</td>
</tr>
<tr>
<td>Photo browsing (mouse &amp; speech) – Photo browsing (mouse &amp; keyboard) (H3)</td>
<td>Negative ranks</td>
<td>15(a)</td>
</tr>
<tr>
<td></td>
<td>Positive ranks</td>
<td>5(b)</td>
</tr>
<tr>
<td></td>
<td>Ties</td>
<td>0(c)</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>20</td>
</tr>
<tr>
<td>Keyword searching (mouse &amp; speech) – Keyword searching (mouse &amp; keyboard) (H4)</td>
<td>Negative ranks</td>
<td>19(a)</td>
</tr>
<tr>
<td></td>
<td>Positive ranks</td>
<td>1(b)</td>
</tr>
<tr>
<td></td>
<td>Ties</td>
<td>0(c)</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>20</td>
</tr>
<tr>
<td>Subjective satisfaction (mouse &amp; speech) – Subjective satisfaction (mouse &amp; keyboard) (H2)</td>
<td>Negative ranks</td>
<td>2(a)</td>
</tr>
<tr>
<td></td>
<td>Positive ranks</td>
<td>15(b)</td>
</tr>
<tr>
<td></td>
<td>Ties</td>
<td>3(c)</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>20</td>
</tr>
</tbody>
</table>

Table 7.3 reveals that the significance level in comparing the different input modalities in search performance and subjective satisfaction are lower than the 0.05 confidence level. The t value, a negative Z value and a corresponding p value of <0.05 indicates a significant difference in search performance and subjective satisfaction between using mouse and keyboard input modalities and using mouse and speech input modalities. This statistic implies that the H1, H2, H3, H4 and H5 respectively can be rejected.

A comparison between the different input modalities in overall search performance alone showed that there was significant difference in search performance between using mouse and keyboard input modalities (M=2.01,
and using mouse and speech input modalities \( [M=1.26, SD=0.31, t(19)=6.13, p=0.000] \). Participants needed significantly less time to complete search tasks when they used mouse and speech input modalities. The standard deviation (SD) values with range from 0.31 to 0.52 represent a tightly clustered distribution around the mean, which implies that the participants' search performances are very similar to each other.

In terms of specific search performance, the result showed that there was a significant difference in browsing between using mouse and keyboard input modalities and using mouse and speech input modalities \( (Z=-2.80 \text{ and Asymp.Sig (2tailed) } = 0.005) \). There was also a significant difference in keyword searching between using mouse and keyboard input modalities and using mouse and speech input modalities \( (Z=-3.31 \text{ and Asymp.Sig (2tailed) } = 0.001) \).

The comparison between different input modalities in visual example search performance showed that there was a significant difference in visual example search performance between using mouse and keyboard input modalities \( (M=0.62, SD=0.35) \) and using mouse and speech input modalities \( [M=0.33, SD=0.17, t(19)=3.85, p=0.001] \). Participants need significantly less time to complete visual example search tasks when they used mouse and speech input modalities. The small SD values with range from 0.35 to 0.62 imply that the participants' visual example search performances are very similar to each other.

The study of subjective satisfaction showed that there was also significant difference in subjective satisfaction between using mouse and keyboard input modalities and using mouse and speech input modalities. Participants were significantly more satisfied with mouse and speech input \( (Z=-2.44 \text{ and Asymp.Sig (2tailed) } = 0.015) \).
7.3.2 Acceptability of FlexPhoReS

Participants' acceptability scores with FlexPhoReS were collected. Means and standard deviations of data collected through the acceptability questionnaires of all participants are shown in Table 7.4. Based on the same data, Figure 7.3 illustrates the comparison in acceptability means among all participants.

Table 7.4: Average acceptability rating and standard deviations of FlexPhoReS among all participants

<table>
<thead>
<tr>
<th>Questionnaires</th>
<th>Average</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1 Suitability of FlexPhoReS system with mouse and keyboard input modalities for photo retrieval tasks</td>
<td>7.35</td>
<td>1.14</td>
</tr>
<tr>
<td>Q2 Suitability of FlexPhoReS system with mouse and speech input modalities for photo retrieval tasks</td>
<td>7.70</td>
<td>0.98</td>
</tr>
<tr>
<td>Q3 Suitability of FlexPhoReS system with both mouse and keyboard and mouse and speech input modalities for photo retrieval tasks</td>
<td>7.80</td>
<td>1.20</td>
</tr>
<tr>
<td>Q4 Both input modalities are complementary each other to retrieve photos</td>
<td>7.45</td>
<td>1.57</td>
</tr>
<tr>
<td>Q5 FlexPhoReS offer better flexibility than a system that only offer one style of input modalities either the mouse and speech support or the mouse and keyboard support.</td>
<td>7.75</td>
<td>1.12</td>
</tr>
<tr>
<td>Q6 FlexPhoReS offer better flexibility on the World Wide Web environment then on the stand-alone environment.</td>
<td>8.00</td>
<td>1.08</td>
</tr>
</tbody>
</table>
Figure 7.3: Acceptability mean of all participants

Analysis of the acceptability of FlexPhoReS revealed that all of the related questions achieved positive scores (above 4.5). The most favourable responses among all participants was related to the flexibility of FlexPhoReS in the World Wide Web environment rather than in a stand-alone environment (Mean=8.00, standard deviation=1.08). The least favourable response was related to suitability of FlexPhoReS system with mouse and keyboard input modalities for photo retrieval tasks (Mean=7.35, standard deviation=1.14).

Figure 7.4 shows in detail that participants agreed that mouse and keyboard input modalities are suitable for photo retrieval tasks. Six participants chose 6 on the scale, five participants chose 7, five participants chose 8 and four participants chose 9.
Figure 7.4: Suitability of mouse and keyboard input modalities of FlexPhoReS for photo retrieval tasks

Figure 7.5 shows in detail that participants agreed that mouse and speech input modalities are suitable for photo retrieval tasks. Two participants chose 6 on the scale, seven participants chose 7, six participants chose 8 and five participants chose 9.

Figure 7.5: Suitability of mouse and speech input modalities of FlexPhoReS photo retrieval tasks

Figure 7.6 shows in detail that participants agreed that FlexPhoReS system with both mouse and keyboard and mouse and speech input modalities are suitable
for photo retrieval tasks. Four participants chose 6 on the scale, four participants chose 7, four participants chose 8 and eight participants chose 9.

![Bar graph showing the majority of participants agreed that both input modalities were complementary to each other when retrieving photos.]

Figure 7.6: Suitability of both input modalities for photo retrieval tasks

The majority of participants also agreed that both input modalities were complementary to each other when retrieving photos. One participant chose 4 on the scale, two participants chose 5, two participants chose 6, four participants chose 7, four participants chose 8 and seven participants chose 9 (see Figure 7.7).

![Bar graph showing that the majority of participants agreed that both input modalities are complementary for photo retrieval tasks.]

Figure 7.7: Both input modalities are complementary for photo retrieval tasks
Figure 7.8 shows in detail that the majority of the participants agreed that FlexPhoReS offers better flexibility over a system that only offers one style of input modalities (either mouse and speech support or the mouse and keyboard support). One participant chose 5 on the scale, eight participants chose 7, three participants chose 8 and seven participants chose 9.

![Bar chart showing the results of the agreement scale for FlexPhoReS flexibility.](chart.png)

Figure 7.8: FlexPhoReS flexibility

Figure 7.9 shows in detail that majority of the participants agreed that FlexPhoReS offers better flexibility by being implemented in the World Wide Web environment. One participant chose 5 on the scale, five participants chose 7, six participants chose 8 and eight participants chose 9.
It is clear that the participants improved search performance by significantly reducing their overall search task completion times when using mouse and speech input modalities. On average the reduction in search performance time was 37.31%. Similarly, participants also needed significantly less time to complete all of the specific search tasks, including browsing, keyword searching and visual example searching when they used mouse and speech input modalities. This result was expected as mouse and speech input modalities involved fewer steps in completing search tasks. With mouse and speech input modalities, the participants simply ‘click/tap and talk’ and the FlexPhoReS system conducts the search. With mouse and keyboard input modalities, participants had to go through several steps with extensive “head-down time” and potentially error-prone keyboard typing or button selections. These findings are consistent with other studies in the field that found similar user performance differences on using speech input with other input channels (Pausch and Leatherby 1991a; Karl et al. 1993).

The participants also reacted positively to their experience with both input modalities. However, the subjective satisfaction results showed that participants
were significantly more satisfied with mouse and speech input modalities than with mouse and keyboard input modalities. With some variations, this result reflected similar findings to those from other studies on user satisfaction on using speech input with another input channel (Karl et al. 1993; Käster et al. 2003).

The study of specific retrieval strategies in user search performance provided some interesting results. The results showed that visual example searching appeared to have the greatest percentage reduction (average reduction: 46.77%) when participants used mouse and speech input modalities. Figure 7.10 shows the average percentage reduction in time between mouse and keyboard and mouse and speech input modalities and retrieval strategies.

![Figure 7.10: Graph of Percent reduction vs. Retrieval strategies](image)

This result was expected, because even though mouse and speech input modalities should speed up participants’ retrieval search performance, most of the participants were very new to visual example searching and found that selecting
and locating photos is not an easy task, especially when they had to browse through a set of photos on the screen. Time was lost due to the fact that the participants had to navigate or browse through the screen in order to select and click or type in the appropriate photo file name when using mouse and keyboard input modalities. Although participants also had to navigate or browse through the screen in order to say the appropriate photo file name when using mouse and speech input modalities, the process required fewer steps.

**Further discussion**

The above results reinforce that overall, participants improved their search performance when using mouse and speech input modalities. Therefore as might be expected, participants were significantly more satisfied with mouse and speech input modalities than with mouse and keyboard input modalities. There are three primary reasons for mouse and speech input being faster overall. First, speech is a natural form of communication and it is simpler to use as compared to mouse and keyboard. Using mouse and keyboard input modalities required that participants spend more time in locating and understanding the structure of the prototype interface and how to perform the search tasks.

Second, speech input can be recognised at a rate faster than many people can type. Time was lost due to the fact that the participants had to type in or navigate before they could execute their search tasks. Speech input rate could have helped any participant who had less well developed motor skills in using mouse and keyboard.

Third, using mouse and keyboard input modalities required that participants might have removed their eyes from screen and keyboard in order to avoid errors. Again time was lost due to the fact that participant had to find his/her place again on the screen or on the keyboard.
The study of acceptability of the input modalities of FlexPhoReS revealed that all of the participants agreed that mouse and keyboard input modalities by themselves are suitable for photo retrieval tasks. They also agreed that mouse and speech input modalities alone are suitable. A higher acceptability rate was given when both input modalities were considered together and the majority of participants agreed that both input modalities are complementary to each other in retrieving photos. Several participants stated that both input modalities were user friendly, practical and easy to use. A number of participants noted that mouse and speech input modalities were more interesting and easier for retrieving photos instead of mouse and keyboard input modalities. They noted, however, that mouse and speech input modalities are very sensitive to noise. Among the suggestions made was that noise reduction is essential to improve FlexPhoReS system performance.

Exploration of the flexibility of the prototype showed that the majority of participants agreed that FlexPhoReS system offers better flexibility than a system that only offers one style of input modalities. In terms of system platform environment, the majority of the participants agreed that FlexPhoReS system offers better flexibility on the World Wide Web environment than on a stand-alone environment.

In choosing their favourite input modalities to use for photo retrieval tasks, participants' responses are illustrated in Figure 7.11.
Mouse and speech input modalities

Mouse and keyboard input modalities

Switch between both input modalities

Don't care

Don't know

Figure 7.11: Participants' choices of FlexPhoReS input modalities for photo retrieval tasks.

The results showed that the majority (10) of the participants chose mouse and speech input modalities. Among the reasons given were that mouse and speech input provides an easier, faster and efficient input method over the alternative. Mouse and speech input modalities are also more convenient and time saving to input information and data. Four participants preferred mouse and keyboard input modalities. Among the reasons given were that they have been familiar with keyboard input method for a very long time and unfamiliar with speech input modalities. Speech recognition errors also could disturb the accuracy of data input. Four participants were happy to switch between both input modalities, giving the reasons that while speech input is easier, typing is also very useful depending on the environment.

Finally, participants were asked to make comments related to the use of both input modalities with the FlexPhoReS system. There was a general consensus among the participants about the use of both input modalities with the prototype. Several participants mentioned that both input modalities were user friendly, practical and easy to use. A number of participants commented on the
comparison between the use of mouse and speech input modalities and mouse and keyboard input modalities. Four participants noted that mouse and speech input modalities were more interesting and easier to retrieve photos instead of mouse and keyboard input modalities.

Other comments included from one participant who noted that because visual example searching requires identification of the sample image, the identification process will take considerable time if the photo database is large. One participant noted that because of not being familiar with the photo filenames, visual example searching seemed slightly rigid to execute.

7.4 Summary

The findings, which are based on the analysis of the data and the user evaluations, were compared against the original aim and objectives. It can be concluded that this research has met the aims and objectives previously outlined. It has:

- Developed a flexible user interface for web based digital photo retrieval model
- Implemented a prototype based on the model
- Evaluated the prototype and found the results of digital photo searchers' search performance, subjective satisfaction and acceptability of the prototype.

The end result also showed that the system's flexibility met the definition of flexibility outlined in Chapter one. The prototype was confirmed to be flexible in terms of the following:
• It allowed digital photo users to utilise both mouse and keyboard input modalities and mouse and speech input modalities to retrieve digital photos through a World Wide Web environment.

• It allowed digital photo users to use photo browsing, keyword searching and visual example searching features to retrieve personal digital photos.

• As it is available across the World Wide Web environment, it allows digital photo users to retrieve digital photos at any time and from any place through a web browser.
CHAPTER 8

CONCLUSIONS AND FURTHER WORK

8.0 Introduction

This chapter presents the overall conclusion of this thesis.

- Section 8.1 presents the research summary and achievements.
- Section 8.2 presents some general conclusions.
- Section 8.3 provides the research implications.
- Section 8.4 presents the limitations and suggestions for further work.

8.1 Research summary and achievements

The aim of this research was to design and evaluate a flexible user interface for a web based personal digital photo retrieval system.

- It employed the Lifecycle model for interaction design system development (Preece et al. 2002, p.186) that consists of several steps: identify needs/establish requirements, (Re) design, build an interactive version and evaluation.

- Chapter 2 reviewed image retrieval evolution, studies related to user queries in image collection, studies of personal digital photos, user interface support for digital photo retrieval and multimodal interfaces. The analysis indicated that none of the existing systems makes use of
multimodal user interface for digital photo retrieval in World Wide Web environment.

- Chapter 3 discussed the research methodology, research design and data collection methodology used in this research.

- Chapter 4 introduced system modelling of the FlexPhoReS system. It explained the proposed model in relation to the literature review and small scale-user study outcomes, task description and process model of the FlexPhoReS.

- Chapter 5 explained the development of the prototype and presented the architecture and components of the system. It also demonstrated the capability of FlexPhoReS.

- The evaluation results of FlexPhoReS were demonstrated in Chapter 6. The prototype developed was evaluated by 20 digital photo participants. Several parameters including logging into the system, understanding related help components and search performance data in time taken to complete search tasks was captured using screen and voice recording software.

- The data gathered from the evaluation were analysed and measured based on the participants’ search performance, subjective satisfaction and acceptability of the prototype. Chapter 7 analysed the research findings and discussions.
8.2 Research conclusions

This section re-examines the information presented in all the previous chapters and presents the general conclusions of this research.

8.2.1 Literature background

This research covered two main areas: personal digital photo retrieval and multimodal user interface. It started with a review of image retrieval evolution. A review from the perspective of first generation and later developments in image retrieval systems was presented. In order to develop a useful image retrieval system, it is important to explore the user aspects of image retrieval. Therefore, the literature review was expanded further to include issues on user studies in image collections: the image user, previous research on user queries in image collections and studies of personal digital photos. It was found that previous research had focused on the needs of specific image user groups and searching behaviour. The studies confirmed that some image users (including personal digital photo users) have very specific needs (Rodden 1999; Frohlich et al. 2002; Rodden and Wood 2003; Crabtree et al. 2004; Davis et al. 2004; Wilhelm et al. 2004; Van House et al. 2005).

With this background, the topic was focused further on the issue of user interface support for digital photo retrieval. The interface design framework for the information retrieval (Shneiderman et al. 1997) was presented along with a discussion of digital photo annotation, user visual access modes and retrieval strategies. Several related systems were presented including QBIC (Flickner et al. 1995), Photobook (Pentland et al. 1996), MARS (Huang et al. 1996), Aria (Lieberman et al. 2001), query by groups (Nakazato et al. 2003), PhotoFinder (Kang and Shneiderman 2000), PhotoTOC (Platt et al. 2003), Show&Tell (2000), AT&T Shoebox (Mills et al. 2000), SmartAlbum (2002), Multimodal interface for CBIR (Käster et al. 2003), MMM (Davis 2004) and MMM2 (Davis et al. 2005).
was found that none of the above models and systems have employed multimodal user interface in retrieving personal digital photo in a web based environment.

Further studies on the subject of a multimodal user interface were reviewed. Classification of multimodal user interface, the advantages of multimodal and speech interface were reviewed. Numerous multimodal user interface systems have been tested, built to demonstrate the usefulness of multimodal interaction. Several studies comparing the effectiveness of speech input with other input channels were reviewed (Schmandt et al. 1990; Pausch and Leatherby 1991a; Pausch and Leatherby 1991b; Karl et al. 1993; Molnar and Kletke 1996; Oviatt et al. 1997; Karat et al. 1999; Christian et al. 2000; Myers et al. 2002; Lamel et al. 2002; Käster et al. 2003; Lahtinen and Peltonen 2005). The above studies revealed that theoretically, such an interface should provide a flexible user interface for a web based personal digital photo retrieval.

To summarise, the most popular use of personal digital photos is sharing online, either via e-mail or web sites and undoubtedly the internet platform will play an important role in transporting personal digital photos collections. With multimodal interface, the weaknesses of one modality are offset by the strengths of another. Current personal digital photo retrieval models and systems have not yet resolved these issues. This research addresses the design issue, proposing a flexible personal digital photo retrieval model and implementing a prototype system.

8.2.2 The aim

The aim of the research was to design and evaluate a flexible user interface for a web based personal digital photo retrieval system which would allow digital photo users more flexibility in performing personal digital photo retrieval.
Specifically, the development of the FlexPhoReS system included:

- The development of the flexible user interface for a web based personal digital photo retrieval process model
- Implementation of a prototype based on the model of the system
- The evaluation of the prototype in order to measure user's search performance, subjective satisfaction and acceptability of the prototype.

The flexibility that this research was looking for in the model and the prototype (as defined in Chapter 1) is as follows:

- The ability to allow digital photo users to utilise mouse and keyboard input modalities and utilise mouse and speech input modalities to retrieve digital photo collection through a World Wide Web environment
- The ability to allow digital photo users to use photo browsing, keyword searching and visual example searching features to retrieve personal digital photos collection.
- The ability to use the prototype across the World Wide Web environment which allowed digital photo users to retrieve digital photos at any time and any place through web browser.

This research was designed to be experimental and had three phases; each phase corresponding to a specific objective of the FlexPhoReS development outlined. The first phase was to develop a model for the flexible personal digital photo retrieval process; the second phase was to develop the flexible personal digital photo retrieval prototype; and the third phase was to evaluate the
prototype. Statistical techniques were used in the analysis, including mean, standard deviation, parametric and non parametric statistical tests.

8.2.3 FlexPhoReS model

The development of the FlexPhoReS model proposed was based on a small-scale user study and incorporated some of the findings reported in the literature (Shneiderman et al. 1997, SALT Forum 2002, Flickner et al. 1995). The model aimed to support the flexibility aspect, that is, the ability to allow digital photo users to use either mouse and keyboard input modalities or mouse and speech input modalities to retrieve personal digital photos in the World Wide Web environment.

The construction of the model started with the conceptual framework of FlexPhoReS which comprised the assumptions, considerations and concepts that were illustrated through the task description with the 'use case' diagram. The small-scale user study and some of the findings reported in the literature review served as the basis for the model. The tasks description and the 'use case' diagram for the proposed system serve as the basis for the process model of FlexPhoReS which is used to describe the structure of prototype system and represents the process flow.

8.2.4 FlexPhoReS prototype

The FlexPhoReS model was the basis for the prototype. The development of the prototype was targeted to accommodate the other two issues in flexibility, namely: the ability to use the prototype across the World Wide Web environment and the ability to let digital photo users choose either mouse and keyboard input modalities or mouse and speech input modalities that fit their requirement and styles to retrieve personal digital photos.
Chapter 8

Conclusions and Further Work

The FlexPhoReS architecture, based on the World Wide Web consists of several viewpoints that are based on a Client-Server Interaction perspective. The components of the FlexPhoReS prototype include user profiles, repository and access control, multimodal interaction, retrieval strategies, photo features and help.

Among the technologies that were used in order to achieve the aim were: the World Wide Web, Speech technologies and SALT (Speech Application Language Tags), Matrix Laboratory (MATLAB), Active Server Pages (ASP) and JavaScript.

8.2.5 Research findings and evaluations

From the analysis and user evaluations, it can be concluded that this research has met the aim outlined in the first chapter. The findings, based on the analysis and the evaluation were compared against the original aim, objective and research questions. This research has developed a flexible personal digital photo retrieval process model, implemented a prototype based on the model, evaluated the prototype in order to measure users search performance, subjective satisfaction and acceptability of the prototype input modalities. The end result also shows that the prototype’s flexibility met the definition of flexibility outlined in Chapter 1.

8.3 Research implications

Since this study is one of the first attempts to integrate multimodal user interface into a web based personal digital photo retrieval system, it produced both theoretical justification and empirical data which might have some practical relevance for the field of digital photo retrieval user interfaces. Accordingly, this section reviews the implications of this study in several respects:
8.3.1 Implications for knowledge

The present study fills a gap in both the digital photo retrieval and multimodal user interface research. Firstly, the contribution is that in reviewing the literature it offers an instructive understanding of user queries in image collection and user interface support for digital photo retrieval. This is important since the development of user interfaces is concerned with the implementation of systems for human use and the study of major phenomena surrounding them. Secondly, the study provided a comprehensive survey of digital photo retrieval evolution, existing digital photo retrieval systems and multimodal user interface design. This study will add a substantial new dimension to the literature. This is because previous studies had predominantly focused on the efficiency of photo retrieval engine components, while there are not many studies concerning multimodal user interface support for digital photo retrieval applications. Therefore, the prototype is a novel approach to user interface for a web based personal digital photo retrieval system. It improves the traditional mouse and keyboard input modalities approaches by considering mouse and speech input modalities as complementary and an alternative. Consequently, the implementation of FlexPhoReS has strengthened the domain of multimodal user interface and web based personal digital photo retrieval systems by integrating them in one system environment.

- Design of the user interface and retrieval strategies

One of the main research objectives was to develop a flexible user interface for web based personal digital photo retrieval system. This resulted in the development of the FlexPhoReS model that re-defined traditional retrieval approaches by means of a novel user interface incorporating speech retrieval.
In the area of digital photo retrieval, there are three state-of-the-art retrieval strategies, i.e., browsing, query by text or keyword, and query by visual example. Most common approaches attempt to provide these retrieval strategies through a graphical user interface with mouse and keyboard, while this research attempts to provide a more multi-dimensional focus with multimodal interaction.

FlexPhoReS advocates the use of the 4 main categories template (who?, what?, when?, where?) for photo browsing and navigation strategies and add query by description or extra note as an additional category to query by text or keyword. Simultaneously with regard to query by visual content, query by photo example was used. Similar to existing system approaches FlexPhoReS provides an interface that is understandable by digital photo users. However, it enhances the structure of the retrieval strategies to make them suitable for linking to a multimodal user interface.

Although many image users have been reluctant to use query by visual content as a retrieval strategy (Eakins et al. 2004), it is believed that easy and intuitive handling could motivate image users to use this method (Käster et al. 2003). The interface should demonstrate the ability to hide the complexity and therefore support broader user communities by the use of multimodal interaction. Although FlexPhoReS only provides mouse and keyboard and mouse and speech input modalities, it is believed that it has made a step towards more intuitive and interactive web-based retrieval for commercial application given that with mouse and speech input modalities, participants were more satisfied and significantly reduced their search completion time in all of the retrieval tasks (including visual retrieval). This may help to set a reasonable research agenda toward natural interaction with unlimited vocabulary in spoken conversational multimodal user interface for the photo retrieval system application.
When projecting from a prototype to a real system, a new set of variables come into play, principally: 1) a large number of photos; 2) systems that know more about users and their preferences and; 3) the need for a vastly increased vocabulary size of words needed for speech retrieval. A viable full-scale system would incorporate a robust, intelligent, fast processor to deal with the increased volume of photos and as reported in the literature, considerable progress is being made in the development of speech interfaces (Bernsen 2001). It is therefore projected that the positive performance profile of FlexPhoReS would be maintained.

- Role of speech in multimodal user interface

Although mouse and keyboard are effective input modalities in most cases, they are limited in many ways and do not provide fully natural communication between humans and computers. Speech has great potential for becoming a key modality in future interactive systems. It holds great promise for simplifying the data entry options in photo retrieval applications. Observers have noted that computer applications will increasingly offer users the opportunity to interact through speech technologies (Bernsen 2001). Given its naturalness, speech is increasingly being used not only as a stand-alone modality, but in combination with others. The FlexPhoReS contribution to multimodal user interface lies in the use of mouse and speech input modalities as an additional or alternative to mouse and keyboard input. It provides traceability of the multimodal user interface in the existing web based (client-server) photo retrieval system architecture. It also shows that speech elements are no longer separate from other fields of research and with speech, FlexPhoReS has offered a more natural language interface for photo retrieval web applications where it will be easier to say a sentence related to a photo as an alternative to typing.
Using speech input, FlexPhoReS provides a new way to allow users to enjoy a level of independence in controlling application menus and instructions. Users can execute commands which are embedded in the application menu structure by saying appropriate words which may be easier than other input devices.

Other than the above contributions, FlexPhoReS also contributes a convenient way to manage speech elements with GUIs which include system dialogues on how to keep track of the users when the system is listening and when it is not listening and convey relevance information to the users. Tap and talk is a proven and widely-accepted mouse and speech input modalities method which FlexPhoReS makes use of to bring immediate and easy interaction between users and system application.

8.3.2 Implications for system developers

Firstly, it essential that web based personal digital photo retrieval system developers pay more attention to user needs when designing user interface. The details and structured experiment in this study (including equipment employed for the procedures and tasks) were unique and should benefit system developers and researchers as a guideline for planning, designing and conducting user interface evaluation in related area of research. Secondly, the study provides a model of a flexible web based personal digital photo retrieval that integrates the multimodal user interface into an end-user interface that researchers and developers can exploit in designing the user interface for information retrieval.

8.3.3 Implications for digital photo searchers

By using the FlexPhoReS system, digital photo searchers may utilise mouse and keyboard input modalities and mouse and speech input modalities to retrieve digital photos through a web browser. The system is also capable of
accommodating users at anytime and anywhere across the World Wide Web networks. Thus, the issue of forcing the user to abandon their favourite platform and install complicated third party software or components does not arise.

8.4 Limitations and suggestions for further work

The research is not without its limitations. During the pre-development of the FlexPhoReS model, a user study was conducted which was limited to a small number of participants and concentrated only on the needs of individuals. The design of the prototype model might have been improved if further analysis and observation had been conducted in participants' natural environments, with more representative participants, for example, focused specifically on the needs of groups or families. This is because most photo collections belong to a household, not an individual. Such a study might also yield very useful qualitative and quantitative results.

In the prototype development and implementation stage, the prototype was constructed using current enabling technology including the World Wide Web, Speech technologies and SALT, MATLAB, ASP and JavaScript. Although the combination of these technologies does provide flexibility for user interaction, the photo retrieval system engine that supports visual example searching would not be efficient enough for a commercial system. In addition, the database contains only a small number of digital photos. The intention was to create a flexible user interface of the prototype and search task times in the evaluation did not take into account the system performance times. The number of photos in the prototype should not make any difference to the human search process even when increased.

Additionally, the system is not yet secure enough and depends too much on the internet infrastructure. Utilising the internet infrastructure may provide security problems for the system. The flexibility provided by the prototype met the aim of
the design model, but it could be improved further. More advanced photo retrieval system engine support, such as advanced browsing and searching (for example, Boolean) should be provided.

Users can access the application through a desktop or laptop but not through their mobile phone or personal digital assistant (PDA). Therefore, the use of multimodal user interface for web based personal digital photo retrieval system in handheld devices including personal digital assistants (PDA) and mobile phone should be investigated further.

The study was also limited to a consideration of retrieval tasks. Before photos can be retrieved, all of the digital photos were manually annotated. The digital photo annotation process is time-consuming and difficult especially when the photos are complex. The digital photo annotation facility with multimodal user interface needs to be further investigated in order to let users organise their own digital photo collection before they can retrieve them easily.

The user evaluation might have involved more participants. The evaluation of user performance was limited to user search performance based on time taken to complete search tasks. The evaluation could have been extended to include broader user performance measurement criteria such as frequency of errors made, time needed to recover from error and frequency of accessing help.

Despite its many caveats, this study has yielded preliminary evidence revealing that the model and the FlexPhoReS prototype already provides a good basis for supporting flexibility in the web based personal digital photo retrieval interaction process.
8.5 Concluding remarks

To conclude, this research has demonstrated that the prototype of flexible user interface for web based personal digital photo retrieval system is acceptable to the users. The prototype is successful in providing a flexible model of the photo retrieval process by offering alternative input modalities through a multimodal user interface. The prototype implementation was successful in providing tools to implement photo retrieval tasks and support the proposed model.

It is clear that all of the participants had significantly improved their digital photo search performance when using mouse and speech input modalities. Participants also were significantly more satisfied with mouse and speech input modalities than with mouse and keyboard input modalities.

The state-of-the-art of speech, image retrieval and World Wide Web technology has reached a level that allows us to build and develop a multimodal user interface for web based personal digital photo retrieval. This interface provides the user with multiple and alternative modes of interfacing with a web based personal digital photo retrieval system beyond the traditional mouse and keyboard input modalities. It allows users to select the modality that best fits the situation and is not likely to replace traditional modes of input, but seems to have a useful place along with other types of input modalities in the system user interface. Therefore the flexibility of a web based personal digital photo retrieval system user interface potentially could provide a usable system for more individuals in more various environments.

As a whole, this research has extended current technology of web based personal digital photo retrieval by providing a flexible user interface for a web based personal digital photo retrieval system. The results and the findings obtained have indicated that the objectives outlined have all been met. As one of the first studies
in this area, it is believed that findings create substantial opportunities for further research.
Bibliography


209


Kay, R., 2002. QuickStudy: system development life cycle, *Computerworld*, [http://www.computerworld.com/id,0,10801,71151,00.html], [accessed 07.08.05].


APPENDIX 1

Interview Questionnaires

Objective:
To understand how digital photo users organise and retrieve their personal digital photo collections

Procedure:
A small number of digital photo users were recruited and asked the same set of questions in a structured interview while recording and taking notes of their answers. The users were asked about their current practices, and then gave their opinions on possible features for organising and retrieving their personal digital photo collections.

1. Explain purpose of interview
2. Ask interviewee's permission to tape-record the interview
3. Write name on interview sheet and record start time
4. Zero the tape counter
5. Switch tape to record
6. Write down key points of interview immediately after over

Respondents' background knowledge on using computers, the Internet and the use of digital photo capture devices.

1. On average, how often do you use a computer?
2. What do you primarily use the computer for?
3. How familiar are you with Windows-based applications, Macintosh based applications, Unix-based applications or other (specify)?
4. On average, how often do you use the Internet?
5. What do you primarily use the Internet for?
6. How did you obtain your personal digital photo collection?
7. Do you have any capturing devices (e.g. digital camera, digital camcorder, etc)?
8. How long have you owned your digital capturing device?
9. What type of storage does your digital capturing device use?
10. On average, how often do you use your digital capturing device?
11. Why do you capture digital photos instead of traditional film photos?
12. Are you satisfied with your digital capturing device? If satisfied, why? If unsatisfied, why not?

Digital photo management system: Respondents' experiences.

1. Have you ever used any software to manage your digital photo collection?
2. If yes, what is the digital photo management software you use?
3. In general are you satisfied with the software? Why yes? And if not, why not?
4. What limitations/difficulties have you found when using the software?
5. What features would you want your digital photo management software to include that it doesn't already have?
6. What aspects are difficult and what are easy about managing your digital photo collection?
7. Do you agree it would be a good idea if you could manage your personal digital photo collection through the web, so you could archive it at any time and anywhere?
8. If yes, why? And if not, why not?

Storage of digital photos

1. How big would you say your collection of digital photos is?
2. What do you do with your digital photos as soon as you get them from your digital camera, scanner or elsewhere?
3. Do you organise good and bad photos separately? If yes what criteria do you use to judge the quality of your digital photo as "good" and as "bad"?
4. How do you organise your good digital photos and bad/unused digital photos?
5. Do you write anything down on the photo, file name, directory folder or elsewhere? If so, what kind of things do you write?

Sharing of digital photos

1. If you have a number of potential users (probably among your family, friends and relatives) for certain digital photos, do you want to share the digital photos with them? If yes, what mechanism might you choose to share the digital photos? And why have you chose this mechanism.
2. If no, why not?
3. Do you agree it would be a good idea if digital photo collections could be as secure as possible to withstand a variety of threats from unauthorised users and hackers?

Retrieval of digital photos

1. On average, how often do you browse your digital photo collection?
2. How easy or hard is it for you to find appropriate digital photos in your collection when you need them?
3. What browsing and query approaches have you used most often? And why have you chosen these approaches?
4. Do you agree it would be a good idea if you could browse a large number of digital photos with visually "similar" photos grouped together? If yes, why? And if not, why not?
5. Do you agree it would be a good idea if you could retrieve digital photos by the text of any extra notes annotated with the digital photo? If yes, why? And if not, why not?
6. There are visual aspects of an image, for example, 3 elements: shape, colour and texture. When you are looking at a particular digital photo, do you agree it would be a good idea if you could retrieve other digital photos which are visually "similar" by shape, colour and texture? If yes, why? If not, why not?
7. Do you agree it would be a good idea if a speech interface was used as an alternative user interface to retrieve digital photo and control the application menu
8. Would you prefer to retrieve digital photos and control the application menu by using the mouse and keyboard or by speech?
APPENDIX 2

Low fidelity prototype evaluation: Structured interview questionnaires

Personal background
1. Respondent name/identification?
2. Gender?
3. Age range?

Computer and internet/web experiences
4. How long have you used computers?
5. How often do you use the computer?
6. How long have you used the internet or web applications?
7. How often do you use the internet or web application?

Number in digital photo collection
8. How big is your collection of digital photos?

Screen design
9. What do you think of the colours used on the screen?
10. What do you think of the characters (fonts) on the screen?
11. What do you think of the amount of information that is displayed on screen?
12. What do you think of the arrangement of information that is displayed on screen?
13. What do you think of sequence of screens?
14. What do you think of the screen design?
15. What do you like about it?
16. What do you do not like about it?
17. What would you like to see from the screen?
18. What would you expect to see on the screen?

Learning how to operate the FlexPhoReS system

Login:
19. How do you find learning to operate the system particularly the login into FlexPhoReS?
20. What do you think of remembering names and use of commands for login?
21. Can the task be performed in a straightforward manner?

Navigate and browse:
22. How do you find learning to operate the system, particularly to control and browse?
23. What do you think of remembering names and use of commands to control and browse?
24. Can the task be performed in a straightforward manner?

Search by keywords:
25. How do you find learning to operate the system particularly to search by keywords?
26. What do you think of remembering names and use of commands to search by keywords?
27. Can the task be performed in a straightforward manner?

Search by visual example:
28. How do you find learning to operate the system particularly to search by visual example?
29. What do you think of remembering names and use of commands to search by visual example?
30. Can the task be performed in a straightforward manner?

Others
31. Is there anything else you want to say?
Appendix 4: Low fidelity FlexPhoReS prototype (Microsoft PowerPoint version)

FlexPhoReS
Low fidelity prototype evaluation
Screen testing

by
N.A Ismail
Department of Information Science,
Loughborough University, United Kingdom

3 Phases
Phase 1: Login to FlexPhoReS
Phase 2: Retrieval strategies
Phase 3: What can the user say?
APPENDIX 5

Recruitment questionnaire

PART 1: Demographics and number of photos in collection
Please tick boxes where appropriate.

1.1 Occupation: ________________________________

1.2 Institution: ________________________________

1.3 Gender: € Male € Female

1.4 Age range:
€ 18 – 24 € 25 – 34 € 35 – 44 € 45 or above

1.5 What is your highest educational background achieved?
€ Primary school € Postgraduate University
€ Secondary school € Others (please specify) ______________
€ Undergraduate University

1.6 On average, how big would you say your collection of digital photos is?
€ Less than 100 € 100 - 200 € 201 - 300
€ 301 - 400 € More than 401

PART 2: Computer and related system experience
Please tick boxes where appropriate.

2.1 How long have you been using computers?
€ Less than 1 year € 1 - 2 years € 3 - 5 years € 6 - 10 years € More than 10 years

2.2 Have you ever used any web applications (i.e. Google/Yahoo image search, eBay) that have digital photo or image retrieval features such as keyword searching, browsing and visual example searching?
€ Yes € No

2.3 If yes, how long have you been using systems with these features?
€ Less than 1 year € 1-2 years € 3-5 years € Over 5 years

2.4 If yes, how often do you use these system features?
€ Only used once before € Regularly, up to 4 times per month € Rarely € More than 5 times per month

2.5 Of the following digital photo retrieval strategies, check those that you have personally used:
€ Searching by keyword € Others (please specify) ______________
€ Searching by browsing € No experience € Searching by visual example
2.6 Of the following input devices and systems, check those that you have used:
   - Keyboard
   - Mouse
   - Microphone
   - Touch screen
   - Pen based computing
   - Microsoft Internet Explorer web browser
   - Voice recognition system
   - Other (please specify) _______

2.7 Have you ever used any web based application with mouse and speech input modalities?
   - Yes
   - No

2.8 If yes, how long have you been using the system?
   - Less than 1 year
   - 1 -2 years
   - Over 5 years

2.9 If yes, how often do you use web applications with mouse and speech input modalities?
   - Only used once before
   - Regularly, up to 4 times per month
   - Rarely
   - More than 5 times per month
Appendix 6

Brief introduction of the test and experimental procedures

In this test we are evaluating a proposed multimodal user interface for an online digital photo retrieval prototype system. The prototype is a web-based photo retrieval system which allows users more flexibility in performing photo retrieval in their personal digital photo collection. The prototype user interface allows users to use either mouse and keyboard input modalities or mouse and speech input modalities to control the application and retrieve personal photos through a web environment.

The test will take approximately 30 to 40 minutes to complete. The feedback that we get during the test will help us to find out whether the proposed user interface features are usable and work well with the users.

We will record information about how you use the prototype software and we will ask you to fill out a questionnaire.

With a given user ID and password, each participant will be allowed 20 minutes for free exploration of the prototype.

During the free exploration, you are encouraged to talk about what you are looking at and what you are thinking about. You can also ask any related questions about the prototype user interface.

The set of tasks will be given to you and are asked to work on your own.

If any task takes more than 15 minutes to complete, you will be asked to stop and proceed to the next task.

If you feel that you are unable to complete a task and want to move on, this is fine. You are also free to consult online help available in the prototype.

After completing all the tasks, you will be asked to complete a questionnaire on search tasks, subjective satisfaction, suitability and flexibility of the prototype user interface.

If you have any questions, you may ask now or at any time during the test.

Thank you.
**APPENDIX 7**

**Evaluation questionnaires on search tasks, user satisfaction, suitability and flexibility of the prototype**

**PART 1: FlexPhoReS with mouse and keyboard input modalities**

Please circle the numbers which most appropriately reflect your impressions about using mouse and keyboard input modalities in FlexPhoReS system. Not Applicable = NA.

<table>
<thead>
<tr>
<th></th>
<th>Overall reactions to the system</th>
<th>terrible</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>Wonderful</th>
<th>NA</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>frustrating</td>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>Satisfying</td>
<td>NA</td>
</tr>
<tr>
<td>3</td>
<td>dull</td>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>Stimulating</td>
<td>NA</td>
</tr>
<tr>
<td>4</td>
<td>difficult</td>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>easy</td>
<td>NA</td>
</tr>
<tr>
<td>5</td>
<td>inadequate</td>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>adequate power</td>
<td>NA</td>
</tr>
<tr>
<td>6</td>
<td>rigid</td>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>flexible</td>
<td>NA</td>
</tr>
<tr>
<td>7</td>
<td>Login into the system</td>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>easy</td>
<td>NA</td>
</tr>
<tr>
<td>8</td>
<td>To know what you can type into the specific textbox</td>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>easy</td>
<td>NA</td>
</tr>
<tr>
<td>9</td>
<td>Retrieve photos by browsing</td>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>easy</td>
<td>NA</td>
</tr>
<tr>
<td>10</td>
<td>Retrieve photos by keyword searching</td>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>easy</td>
<td>NA</td>
</tr>
<tr>
<td>11</td>
<td>Retrieve photos by visual example searching</td>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>easy</td>
<td>NA</td>
</tr>
</tbody>
</table>

Please write your comments about the mouse and keyboard input modalities with the system here:

________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
PART 2: FlexPhoReS with mouse and speech input modalities
Please circle the numbers which most appropriately reflect your impressions about using mouse and speech input modalities in FlexPhoReS system. Not Applicable = NA.

<table>
<thead>
<tr>
<th></th>
<th>Overall reactions to the system</th>
<th>terrible</th>
<th>1 2 3 4 5 6 7 8 9</th>
<th>Wonderful</th>
<th>NA</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>frustrating</td>
<td>1 2 3 4 5 6 7 8 9</td>
<td>Satisfying</td>
<td>NA</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>dull</td>
<td>1 2 3 4 5 6 7 8 9</td>
<td>Stimulating</td>
<td>NA</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>difficult</td>
<td>1 2 3 4 5 6 7 8 9</td>
<td>easy</td>
<td>NA</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>inadequate power</td>
<td>1 2 3 4 5 6 7 8 9</td>
<td>adequate power</td>
<td>NA</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>rigid</td>
<td>1 2 3 4 5 6 7 8 9</td>
<td>flexible</td>
<td>NA</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Login into the system</td>
<td>difficult</td>
<td>1 2 3 4 5 6 7 8 9</td>
<td>easy</td>
<td>NA</td>
</tr>
<tr>
<td>8</td>
<td>To know what you speak into the specific textbox.</td>
<td>difficult</td>
<td>1 2 3 4 5 6 7 8 9</td>
<td>easy</td>
<td>NA</td>
</tr>
<tr>
<td>9</td>
<td>Retrieve photos by browsing</td>
<td>difficult</td>
<td>1 2 3 4 5 6 7 8 9</td>
<td>easy</td>
<td>NA</td>
</tr>
<tr>
<td>10</td>
<td>Retrieve photos by keyword searching</td>
<td>difficult</td>
<td>1 2 3 4 5 6 7 8 9</td>
<td>easy</td>
<td>NA</td>
</tr>
<tr>
<td>11</td>
<td>Retrieve photos by visual example searching</td>
<td>difficult</td>
<td>1 2 3 4 5 6 7 8 9</td>
<td>easy</td>
<td>NA</td>
</tr>
</tbody>
</table>

Please write your comments about the mouse and speech input modalities with the system here:

________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
PART 3: Suitability and Flexibility
Please tick boxes where appropriate and circle the numbers which most appropriately reflect your impressions about FlexPhoReS flexibility. Not Applicable = NA.

<table>
<thead>
<tr>
<th></th>
<th>Mouse and keyboard input modalities are suitable for photo retrieval process</th>
<th></th>
<th>Mouse and speech input modalities are suitable for photo retrieval process</th>
<th></th>
<th>Both input modalities i) mouse and speech; and ii) mouse and keyboard input modalities are suitable for photo retrieval process</th>
<th></th>
<th>Both input modalities are complementary each other to retrieve photos</th>
<th></th>
<th>FlexPhoReS offers more flexibility than a system that only offers one style of input modalities either mouse and speech support or mouse and keyboard support</th>
<th></th>
<th>FlexPhoReS offers more flexibility on the World Wide Web environment than in a stand-alone environment</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>strongly 1 2 3 4 5 6 7 8 9 strongly disagree agree NA</td>
<td>2</td>
<td>strongly 1 2 3 4 5 6 7 8 9 strongly disagree agree NA</td>
<td>3</td>
<td>strongly 1 2 3 4 5 6 7 8 9 strongly disagree agree NA</td>
<td>4</td>
<td>strongly 1 2 3 4 5 6 7 8 9 strongly disagree agree NA</td>
<td>5</td>
<td>strongly 1 2 3 4 5 6 7 8 9 strongly disagree agree NA</td>
<td>6</td>
<td>strongly 1 2 3 4 5 6 7 8 9 strongly disagree agree NA</td>
<td></td>
</tr>
</tbody>
</table>

Others
Please tick boxes where appropriate

1a In FlexPhoReS, which input modalities (mouse and speech or mouse and keyboard) did you like the best to use for photo retrieval and why?
- Mouse and speech input modalities
- Don't care
- Mouse and keyboard input modalities
- Don't know
- Switch between both input modalities

1b Why? ____________________________________________________

Please write any other comments you had about the system here:
__________________________________________________________
__________________________________________________________
__________________________________________________________

- Thank you -
APPENDIX 8
Characteristics of Participants

<table>
<thead>
<tr>
<th>Participants</th>
<th>Occupation</th>
<th>Age group</th>
<th>Gender</th>
<th>Computer use</th>
<th>Educational background</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Research student</td>
<td>25 to 34</td>
<td>Male</td>
<td>More than 10 years</td>
<td>Postgraduate University</td>
</tr>
<tr>
<td>2</td>
<td>Machine operator</td>
<td>35 to 44</td>
<td>Male</td>
<td>3 to 5 years</td>
<td>Secondary school</td>
</tr>
<tr>
<td>3</td>
<td>Research student</td>
<td>25 to 34</td>
<td>Male</td>
<td>More than 10 years</td>
<td>Postgraduate University</td>
</tr>
<tr>
<td>4</td>
<td>Lecturer</td>
<td>45 or above</td>
<td>Female</td>
<td>More than 10 years</td>
<td>Postgraduate University</td>
</tr>
<tr>
<td>5</td>
<td>Research student</td>
<td>35 to 44</td>
<td>Male</td>
<td>More than 10 years</td>
<td>Postgraduate University</td>
</tr>
<tr>
<td>6</td>
<td>Postgraduate student</td>
<td>18-24</td>
<td>Male</td>
<td>3-5 years</td>
<td>Undergraduate University</td>
</tr>
<tr>
<td>7</td>
<td>Research student</td>
<td>35 to 44</td>
<td>Female</td>
<td>More than 10 years</td>
<td>Postgraduate University</td>
</tr>
<tr>
<td>8</td>
<td>Research student</td>
<td>25 to 34</td>
<td>Male</td>
<td>More than 10 years</td>
<td>Postgraduate University</td>
</tr>
<tr>
<td>9</td>
<td>Postgraduate student</td>
<td>25 to 34</td>
<td>Male</td>
<td>3 to 5 years</td>
<td>Undergraduate University</td>
</tr>
<tr>
<td>10</td>
<td>Postgraduate student</td>
<td>35 to 44</td>
<td>Male</td>
<td>More than 10 years</td>
<td>Undergraduate University</td>
</tr>
<tr>
<td>11</td>
<td>Research student</td>
<td>45 or above</td>
<td>Male</td>
<td>More than 10 years</td>
<td>Postgraduate University</td>
</tr>
<tr>
<td>12</td>
<td>Automotive consultant</td>
<td>45 or above</td>
<td>Male</td>
<td>More than 10 years</td>
<td>Postgraduate university</td>
</tr>
<tr>
<td>13</td>
<td>Teacher</td>
<td>25-34</td>
<td>Female</td>
<td>More than 10 years</td>
<td>Undergraduate University</td>
</tr>
<tr>
<td>14</td>
<td>Research student</td>
<td>35-44</td>
<td>Male</td>
<td>More than 10 years</td>
<td>Postgraduate university</td>
</tr>
<tr>
<td>15</td>
<td>Postgraduate Student</td>
<td>18-24</td>
<td>Male</td>
<td>6 to 10 years</td>
<td>Undergraduate university</td>
</tr>
<tr>
<td>16</td>
<td>Machine operator</td>
<td>25-34</td>
<td>Male</td>
<td>3 to 5 years</td>
<td>Undergraduate university</td>
</tr>
<tr>
<td>17</td>
<td>Research student</td>
<td>25-34</td>
<td>Male</td>
<td>More than 10 years</td>
<td>Postgraduate university</td>
</tr>
<tr>
<td>18</td>
<td>Research student</td>
<td>25-34</td>
<td>Male</td>
<td>6-10 years</td>
<td>Undergraduate University</td>
</tr>
<tr>
<td>19</td>
<td>Undergraduate student</td>
<td>18-24</td>
<td>Male</td>
<td>More than 10 years</td>
<td>Diploma</td>
</tr>
<tr>
<td>20</td>
<td>Research student</td>
<td>35-44</td>
<td>Male</td>
<td>More than 10 years</td>
<td>Postgraduate university</td>
</tr>
</tbody>
</table>
APPENDIX 9i

Normal Q-Q Plot of total MK search completion time

Normal Q-Q Plot of total MS search completion time

Tests of Normality

<table>
<thead>
<tr>
<th></th>
<th>Kolmogorov-Smirnov(a)</th>
<th>Shapiro-Wilk</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Statistic</td>
<td>df</td>
</tr>
<tr>
<td>total MK search</td>
<td>.128</td>
<td>20</td>
</tr>
<tr>
<td>completion time</td>
<td></td>
<td></td>
</tr>
<tr>
<td>total MS search</td>
<td>.132</td>
<td>20</td>
</tr>
<tr>
<td>completion time</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* This is a lower bound of the true significance.
a Lilliefors Significance Correction
APPENDIX 9ii

Normal Q-Q Plot of MK browsing

Normal Q-Q Plot of MS browsing

Histogram

Histogram

Tests of Normality

<table>
<thead>
<tr>
<th></th>
<th>Kolmogorov-Smirnov(a)</th>
<th>Shapiro-Wilk</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Statistic</td>
<td>df</td>
</tr>
<tr>
<td>MK browsing</td>
<td>.089</td>
<td>20</td>
</tr>
<tr>
<td>MS browsing</td>
<td>.223</td>
<td>20</td>
</tr>
</tbody>
</table>

* This is a lower bound of the true significance.

a Lilliefors Significance Correction
Tests of Normality

<table>
<thead>
<tr>
<th></th>
<th>Kolmogorov-Smirnov(a)</th>
<th>Shapiro-Wilk</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Statistic</td>
<td>df</td>
</tr>
<tr>
<td>MK keyword</td>
<td>.115</td>
<td>20</td>
</tr>
<tr>
<td>MS keyword</td>
<td>.215</td>
<td>20</td>
</tr>
</tbody>
</table>

* This is a lower bound of the true significance.

a Lilliefors Significance Correction
APPENDIX 9iv

Tests of Normality

<table>
<thead>
<tr>
<th></th>
<th>Kolmogorov-Smirnov(a)</th>
<th>Shapiro-Wilk</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Statistic</td>
<td>df</td>
</tr>
<tr>
<td>MK Vis example</td>
<td>.191</td>
<td>20</td>
</tr>
<tr>
<td>MS Vis example</td>
<td>.128</td>
<td>20</td>
</tr>
</tbody>
</table>

* This is a lower bound of the true significance.

a Lilliefors Significance Correction
Appendix 10: FlexPhoReS source codes files (CDROM attached) and services

- FlexPhoReS page (home.asp)
  - main.asp
  - login.asp
  - whatcaniseay_id.asp
  - profile.asp
  - profile_source.asp
  - about.asp
  - about_source.asp
  - help.asp
  - help_source.asp
  - logout.asp
  - browse_event.asp
  - browse_places.asp
  - browse_subjects.asp
  - browse_times.asp
  - whatcanisay1.asp
  - whatcanisay1_source.asp
  - keyword_search.asp
  - whatcanisay2.asp
  - whatcanisay2_source.asp
  - whatcanisay3.asp
  - flexphores.m
  - matweb.conf
  - matweb.exe
  - visualout2.asp
  - whatcanisay3_source.asp

Microsoft Internet Information Services

- Web Server operating system
  - MATLAB Server services
  - Data Sources (ODBC)

- Client operating system
  - Microsoft Windows XP or Microsoft Windows 2000

- Speech engine
  - Microsoft speech recognition
  - Microsoft speech synthesizer

- Multimodal web browser
  - Internet Explorer version 6.0 with speech add-in
The CD-ROM disc included in this thesis contains the FlexPhoReS prototype source codes files. Server site configurations are needed in order to run the whole system in web environment.